自由エネルギー原理に基づく他者の願望推定と共同行動の発生

Formulation of Desire Inference and Emergence of Cooperative Behavior Based on the Free Energy Principle

山口 拓巳 [1*]　竹内 勇剛 [1]
Takumi Yamaguchi[1]　Yugo Takeuchi[1]

[1] 静岡大学
[1] Shizuoka University

Abstract: 他者と円滑にインタラクションが可能な自律エージェントの開発は人工知能分野において重要な課題となっている．このようなエージェントの開発にとって重要な要素の一つは他者の行動の要因となる，願望や信念などの心的状態を推定することである．本稿では自由エネルギー原理とそれに基づく能動的推論を用いて，他者の願望推定と共同行動の発生を数理的にモデル化することに挑戦する．自由エネルギー原理は生物が環境を認識し，行動を選択するための理論であり，ベイズ脳仮説に基づいて人間が環境を認識するための理論である．このモデルは推論と行動を同一の理論の上で扱えるため，行動設計までの流れを明確に行うことができるだけでなく，数理的に記述されており，行動決定の透明性や説明可能性という問題に対して有効であると考えられる．共同行動の発生をモデル化するために，他者の内部状態をエージェントが内部でシミュレートし，その結果をもとに他者の行動計画を推定し，最終的に自己の行動計画を立てるモデルを提案する．このモデルにおける各種推論行為は自由エネルギー原理に基づいて行う．提案したモデルの妥当性を検証するためにシミュレーション実験を行い，提案したモデルが共同行動の発生に必要な情報処理を行うことができることを示した．しかしながら提案モデル内で行われていた願望推定の精度は低く，今後の改良が必要であることも明らかになった．

1　はじめに

近年，他者と円滑にインタラクションが可能な自律エージェントの開発は人工知能分野において重要な課題となっている．このようなエージェントの開発にとって重要な要素の一つは他エージェントの行動の要因となる，願望や信念などの心的状態を推定することである．他者の心的状態の推定とは，自身の中に他者のモデルを構築することと等価であると言え，より一般的には他者の行動や言動といった観測可能な情報を入力とし，その心的状態という観測不可能な状態（隠れ状態）を出力するような関数であると言い換えることができる．

人間の行う共同行動を成功させるためには，他者の信念や願望の推定，それを通した他者の行動の推測が必要となる [1]．したがって共同行動は前述したように他者の心的状態の推定に基づいて行われる行為の一つであると言える．ここで本稿で扱う共同行動とは，中央集権的に指示を与えることなく，各個体が自律的に判断した結果，協力して目標を達成する分散的なシステムにおける共同行動を指す．分散型の共同行動を行うシステムは例えば自動運転システムの実装や，ヒューマノイドロボットへの適用，災害時の救護ロボットへの応用が考えられる [2][3][4].

他者の心的状態の推定に関して，人間が行う推論について心の理論（Theory of Mind）と呼ばれる枠組みが提唱されている．心の理論とは他者が自身とは異なる心的状態を持っていることを理解し，推測できる能力のことを指し，他者の行動を理解するために必要な能力である．心の理論に関連する研究は多岐にわたるが，近年は協調エージェントの設計という観点から実装的な研究が行われている．例えばベイジアン ToM として，ベイズ理論を用いて他者の信念や欲求を推定するモデルが提案されており，ベイズ理論による他者の願望推定が人間が行う願望推定の結果と類似しているとして提案されている [5][6]．他にも逆強化学習による心の理論や心の理論を用いた AI の設計など，心の理論を用いた協調エージェントの設計に関する研究が行わ

*連絡先：静岡大学情報学部
〒 432-8011　静岡県浜松市中央区城北 3-5-1
E-mail: yamaguchi.takumi.20@shizuoka.ac.jp

れている [7][8]．しかしながら現行の研究では，推論された他者の内部状態を用いた行動設計までの流れが明確に示されていない．また行動決定まで実行することのできるモデルの多くは機械学習ベースの手法を多く用いており，その内部の推論過程がブラックボックスとして扱われていることが多く，透明性や説明可能性という点で問題がある [9]．

そこで本稿では自由エネルギー原理とそれに基づく能動的推論を用いて，他者の願望推定と共同行動の発生を数理的にモデル化することに挑戦する [10][11]．自由エネルギー原理は生物が環境を認識し，行動を選択するための理論であり，ベイズ脳仮説に基づいて人間が環境を認識するための理論である．このモデルは推論と行動を同一の理論の上で扱えるため，行動設計までの流れを明確に行うことができるだけでなく，数理的に記述されており，前述した問題である行動決定の透明性や説明可能性に対して有効であると考えられる．共同行動の発生をモデル化するために，他者の内部状態をエージェントが内部でシミュレートし，その結果をもとに他者の行動計画を推定し，最終的に自己の行動計画を立てるモデルを提案する．このモデルにおける各種推論行為は自由エネルギー原理に基づいて行う．提案したモデルの妥当性を検証するためにシミュレーション実験を行い，提案したモデルが共同行動の発生に必要な情報処理を行うことができることを示す．

2　自由エネルギー原理

まずはじめに自由エネルギー原理について，その概要を説明する．先に以下で使用する記号について表 1 に示す．

2.1　隠れ状態の推定

自由エネルギー原理ではベイズ脳仮説に基づいて，人間がベイズ推論（あるいはその近似手法）を用いて環境を認識するという仮説である．例えば生物が観測値として o を獲得し，その観測値の原因となる隠れ状態 s を推論したいとする．このときベイズの法則に従えば，ある観測値が得られたときの隠れ状態の確率である $P(s|o)$ は以下の式 1 で表すことができる．

$$P(s|o) = \frac{P(o,s)}{P(o)} \tag{1}$$

このとき完全なベイズ推論を行うためには $P(o)$ が既知である必要がある．しかしながら実世界においてこのような状況はほとんど存在せず、様々な近似手法が考えられている。その中で変分ベイズ推論と呼ばれるものがある．変分ベイズ推論では事後確率を直接導出するのではなく，隠れ状態に関する任意の確率分布（例えば $Q(s)$）を用意し，これと事後確率分布（$P(s|o)$）の確率分布の類似度を近づけるように任意の確率分布を変化させることでベイズ推論を近似しようとする手法である．実際には以下の式 2 で表される変分自由エネルギーという指標を小さくするようなパラメータ θ を求める手法である．

$$\begin{aligned} \theta^* &= \underset{\theta}{\operatorname{argmin}} \mathcal{F} \\ \mathcal{F} &= D_{KL}[Q(s;\theta)||P(s|o)] - \ln P(o) \\ &= \mathbb{E}_Q[\ln Q(s;\theta) - \ln P(s,o)] \end{aligned} \tag{2}$$

このとき，エージェントとそれを取り巻く環境状態の関係性と推論を図 1 の青く塗りつぶされている部分に表している．エージェントが持つ環境状態に関する認識を $Q(s;\theta)$ として変分ベイズ推論の手法をエージェントに適応している．これはエージェントが自身の持つ環境に関する仮説（内部状態 $Q(s;\theta)$）と知覚から推測した環境状態の推論（尤度 $P(s|o)$）を比較し，これを最小化するように内部状態を変化させることで環境状態を推論することを示している．

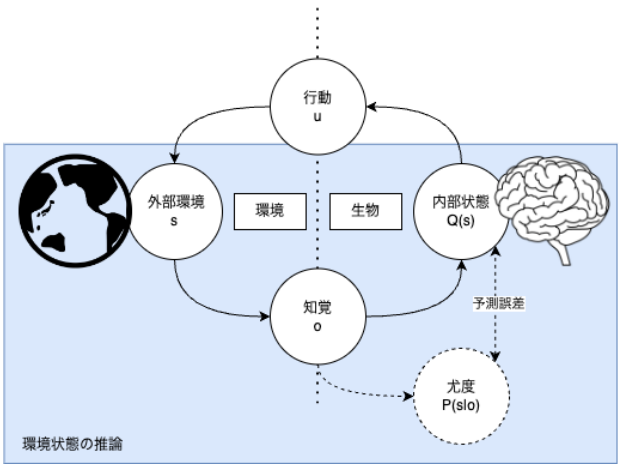

図 1: 自由エネルギー原理における状態推論の概要図．エージェントは自身の持つ環境に関する仮説（内部状態）と知覚から推測した環境状態の推論（尤度）を比較し，これを最小化するように内部状態を変化させる．

2.2　行動の決定（能動的推論）

自由エネルギー原理に基づく行動の制御理論を能動的推論と呼ぶ．この理論では生物が前述の変分自由エネルギーを最小化するように行動を計画するとする理論である．ここでエージェントが変分自由エネルギーを最小化するためには，環境状態の変化をエージェン

表 1: 使用する記号の説明

記号	説明	備考
s	環境状態	
o	観測値	
u	行動（制御状態）	
π	行動方策	$\pi = (u_0, u_1, ...)$
$\hat{o}$	選好状態	
$Q(s)$	環境状態に関する認識	
$\hat{P}(o)$	選好状態分布	選好状態の観測値に関する分布
$\mathcal{F}$	変分自由エネルギー	予測誤差
$\mathbf{G}$	期待自由エネルギー	選好状態との差
$\mathbb{E}$	期待値	
D_{KL}	カルバック・ライブラー情報量	2 つの確率分布間の類似度を測る

トが予測し，その後の状態での自由エネルギーが最小のものを選択することが良い．

変化後の変分自由エネルギーの予測値を期待自由エネルギー（$\mathbf{G}$）と呼び，以下の式 3 で表される．

$$\begin{aligned}&\mathbf{G}(o_{1:T}, s_{1:T}, \pi)\\&= \mathbb{E}_Q[\ln Q(s_{1:T}, \pi) - \ln \hat{P}(o_{1:T}, s_{1:T}, \pi)]\end{aligned} \tag{3}$$

このとき π はエージェントの行動方策と呼ばれ，エージェントが採択する行動の組み合わせを表している．すなわち式 3 はエージェントが特定の行動方策 π を採択したとき，その後の状態における自身の仮説と観測値から予測される環境状態の差がどの程度になるのかという指標と言える．特にある時点 τ において、エージェントが特定の行動方策 π を採択したときの期待自由エネルギーは以下の式 4 で表される．

$$\begin{aligned}&\mathbf{G}_\tau(\pi)\\&= \mathbb{E}_{Q(o_\tau, s_\tau|\pi)}[\ln Q(s_\tau|\pi) - \ln \hat{P}(o_\tau, s_\tau|\pi)]\end{aligned} \tag{4}$$

ここで記述されている $\hat{P}(o_\tau)$ はエージェントにとっての選好状態を表しており、エージェントが特定の観測値を得ることに対してどのような選好を持っているかを表している．これによって期待自由エネルギーの計算においては（1）エージェントの環境状態の予測を正当化することのできる観測値を得ることと（2）エージェントの選好状態を満たすことの双方を考慮することが可能となり，エージェントにとっての報酬最大化と環境状態の不確実性の最小化を同時に達成することが可能な方策を決定する指標となる．この選好状態はエージェントが持つ内部状態であり、エージェントが持つ環境状態に関する認識とは異なるものである．

以上のようにして算出された期待自由エネルギーを用いて、行動方策の選択を行う．以下の式 5 で表されるように、エージェントは期待自由エネルギーが最小となるような行動方策を選択することで、エージェントにとっての選好状態を達成しつつ，自由エネルギーを最小化するような行動を選択する．

$$P(\pi) = \sigma(\mathbf{G}) \tag{5}$$

2.3 pymdp の概要

pymdp は Python で実装された自由エネルギー原理に基づく能動的推論を行うためのライブラリである [12]．このライブラリでは部分観測マルコフ決定過程（POMDP）を用いて，エージェントが環境とやり取りする際の行動選択を行う．本ライブラリは POMDP 生成モデルをユーザが仮定することで能動的推論に基づく生物の行動選択をシミュレートすることが可能である．生成モデルの構築には以下の 3 つのモデルが必要となる．

1. 観測尤度関数：$P(o|s)$
2. 状態遷移モデル：$P(s_{t+1}|s_t, a_t)$
3. 選好分布：$P(\hat{o})$

1 の観測尤度関数は特定の環境状態 s_t においてエージェントが観測 o_t を得る確率を表す．2 の状態遷移モデルは特定の環境状態 s_t において，ある制御状態（エージェントの採択した行動）u_t を取ったとき，次の環境状態 s_{t+1} に遷移する確率を表す．3 の選好分布はエージェントの持つ観測値の事前分布である．これは環境中に含まれるエージェントがその環境中で生存するためには自身にとって都合の良い状態（選好状態）を維持する必要があるためである．

3 他者の行動予測のモデル

人間は他者の願望を，行動や言動といった限られた情報から推論することが可能である．心の理論におけ

るシミュレーション理論を用いることで，2.3 節で述べた生成モデルを用いて他者の願望を推定することが可能であることを示す．ここで心の理論におけるシミュレーション理論とは他者の行動を推論する際に，自身が同じ状況に置かれた場合にどのような行動を取るかを想定することで他者の行動を推論する理論である [13]．これはメンタライジングプロセスとも呼ばれ，他者の行動を推論するための一つの手法として提案されている [14]．既存の研究では他者の内部状態の推定までを扱うことのできる理論は提案されていたものの，行動の予測は他の領域に含まれてしまっていた．自由エネルギー原理を用いることによって能動的推論の枠組みを使用することが可能となり，統一の理論の上で行動の予測までを取り扱うことができるようになる．これが本研究の特徴である．これらの理論をもと他者の願望推定は以下の手順で行うことができると考える．

1. 他者の観測値の推論：自身の持つ隠れ状態の事後分布を用いて他者の観測値の予測を行う．
2. 他者の信念の推論：他者の観測値の推論を行った結果を用いて他者の信念の推論を行う．
3. 他者の行動の推論：他者の信念の推論を行った結果を用いて他者の行動の推論を行う．
4. 他者の願望の推論：他者の行動の推論を行った結果を用いて他者の願望の推論を行う．

2.2 節で示したように，能動的推論の枠組みではエージェントは期待自由エネルギーを最小化するような行動方策を選択することで自由エネルギーを最小化するような行動を選択する．期待自由エネルギーが算出することができるのは特定の信念の状態において，エージェントの選好状態が定義されている場合である．他者の行動を推論するためには，他者の信念の状態の推論結果と，他のエージェントの選好状態が必要である．

ここで提案するモデルでは，他者の選好状態のモデルを複数用意し，推論した他者の信念の状態に対してそれぞれの選好状態を用いて期待自由エネルギーを算出することで，他者の選好状態を隠れ状態とする尤度関数（以降では行動の尤もらしさを表す確率分布であることから行動尤度関数と呼称する,）を求める．これは例えば他者が協力しているならこのような行動をとるし，競争しているならこのような行動をとるといったように，他者の選好状態に応じた行動予測をすることと等しいと考える．この計算により，他者の行動を観測値，願望を隠れ状態としてベイズの枠組み（すなわち自由エネルギー原理）で他者の行動を推論することが可能となる．

上述した手順を図式化したものを図 2 に示す．以降の節では $\hat{x}$ を x に関する他者の推定値を表すとする．

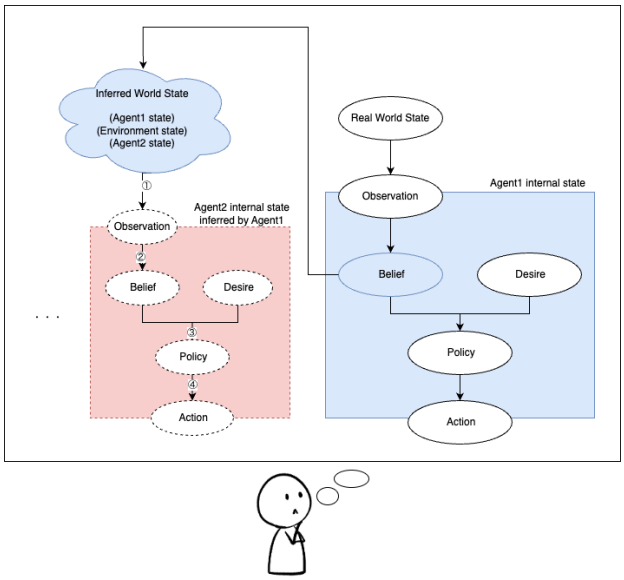

図 2: 提案モデルの概要．図中の①から④は他者の願望推定の手順として述べた各手順の番号と対応づけられている．

3.1 隠れ状態の再定義

他者の信念を推定するためには他者の観測値が必要であるが，これは直接手に入れることはできない．そこで自身が推論した隠れ状態の分布を用いて再生成すると考える．このとき元の隠れ状態集合が以下のように定義されていたとする．

$$\mathbf{s} = \{s^{myself}, s^{environment}, s^{other}\}$$

これは 1 節で述べた他者を含む環境における不確実な状態をそれぞれの次元で表現している．

2.3 節で述べた生成モデルを用いて，自身の信念から他者の観測値を予測するためには自身に関する状態と他者に関する状態を入れ替える必要がある．このときの隠れ状態の集合は以下のように再定義される．

$$\hat{\mathbf{s}} = \{s^{other}, s^{environment}, s^{myself}\}$$

これを用いて次節の他者の観測値の予測が可能となる．

3.2 他者の観測値の予測

3.1 節で述べたように隠れ状態集合を再定義することで，観測尤度関数を用いて他者の観測値の予測を行うことが可能となる．このときの観測値の確率分布は以下のように表される．

$$P(\hat{\mathbf{o}}_\tau) = \sum_{\hat{\mathbf{s}}} \mathbb{E}_{Q(\hat{\mathbf{s}}_\tau)} \left[P(\mathbf{o}_\tau | \hat{\mathbf{s}}_\tau) \right] \tag{6}$$

ここで算出した観測値の確率分布を用いて，他者の信念の推定を行う．

3.3 他者の信念の予測

一般的な環境における自由エネルギーの最小化のためには同時に発生しうるある一つの観測値の組み合わせから隠れ状態を推論する．しかしながら自身の信念から観測値を予測する場合，特定の観測値は存在せず，複数の観測値についてそれぞれ確率が割り当てられることになる．この確率を重みとして各観測値から求められるパラメータの総和を取ることで，他者の信念を推論することができると考える．推論結果の確率分布のパラメータ θ^* 以下のように表される．

$$\theta^* = \sum_{\hat{\mathbf{o}}} P(\hat{\mathbf{o}}) \underset{\theta}{\mathrm{argmin}} \mathcal{F}$$

$$\begin{aligned} \mathcal{F} &= D_{KL}[Q(\hat{\mathbf{s}};\theta)||P(\hat{\mathbf{s}}|\hat{\mathbf{o}})] - \ln P(\hat{\mathbf{o}}) \\ &= \mathbb{E}_Q[\ln Q(\hat{\mathbf{s}};\theta) - \ln P(\hat{\mathbf{s}},\hat{\mathbf{o}})] \end{aligned} \tag{7}$$

3.4 方策を予測

3.3 節で求めた他者の信念に対して，とり得るすべての選好状態に対して期待自由エネルギーを算出することで，各選好状態において発生させる行動の確率を求めることができる．ここで求めた確率は他者の信念を条件とした行動の確率であると言え願望を条件とした行動の確率分布 $P(Action|Desire)$ を表していると言える．本稿ではこれを行動尤度関数と呼び，隠れ状態を選好，観測値を行動としてベイズの枠組みで他者の選好を推論することができることを提案する．

まず他者の選好状態集合を $\mathbf{D} = (d_0, d_1, d_2, ...)$ と表す．ここで d_n は例えば協調型や敵対型といったような各選好状態分布を表している．式 4 で表される期待自由エネルギーについて 3.3 節で求めた他者の信念と d_n を用いて算出した期待自由エネルギーを $\hat{\mathbf{G}}_n$ とする．この $\hat{\mathbf{G}}_n$ によって，各選好状態において特定の行動方策を取る確率は $P(\pi|d_n) = \sigma(\hat{\mathbf{G}}_n)$ となる．各行動方策は $\pi = (u_0, u_1, u_2, ...)$ として各時間における行動を表しているため，各行動方策における第一の行動を取る確率を求めることで $P(Action|Desire)$ で表される行動尤度関数を求めることができる．これは以下の式 8 で表される．

$$P(u|d_n) = \sum_{\pi} P(u_0 = u|d_n) \tag{8}$$

3.5 他者の願望の推論

3.4 節で求めた行動尤度関数を用いて，他者の行動を観測値，願望を隠れ状態としてベイズの枠組みで他者の願望を推論することが可能となる．具体的には自由エネルギーと同様の枠組みとして，他者の行動を観測値，願望を隠れ状態として以下の式 9 で表される自由エネルギーを最小化するような願望を推論することが可能である．

$$\theta^* = \underset{\theta}{\mathrm{argmin}} \mathcal{F}$$

$$\begin{aligned} \mathcal{F} &= D_{KL}[Q(d;\theta)||P(d|u)] - \ln P(u) \\ &= \mathbb{E}_Q[\ln Q(d;\theta) - \ln P(d,u)] \end{aligned} \tag{9}$$

3.6 環境状態の変化の予測

3.5 節で求めた他者の願望を用いて，他者に関する環境状態の変化を予測することが可能となる．これは 3.2 節で述べた他者の観測値の予測と同様の手法で行うことが可能である．このときの行動の確率分布は以下のように表される．

$$P(u_\tau) = \sum_{d \in \mathbf{D}} \mathbb{E}_{Q(d)}\left[P(u_\tau|d)\right] \tag{10}$$

4 シミュレーション実験

本実験では 3 章で提案したモデルが他エージェントの願望推定，並びに共同行動の創発のモデルとして適当かどうかを検証する．そのために本稿では待ち合わせのシミュレーションを行い，エージェントが共同行動を行う際に他者の願望を推定し，それをもとに行動計画を立てることができるかを検証する．エージェントは 2 体存在し，それぞれが他者の行動を観測し，その行動をもとに他者の願望を推定し，それに応じた行動計画を立てることが可能であると仮定する．

また各エージェントは願望のタイプとして選好状態として協力と競争の 2 つを持つ可能性があることと仮定する．これによって自動的に共同行動が発生するような状況ではなく，相手エージェントの選好状態を予測することで自身に近づいてこようとしている（つまり協調しようとしている）エージェントなのか，離れて行こうとしている（つまり競争しようとしている）エージェントなのかを推定し，それに応じた戦略を取る必要がある．また本来図 2 に示したような推論は再帰的に定義することができるが，本稿では実装上の都合により他者が予測する自身の行動の予測までは行えていない．

4.1 実験環境

実験を行った機材は以下の表 2 の通りである．

4.2　シミュレーション環境

実際にシミュレーションを行った環境は以下の図 3 に示すような環境である．ここで灰色のマスは移動可能なマスを意味し，青色のマスは移動不可能なマスを意味する．このような壁となるマスを導入することでエージェントが壁越しに行われる他者の行動を予測し，協調するエージェントであれば同様の向きに動く，競争するエージェントであれば逆の向きに動くといったような行動を取ることができると考える．

また今回の実験ではエージェントの初期位置は固定されており，Agent1 は 3 行 1 列，Agent2 は 3 行 5 列に配置されている．各エージェントが正しく他者の選好状態と相手の位置を推定できた場合，2 エージェントは壁越しに片方が上方向に動いたらもう片方も上方向に動くといったように，相手に合わせて近づこうとする行動が観察されることが期待される．

4.3　エージェントの設定

エージェントには以下のような観測値が与えらえる．また両エージェントは期待自由エネルギーの算出の際には自身の行動方策を事前に任意の時間ステップ分作成し，全ての行動方策に対して期待自由エネルギーを算出する．今回の実験ではエージェントの行動方策の大きさは 6 としており，すなわち 6 回先の行動までを計画し，その行動方策の中で最も期待自由エネルギーが小さい行動を選択することになる．

4.3.1　観測値

エージェントに与えられる情報は以下の通りである．

- 自身の位置（正確な位置）
- 相手との距離（相手の位置に到達するまでに必要な移動回数）
- 相手の行動

表 2: 実験環境

ハードウェア構成		
OS	macOS Sequoia	15.1.1
CPU	Apple M3	
RAM	16GB	
ソフトウェア構成	**version**	
Python	3.11.10	
pymdp	0.0.7.1	

自身の位置と相手との距離によって環境状態を推定し，相手の行動によって選好状態を推定することが可能であると考える．

4.3.2　行動

エージェントは以下の行動を取ることができる．

- 上に移動
- 下に移動
- 右に移動
- 左に移動
- 待機

これらの行動を取ることでエージェントは相手エージェントとの距離を縮めたり離したりすることが可能である．

5　結果と考察

まず互いの心的状態に関する推論を一切行わない場合の結果を図 4 に示す．シミュレーション結果は上限である Step50 まで行ったが，最終的には待ち合わせタスクの成功を達成することができなかった．このときのエージェントは互いの行動に関する予測を一切行わない（つまり相手は常に同じ位置に留まり続ける）と考えるため，Step1 と Step2 の動作を繰り返すことになる．片方が相手に近づくように動作すれば，もう片方は遠ざかるように動作し，結果として互いにすれ違いを発生させることになる．

これと比較して，互いの心的状態に関する推論を行った場合の結果を図 5 に示す．この場合，待ち合わせタスクの成功を達成することができた．

Step1 から 10 辺りまでは互いにランダムに動作しているように見える．Step11 から 14 にかけて各エージェ

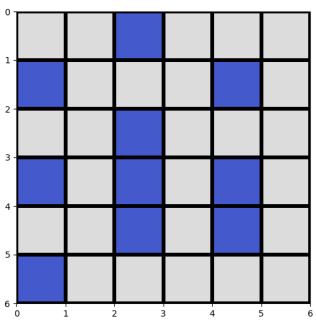

図 3: シミュレーション実験の環境

ントは相手エージェントに近づくように動作しており，Step15 でエージェントはすれ違いを発生させ，待ち合わせタスクの成功を一度逃している．その後の Step16 からは青色のエージェント（エージェント 2）が赤色のエージェント（エージェント 1）から遠ざかるように動作している．それに対してエージェント 1 はエージェント 2 を追いかけるような動作をしている．ここでなぜエージェント 2 がエージェント 1 から遠ざかるような行動をとったのかを考察するために，各 Step における各エージェントの内部状態を以下の図 6 から図 8 に示す．その結果 Step15 でのすれ違いが発生した際にエージェント 2 はエージェント 1 の位置を誤って推論しており，それに応じた結果がこの遠ざかる行動につながっていることがわかる．

また位置状態の推論だけでなく，他エージェントの願望推定の結果にも注目すると両エージェント共に相手の選好状態が敵対であると推定していることがわかる．この原因について考察する．原因の一つとして他エージェントの願望状態の推定において，他エージェントが予測する自身の行動を考慮していないことが考えられる．例えば Step1 から 4 に関して，エージェント 1 はエージェント 2 の移動方向を推定する際に重要となるのはエージェント 2 の推定するエージェント 1 の位置である．エージェント 2 がエージェント 1 が上に行くと予測している場合，エージェント 2 が協調的であるならばエージェント 2 は上に行くべきということを考えることができるが，これは無限に再帰する構造が必要となる．本稿で扱ったモデルではこのような再帰的な構造を再現はしておらず，結果として誤推定が生じていると考えられる．

6 議論

本稿で行った実験では他者の願望を推定し，それに応じた行動計画をとることで共同行動の創発につながると考え，その可能性を示そうとした．実験結果より，他者の心的状態を推定することで共同行動が必要なタスクを成功させることができるようになったことから，他者の願望推定と，それに伴う行動推定という心的状態の推論を行うことが共同行動の創発に寄与する可能性を示唆した．しかしながら現状のエージェントの推論結果を観察すると，その精度は低いと言える．その要因として，他者の選好状態を推論する際に相手エージェントが予測する自身の行動を考慮していないことが挙げられる．このような再帰的な構造の実装を行うことにより，推論の精度を高めることが課題として挙げられる．

また本実験の問題として，設計した実験条件の単調さが挙げられる．本稿においてエージェントが報酬を獲得できる場面は他者との待ち合わせが成功した場合のみであり，協調をしないことによる報酬の損失が存在しない．そのためエージェントの協調行動の発生は必然であると言える．協調行動の創発に関する研究においては，ゲーム理論の文脈で検討されている StugHunt ゲームのような環境を用いることが考えられる [15]．StugHunt ゲームとは，個人の選択が他者の行動に依存する状況をモデル化するゲームであり，各エージェントは「鹿を狩る（協力）」または「ウサギを狩る（単独行動）」のどちらかを選択する．両者が協力して鹿を狩れば高い報酬を得られるが，片方が協力しない場合，協力した側は何も得られない．一方で，ウサギを選べば確実に小さな報酬を得ることができる．したがって，協力は大きな利益をもたらすが，相手を信頼できなければリスク回避のために単独行動を行うという，協力と競争のトレードオフを考慮することができる．このような環境において，本稿で示したモデルによって他者の心的状態を推定することが協調行動を促進することができるかを検証することが今後の課題として挙げられる．

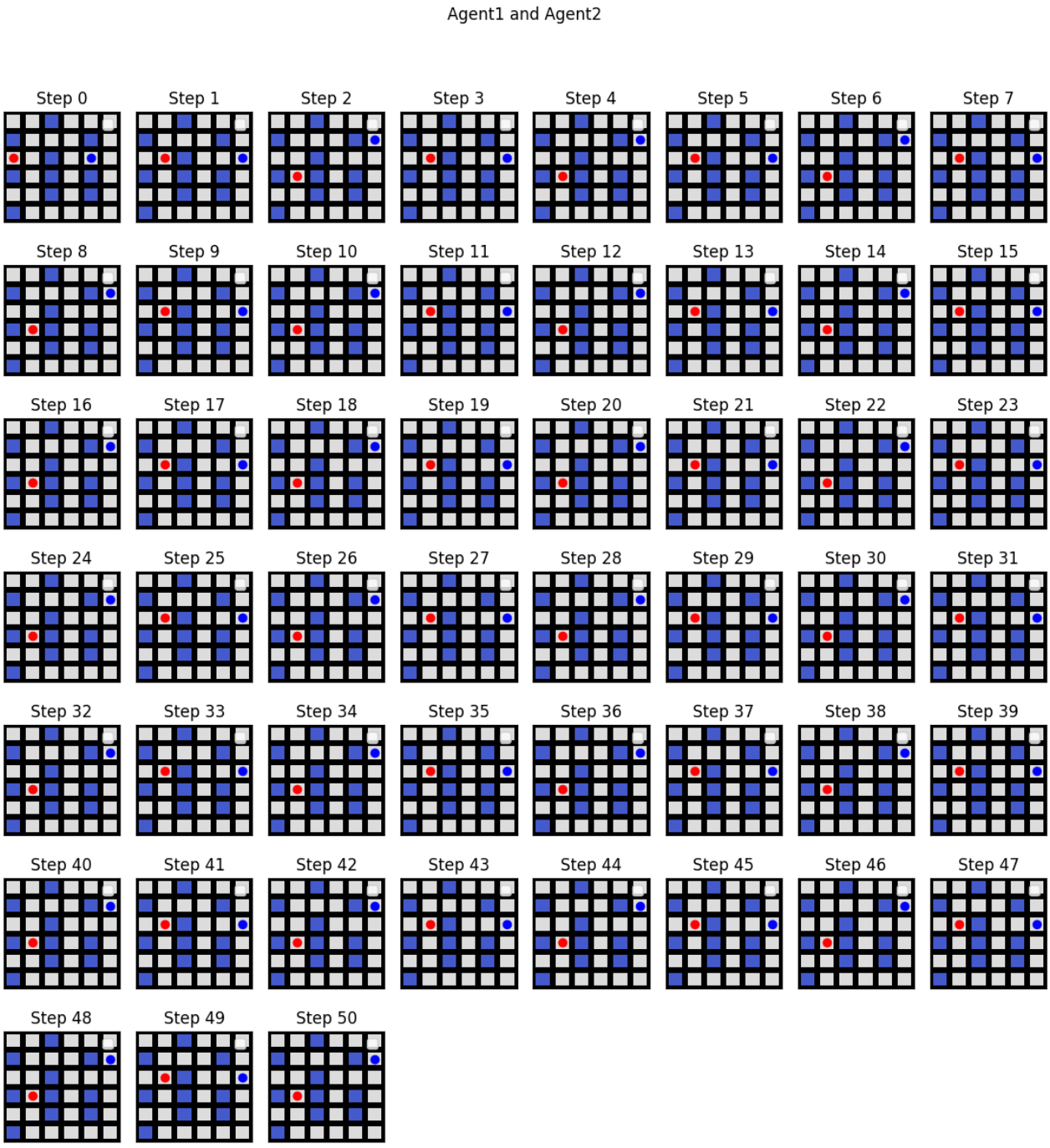

図 4: 他エージェントの心的状態を推論しないエージェント同士のシミュレーション結果

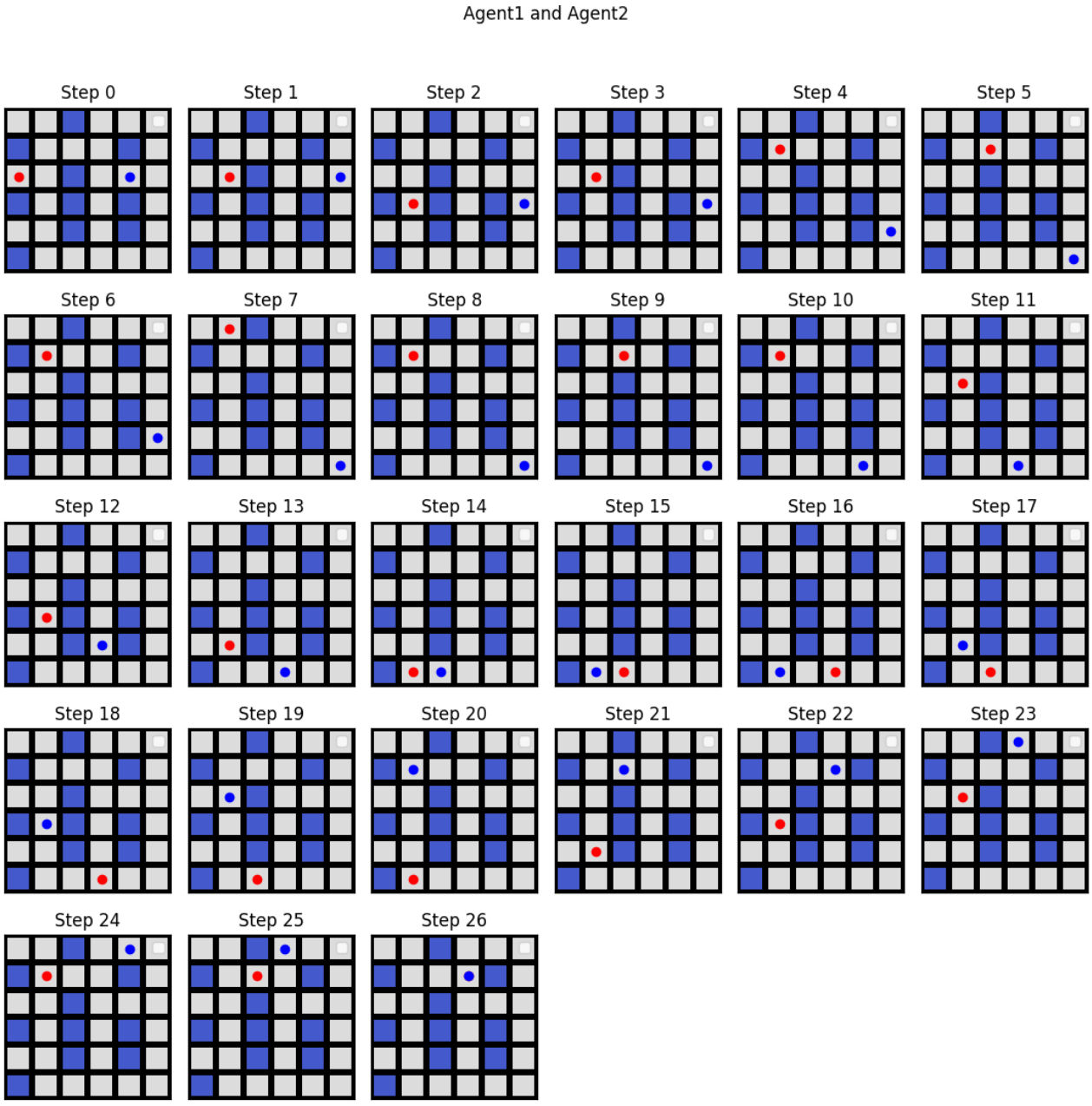

図 5: 協力するエージェント同士のシミュレーション結果

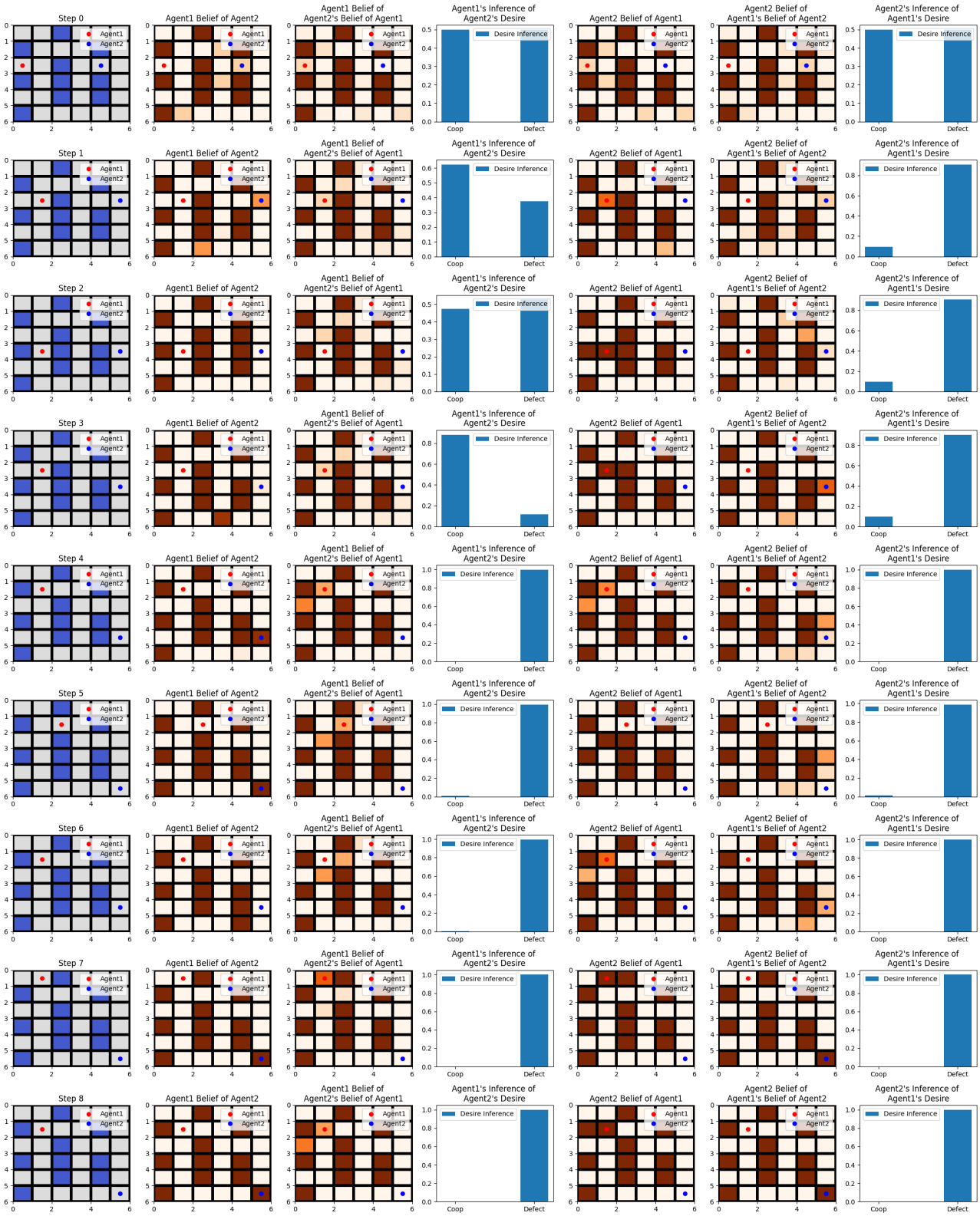

図 6: step0 から 8 までの各エージェントの内部状態

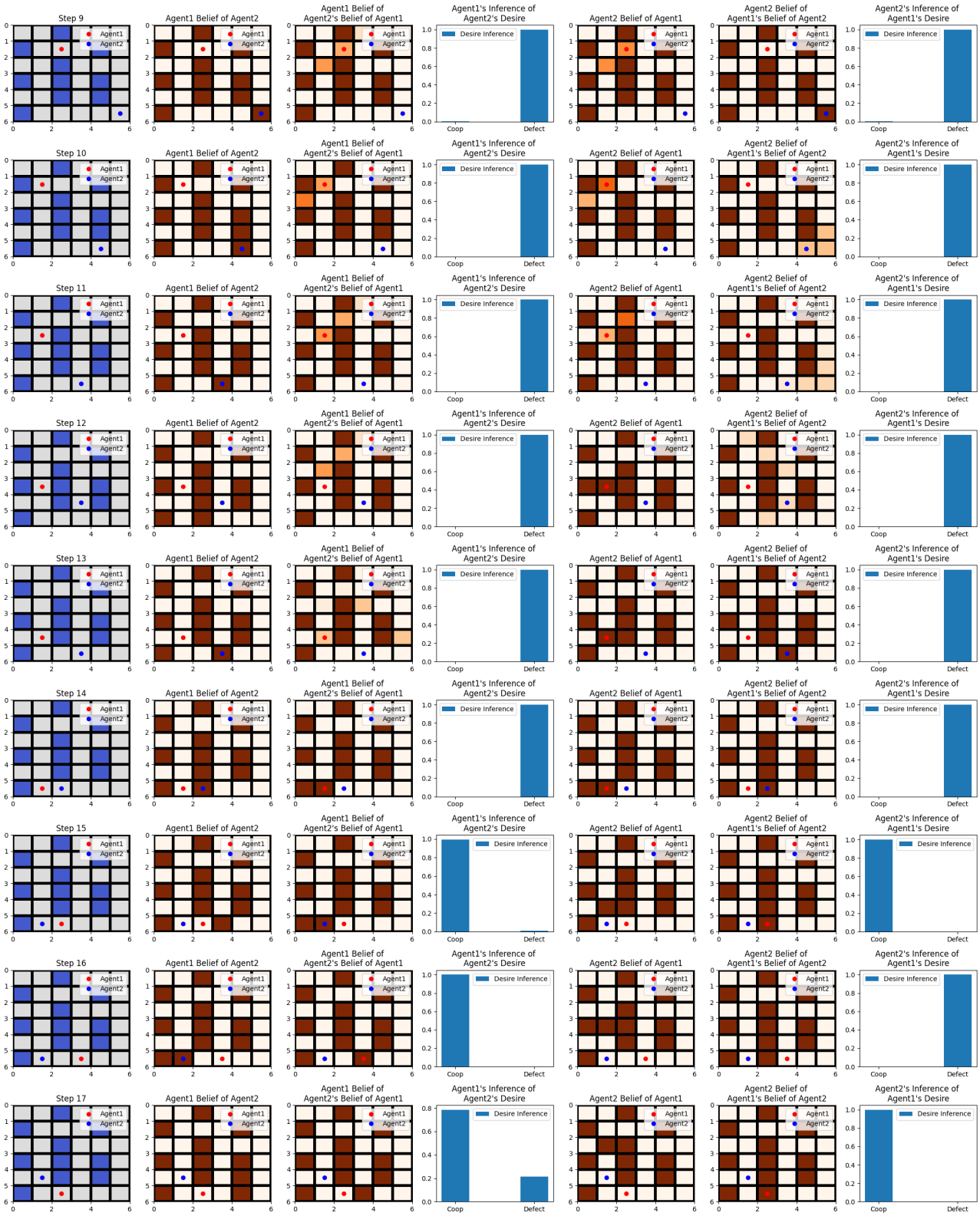

図 7: step9 から 17 までの各エージェントの内部状態

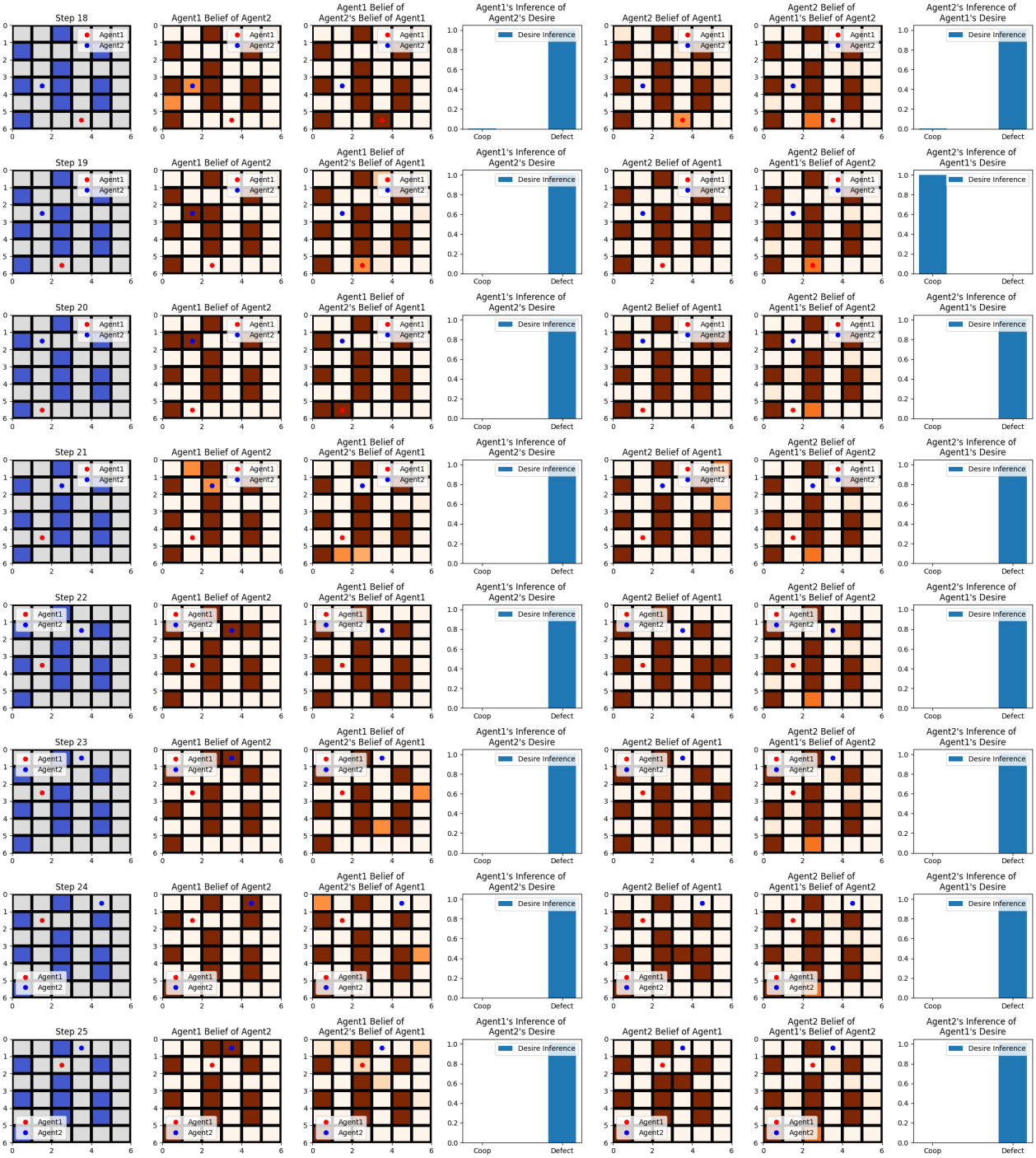

図 8: step18 から 25 までの各エージェントの内部状態

参考文献

[1] Sebanz, N., Bekkering, H. & Knoblich, G. (2006). Joint action: bodies and minds moving together. Trends in Cognitive Sciences, 10(2), 70-76.

[2] Sadigh, D., Sastry, S., Seshia, S. A. & Dragan, A. D. (2016). Planning for Autonomous Cars that Leverage Effects on Human Actions. Robotics: Science and Systems XII.

[3] Scassellati, B. (2002). Theory of Mind for a Humanoid Robot. Autonomous Robots, 12(1), 13-24.

[4] Allen-Williams, M. & Jennings, N. R. (2009). Computational Intelligence, Collaboration, Fusion and Emergence. Intelligent Systems Reference Library, 321-360.

[5] Stacy, S., Gong, S., Parab, A., Zhao, M., Jiang, K. & Gao, T. (2024). A Bayesian theory of mind approach to modeling cooperation and communication. Wiley Interdisciplinary Reviews: Computational Statistics, 16(1).

[6] Baker, C. L., Saxe, R. R. & Tenenbaum, J. B. (2011). Bayesian Theory of Mind: Modeling Joint Belief-Desire Attribution. Cognitive Science.

[7] Jara-Ettinger, J. (2019). Theory of mind as inverse reinforcement learning. Current Opinion in Behavioral Sciences, 29, 105-110.

[8] Cuzzolin, F., Morelli, A., C?rstea, B. & Sahakian, B. J. (2020). Knowing me, knowing you: theory of mind in AI. Psychological Medicine, 50(7), 1057-1061.

[9] Albrecht, S. V. & Stone, P. (2018). Autonomous agents modelling other agents: A comprehensive survey and open problems. Artificial Intelligence, 258, 66-95.

[10] Friston, K. (2010). The free-energy principle: a unified brain theory? Nature Reviews Neuroscience, 11(2), 127-138.

[11] Pezzulo, G., Rigoli, F. & Friston, K. (2015). Active Inference, homeostatic regulation and adaptive behavioural control. Progress in Neurobiology, 134, 17-35.

[12] Heins, C., Millidge, B., Demekas, D., Klein, B., Friston, K., Couzin, I. & Tschantz, A. (2022). pymdp: A Python library for active inference in discrete state spaces. arXiv.

[13] Apperly, I. A. (2008). Beyond Simulation-Theory and Theory-Theory: Why social cognitive neuroscience should use its own concepts to study " theory of mind. " Cognition, 107(1), 266-283.

[14] Kopp, S. & Kr?mer, N. (2021). Revisiting Human-Agent Communication: The Importance of Joint Co-construction and Understanding Mental States. Frontiers in Psychology, 12, 580955.

[15] Skyrms, B. (2003). The Stag Hunt. In The Stag Hunt and the Evolution of Social Structure (pp. 1-14). chapter, Cambridge: Cambridge University Press.