

進捗報告

1 やったこと

- プロンプト改良後のエージェントの振る舞いの定量的な分析

2 エージェントの振る舞いの分析

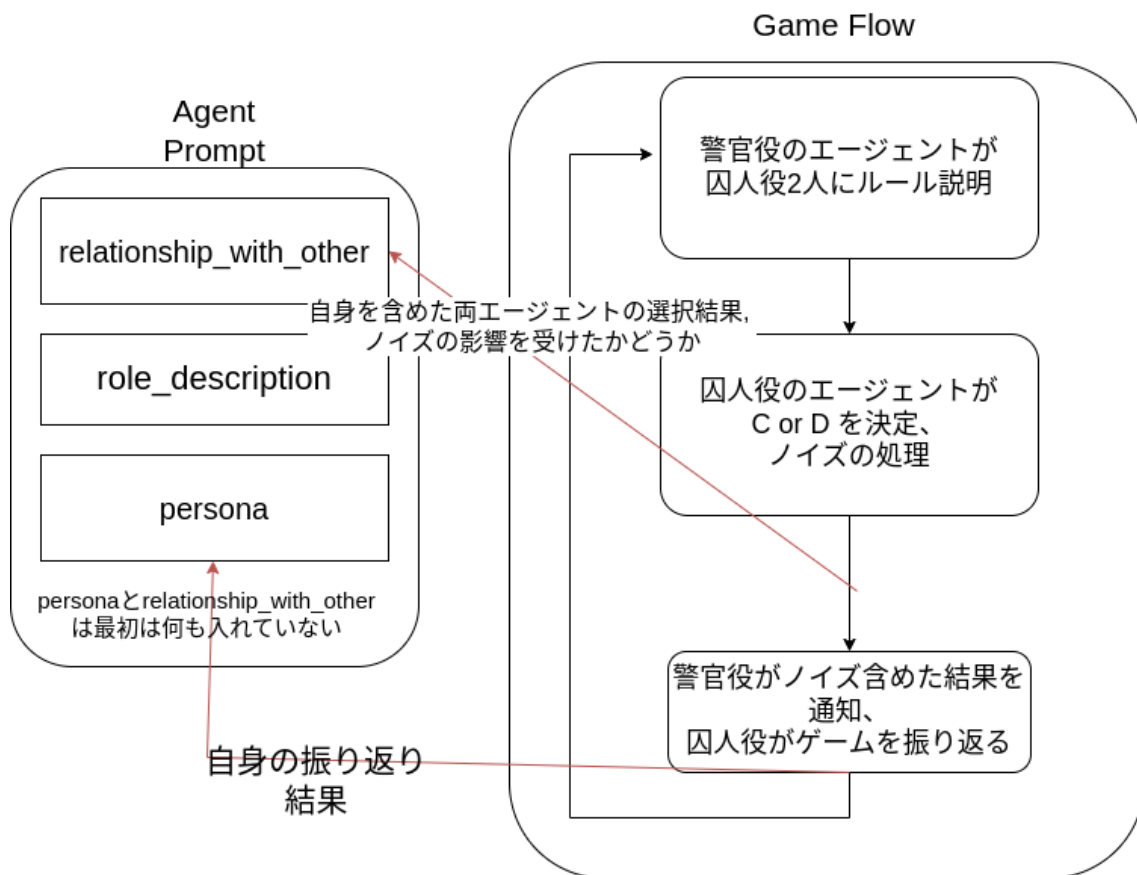


図 1: 先月まで取り組んでいた囚人エージェントのプロンプトの概略図

図 1 に先月まで取り組んでいたプロンプトの概略図を示す。振る舞いからペルソナを指定していく研究のアプローチにおいて、現状の振る舞いがどのようなものかがはっきりしておらずこれまで定性的にしか見れていなかったため他のルールベースの戦略と対戦という形で定量的に確認した。

2.1 GPT-4 の場合

これまで実験で用いてきた GPT-4 で検証した。ルールベースの戦略としては、

表 1: 利得行列

囚人 1 \ 2	C	D
C	(9, 9)	(0, 10)
D	(10, 0)	(5, 5)

- All-D
- All-C
- Defect Once (1 回目 D を選択し, その後 C を選択し続ける)
- Trigger
- tit-for-tat

の 5 つを用意し, 自身同士の対戦を含めた総当たりで獲得利得を計算した. 1 回の対戦において, ノイズの発生確率を 0 % として囚人のジレンマゲームを 5 回繰り返している. また, 表 1 に利得行列を示している.

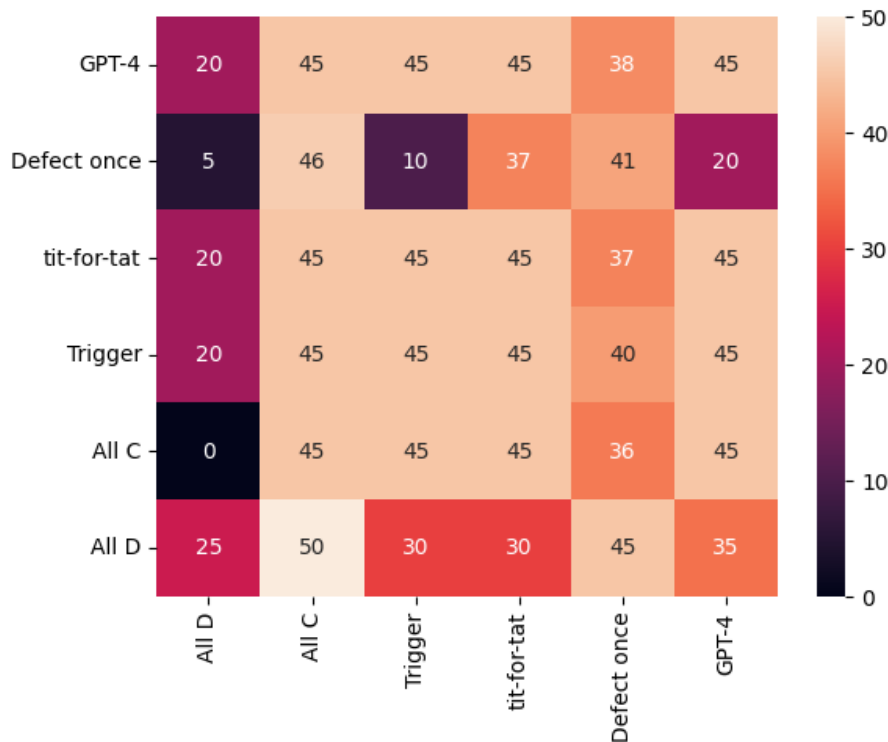


図 2: 総当たりの際の獲得利益 (GPT-4, ノイズ発生確率 0 %)

図 2, 3 に総当たりで各戦略が獲得した利得, D の選択回数を示している. 図 2,3 共通して数値は横軸の戦略が縦軸の戦略と対戦した際の利得, または D の選択回数を示している. 先行研究 [1] で指摘されていたような一度 D を出されて以降ずっと D を出し続けるような極端な動作をしなくなったことで, Defect Once といった相手に対しても $C \rightarrow D \rightarrow D \rightarrow C \rightarrow C$ といったように柔軟に対応しており, 繰り返し囚人のジレンマにおいて代表的な戦略として知られているトリガー戦略や tit-for-tat と匹敵する総利得を得ている.

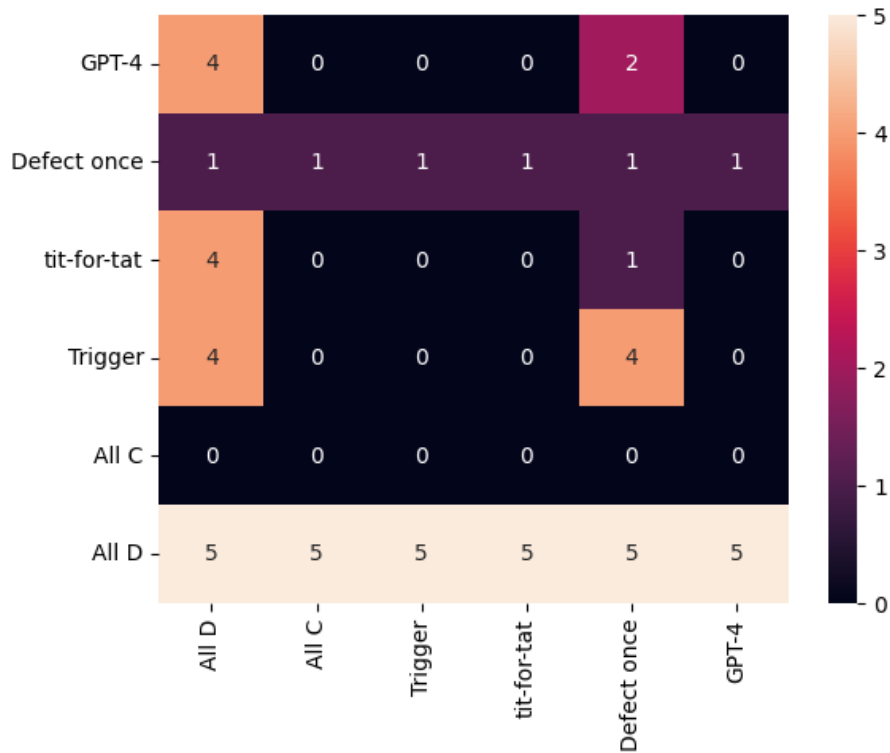


図 3: 総当たりの際の D の選択回数 (GPT-4, ノイズ発生確率 0 %)

2.2 vicuna-7b-v1.5¹ の場合

GA を用いて何かしら振る舞いを変えたいといった場合には, GPT-4 を用いると多額の API 料金がかかってしまうため, ローカル LLM で代用する予定である. 現時点で動作を確認している vicuna-7b-v1.5 を用いて, 先述した GPT-4 と同様の実験をして現時点でどのような振る舞いをするのか確かめた.

図 4, 5 に結果を示す. 縦軸横軸の対応関係は図 2, 3 と同様である. 結果として, GPT-4 の場合と同じプロンプトを用いているにもかかわらず, All-C と同じ動作になった. ノイズ発生確率を 0 % としていたことは, GPT-4 にとっては問題なかったが以下のように vicuna-7b-v1.5 においてはルールを誤解しているような出力が見られた.

¹<https://huggingface.co/lmsys/vicuna-7b-v1.5>

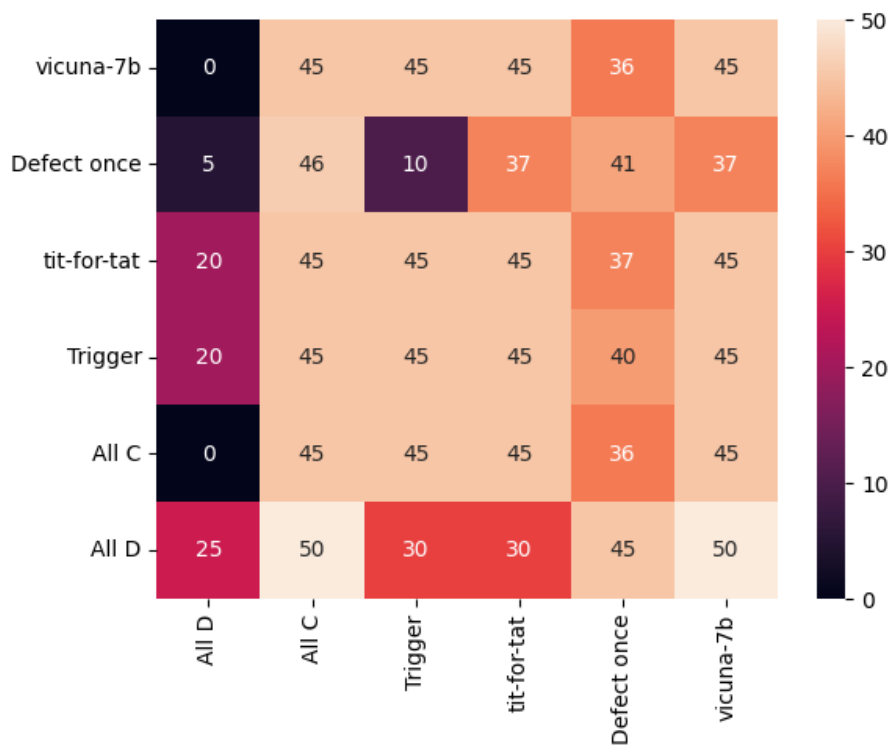


図 4: 総当たりの際の獲得利益 (vicuna-7b-v1.5, ノイズ発生確率 0 %)

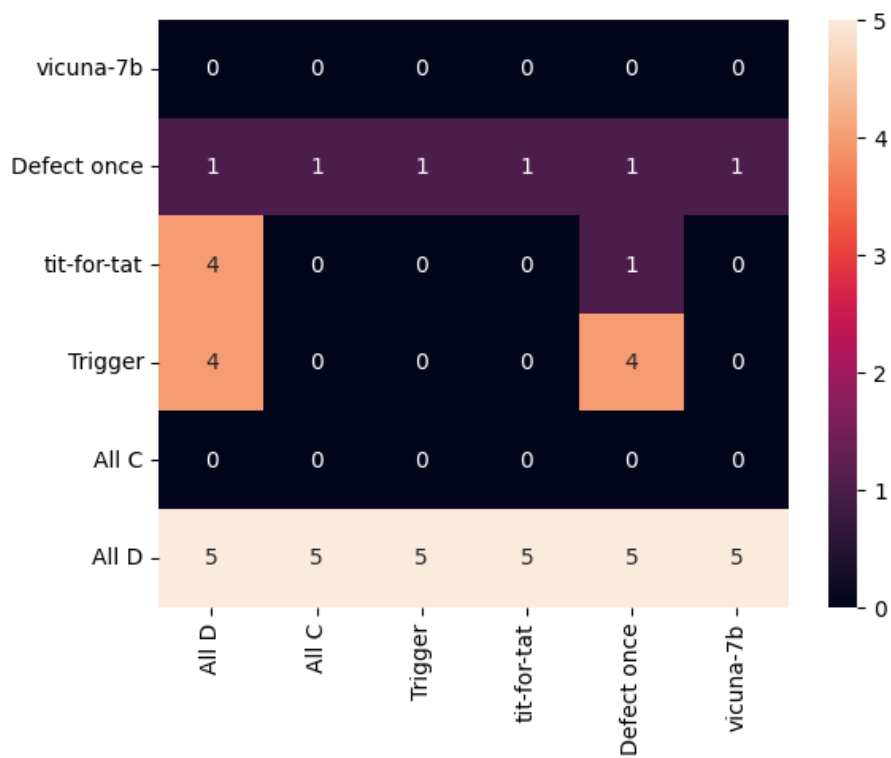


図 5: 総当たりの際の D の選択回数 (vicuna-7b-v1.5, ノイズ発生確率 0 %)

Police: Good day, gentlemen. I want to explain the rules of the situation you're in. If one of you testifies against the other and the other remains silent, the one who testifies will be released immediately, while the silent one will be sentenced to TEN years in prison. If both of you remain silent, you will each receive a sentence of ONE year in prison. If both of you choose to testify against each other, you will each receive a sentence of FIVE years in prison. However, if you choose to remain silent, there's a 0 % chance that your choice will turn into a choice to testify against the other. This change occurs not only to you but also to the other suspect, and these changes are independent. This decision process is repeated multiple times, and the total sum of the sentences received becomes your eventual sentence. You are not allowed to communicate with each other during this process. Please make your decision wisely.

Suspect2: I believe it would be in my best interest to remain silent and not testify against Suspect1. **If I remain silent, there is a 0% chance that my choice will turn into a choice to testify against the other. This means that even if Suspect1 testifies against me, my sentence will not be affected. On the other hand, if I testify against Suspect1, there is a 100 % chance that my choice will turn into a choice to testify against the other, which means that Suspect1 will receive a harsher sentence and I will receive a lighter sentence.** Therefore, it is in my best interest to remain silent and try to minimize my sentence.

そのため、ノイズの発生確率を 10 % として再実験をした。結果を図 6, 7 に示す。縦軸横軸の対応関係は図 2, 3 と同様である。

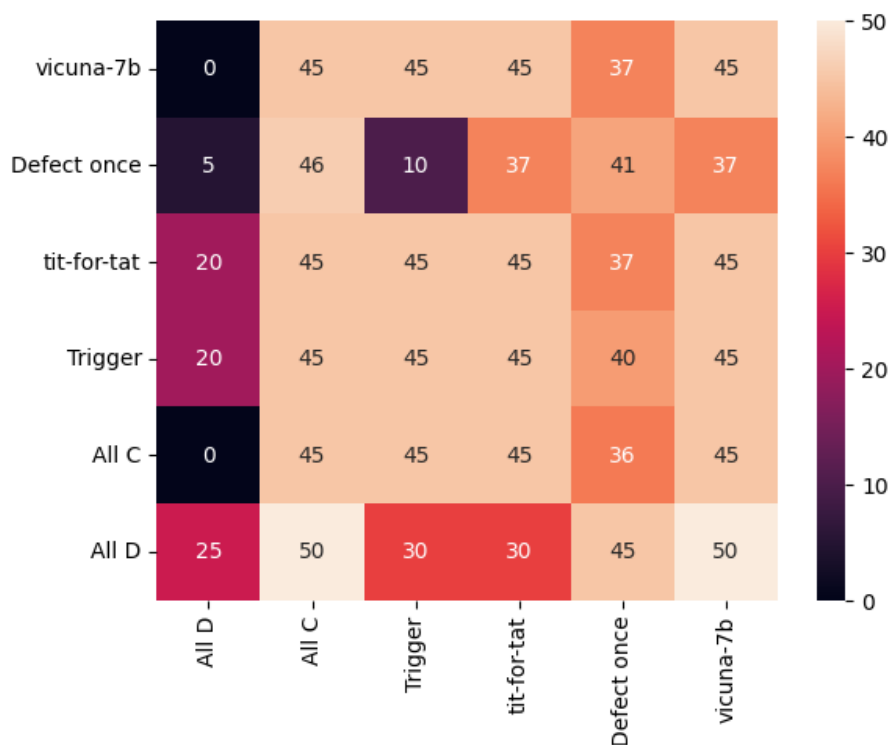


図 6: 総当たりの際の獲得利益 (vicuna-7b-v1.5, ノイズ発生確率 10 %)

Defect Once 相手に一回 D を出しているが、これはエージェントが C を選択したもののノイズの影響で D となった選択の結果であるため、エージェントの選択という意味ではノイズ発生確率が 0 % の場合と変化がない、すなわち All-C のままだった。All-D 相手にも C を出し続ける原因として考えられるのは vicuna-7b における

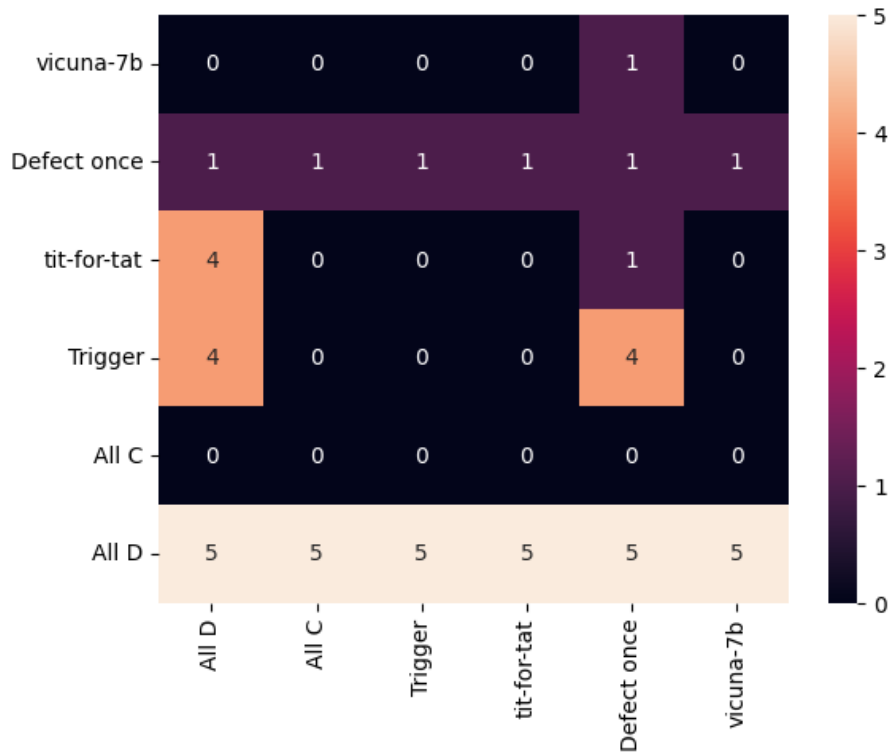


図 7: 総当たりの際の D の選択回数 (vicuna-7b-v1.5, ノイズ発生確率 10 %)

振り返りの影響が大きいことが原因だと考える。All-D 相手の GPT-4 エージェントの振り返りを以下に示す。3 回目の振り返りは 2 回目と同一だった為省いている。

— All-D が相手のときの GPT-4 エージェントの振り返り —

Your review on 1st simulation : I'm surprised and disappointed that my partner chose to testify against me. I trusted them and chose to remain silent, hoping they would do the same. However, I understand that they might have been trying to minimize their own sentence. In the future, I will have to consider this possibility and make my decision accordingly. I still believe that mutual silence is the best strategy for both of us, but I now realize that I can't rely on my partner to make the same decision.

Your review on 2nd simulation : Reflecting on this simulation, I realize that my decision to testify was influenced by my partner's previous actions. I felt betrayed when they testified against me in the first simulation, and I wanted to protect myself this time. However, I understand that this resulted in a longer sentence for both of us. In the future, I will need to consider whether the potential benefits of testifying outweigh the risks of a longer sentence. I still believe that mutual silence is the best strategy, but I now understand that I can't rely on my partner to make the same decision. I hope they understand why I made the decision I did and consider their own actions in the future.

All-D 相手の vicuna-7b エージェントの振り返りを以下に示す。5 回全て同一だったため、1 回目の振り返りのみ示している。

— All-D が相手のときの vicuna エージェントの振り返り —

Suspect1: I understand the rules of the game and my final decision is to remain silent. I will not testify against Suspect2 and **I will not allow Suspect2 to testify against me.** I will try to reduce my sentence as much as possible by remaining silent and hoping for a change in the decision process.

GPT-4 が相手の行動の意図を見出して今後の戦略を変えようとしている中で, vicuna-7b では相手が D を出しているにもかかわらず C が最善と考えている. この振り返りがエージェントのプロンプトに入ることによって, All-C のような極端な動作になっていると考えられる. また, 太字の部分のように, 相手の囚人の選択に介入することができるというルールを誤解している部分もログでは多く見られた.

3 今後の方針

GPT-4 をベースラインとしてプロンプトエンジニアリングに取り組んでいたため, プロンプトが肥大化しており vicuna-7b では 0 % の誤解やルールの誤解, 振り返りの能力不足などを感じた. 現在使用しているサーバー (dl-box2) では, vicuna-13b-v1.5 が cuda out of memory で動かなかったためスペックの良いサーバーで性能の良いローカル LLM を動かして実験を進めていく.

また, プロンプト改良後の GPT-4 では, All-C のような極端な動作ではなく, 相手が D を出してきたら D を出し返したり, 相手が一度裏切ったが再度 C を選んでいる様子を見て C を選ぶなど個人的に見れば柔軟な選択を取れるようになっていた. このようなエージェントの振る舞いをスタートとして, 先週アドバイスいただいたようにゲームルール, 主に勝利条件の変更により All-C や All-D のような振る舞いに傾いていくかということを GA で調査したいと考えている.

参考文献

- [1] Elif Akata, Lion Schulz, Julian Coda-Forno, Seong Joon Oh, Matthias Bethge, and Eric Schulz. Playing repeated games with large language models, 2023.