

相互作用から目的が立ち上がるとき

～ 生命と目的の記号論へ向けたロボティクスと計算モデルからの話題提供～



堀井 隆斗

大阪大学大学院 基礎工学研究科 システム創成専攻

大阪大学 先導的学際研究機構

東京大学 IRCN

日本記号学会第46回大会 第1セッション「生命と目的の記号論」

@二松学舎大学 九段キャンパス 2026/7/11

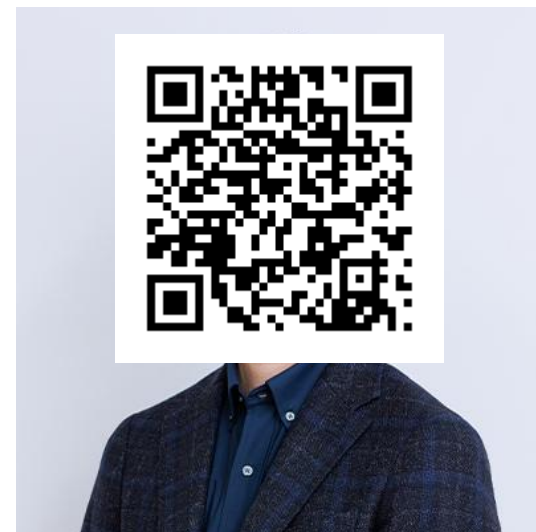
自己紹介

氏名：堀井 隆斗（ほりい たかと）

現在：大阪大学大学院基礎工学研究科 准教授

東京大学IRC� 連携研究者

KGモーターズ株式会社 研究開発部フェロー



略歴：

'11~'17 大阪大学大学院工学研究科, 博士前期・後期課程

'17~'19 電気通信大学, 特任研究員・特任助教

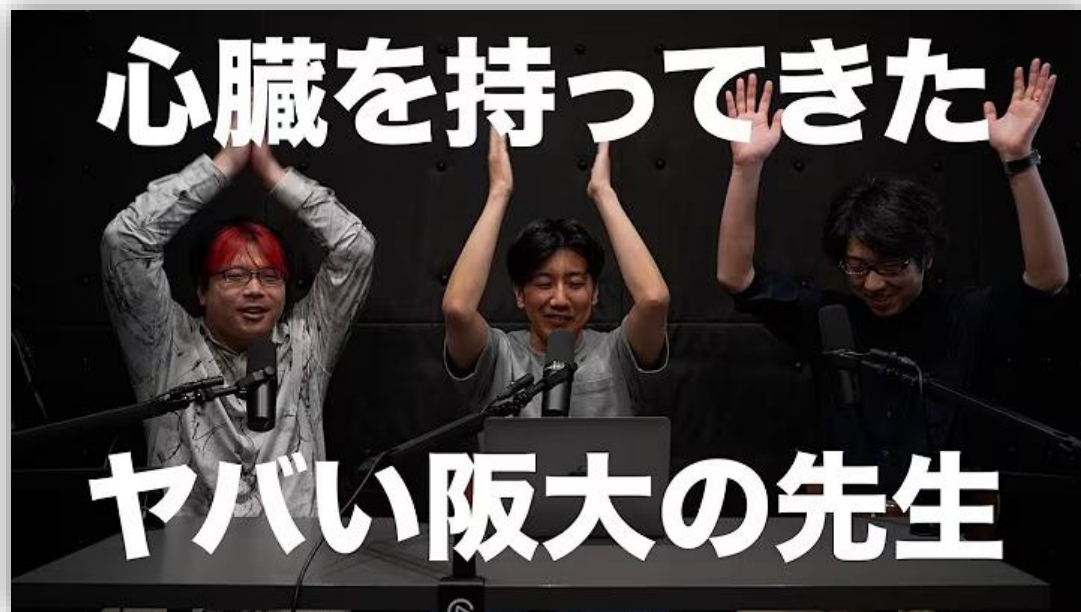
'19~'20 大阪大学基礎工学研究科, 助教

'20~ 大阪大学基礎工学研究科, 講師

'24~ 大阪大学基礎工学研究科, 准教授



YouTube動画みてね！



身体情報から創発した記号としての感情



堀井 隆斗

大阪大学大学院 基礎工学研究科 システム創成専攻 講師
大阪大学 先進的学際研究機構 兼任教員
東京大学 IRCN 協力研究員

第20回 Language and Robotics研究会 @Zoom 2023/12/15



YouTube動画みてね！



この話題提供で共有したい入り口

- ・「ともに生きる」とは、同じ記号を持つことではなく、互いの環世界をずらしながら調整し続けることではないか。



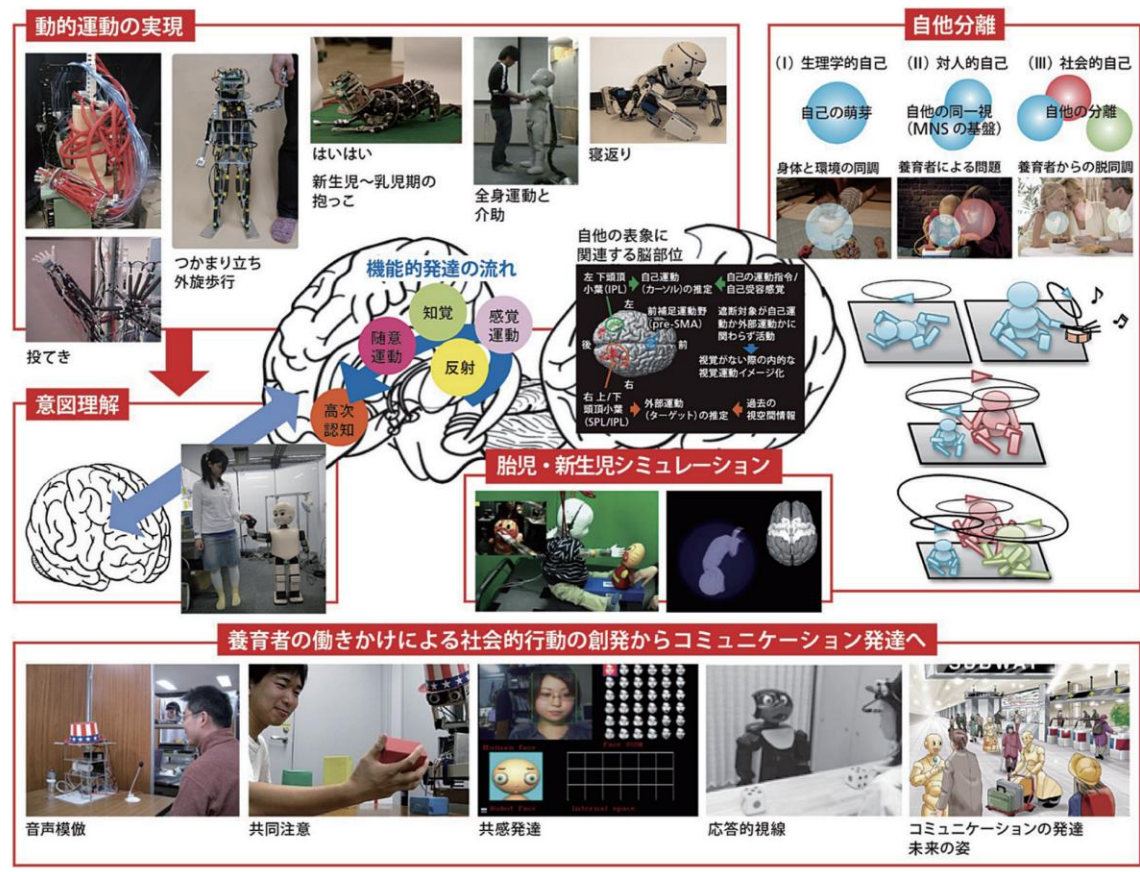
- ・記号は「外界を写すラベル」ではなく、身体・他者・環境のあいだで予測と行為を調整する媒体として立ち上がる。



構成論的ロボティクス：作って、壊して、見える化する

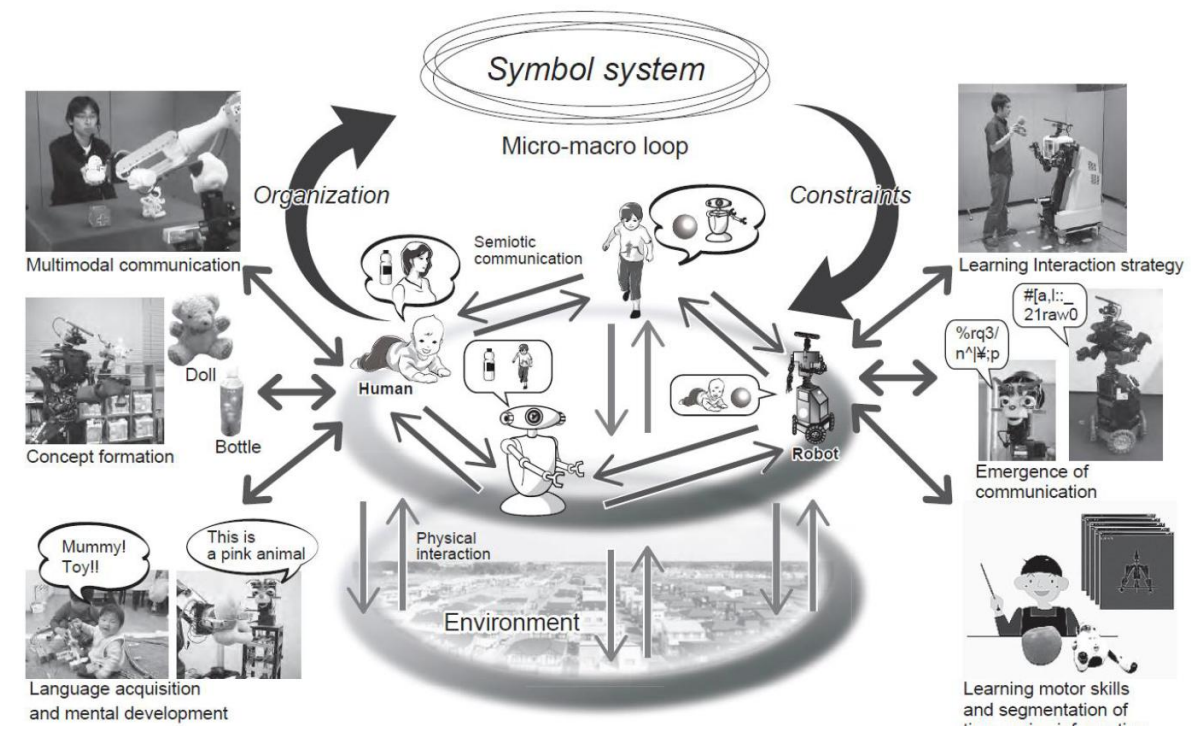
・ 認知発達ロボティクス [浅田, '10][Asada+, '09]

－ 発達科学とロボティクスの
相互理解を目指す学際的取り組み



・ 記号創発ロボティクス [Taniguchi+, '16, 19]

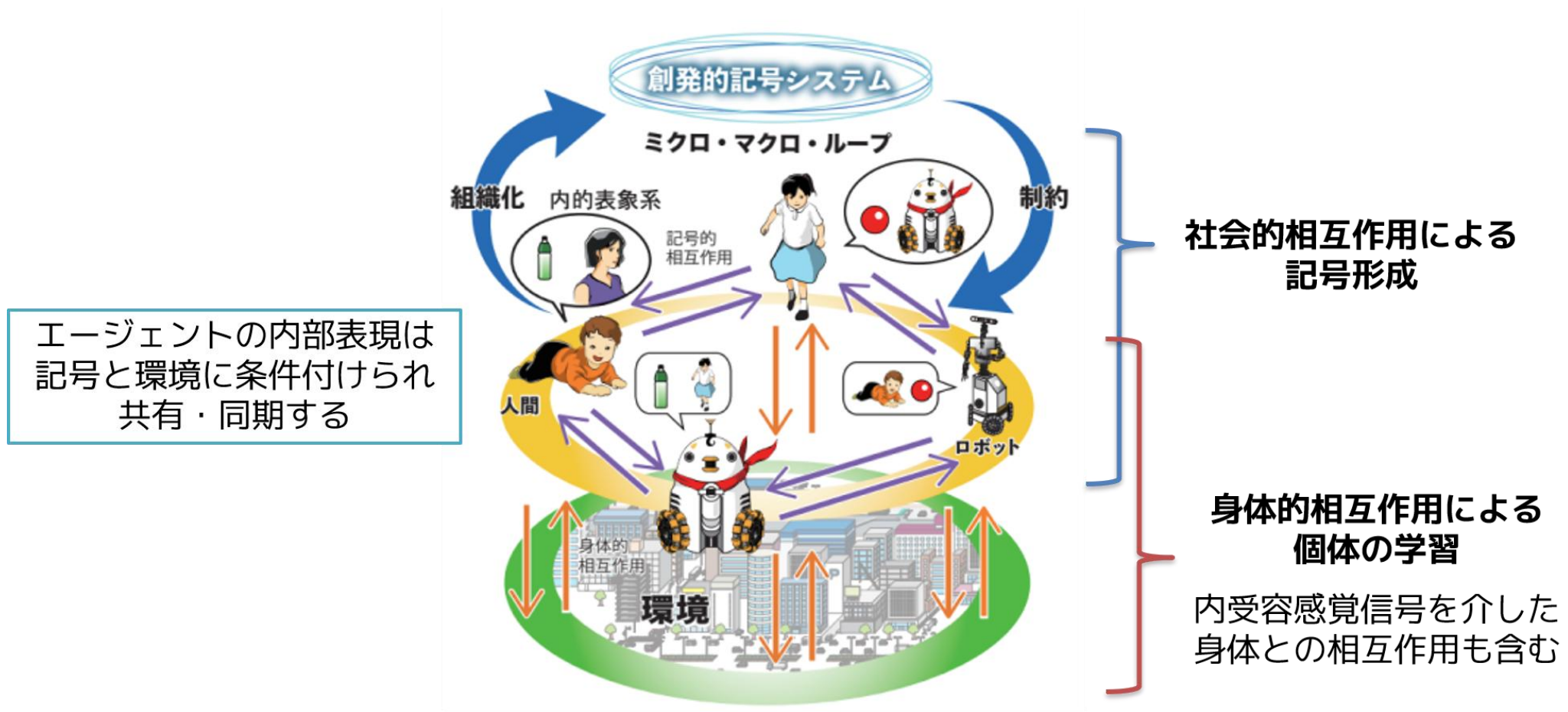
－ センサ・行為・言語の循環から内部表象と社会的
記号システムが同時に形成される過程を扱う



CPC：意味は「集合的な予測符号化」として生まれる

- **CPC（Collective Predictive Coding）：集合的予測符号化仮説** [Taniguchi+, 24]

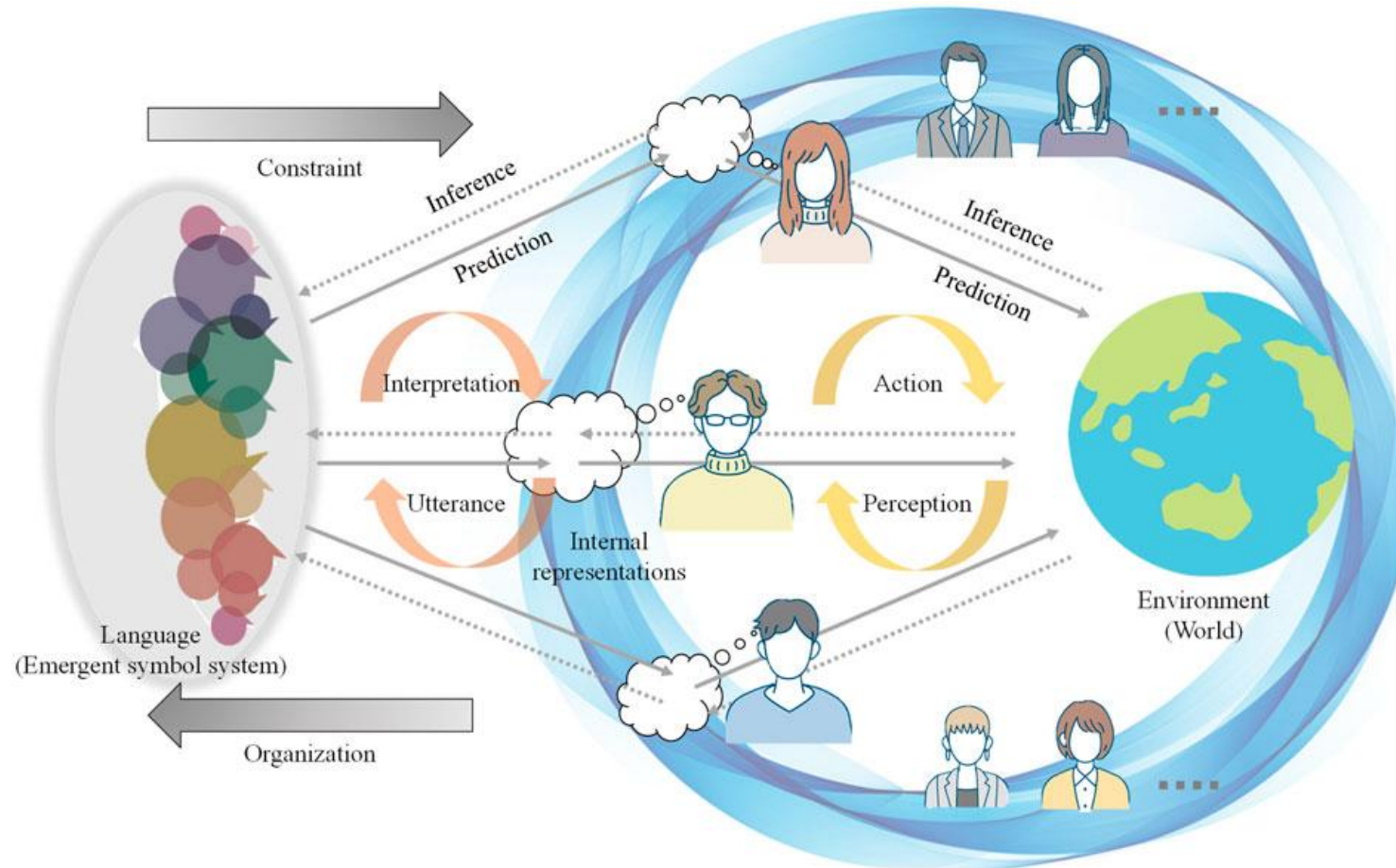
- 個体の中で閉じた推論だけではなく 「他者との符号化のループ」が内部表象を変える



CPC：意味は「集合的な予測符号化」として生まれる

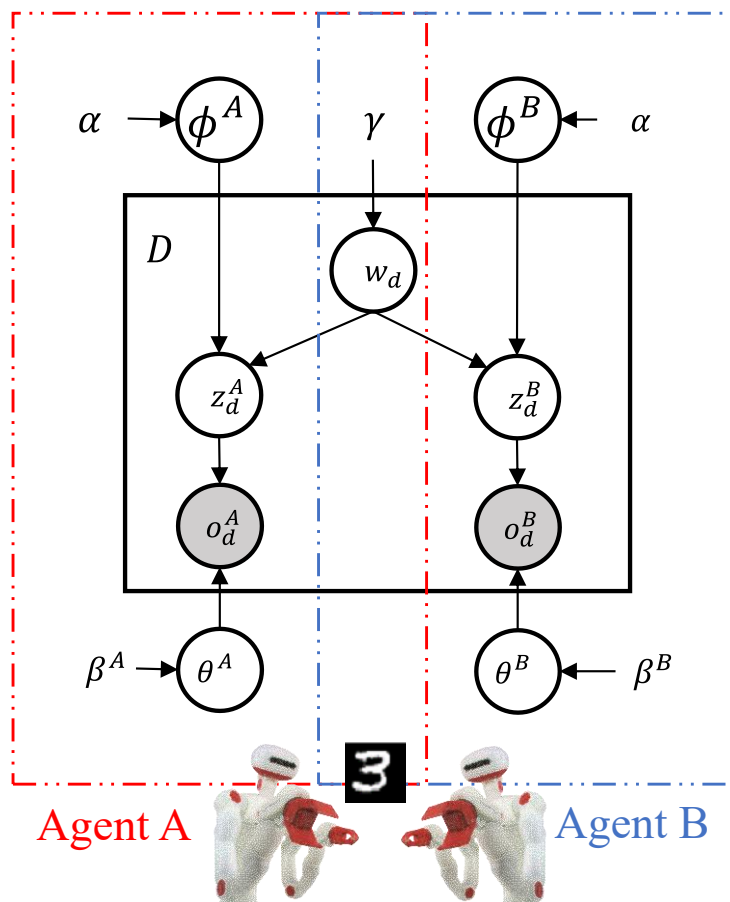
- **CPC（Collective Predictive Coding）：集合的予測符号化仮説** [Taniguchi+, 24]

- 個体の中で閉じた推論だけではなく 「他者との符号化のループ」が内部表象を変える



CPCの計算モデルとしての名付けゲーム

- 記号のやり取りによって他者と共有可能な多感覚カテゴリを形成 [Taniguchi+, 23]



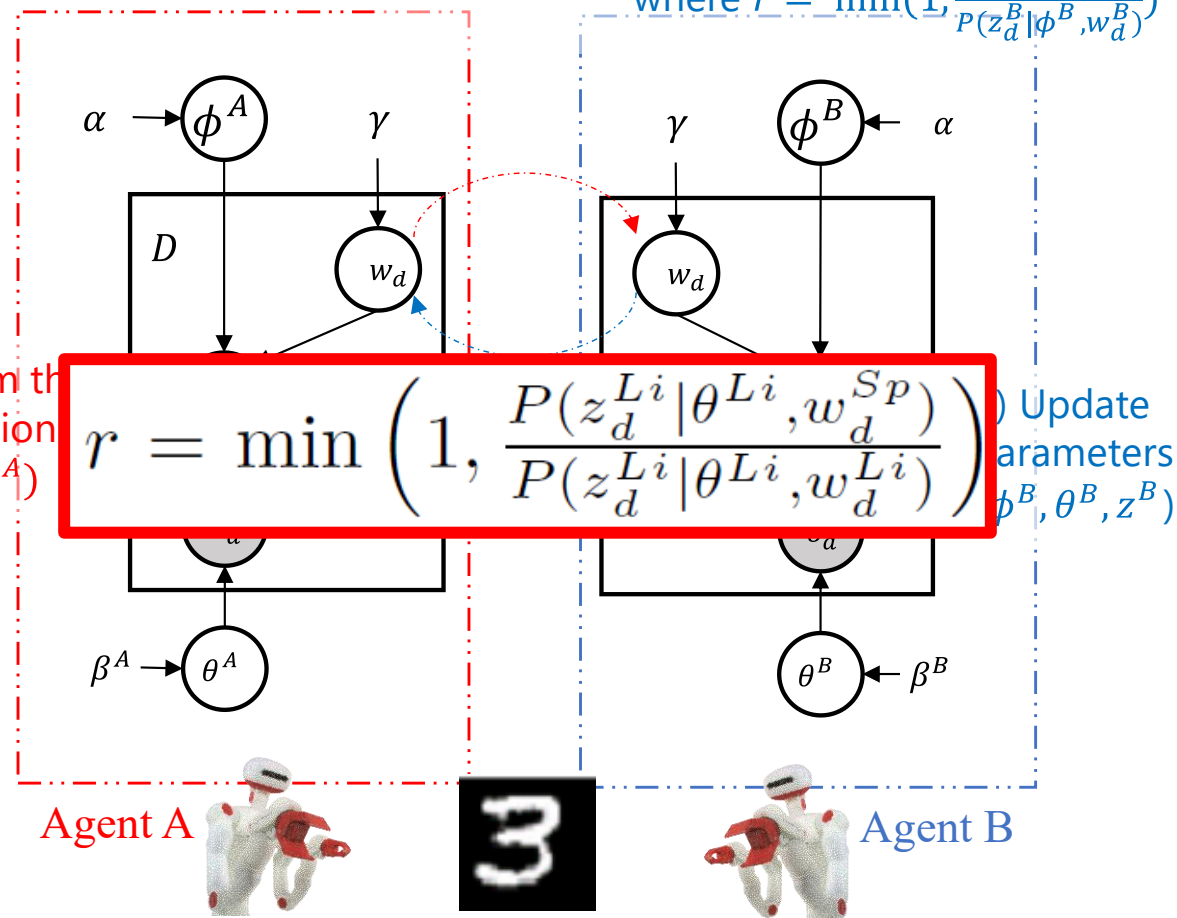
Decomposition

(1) Sample z_d^A from the posterior distribution $P(z_d^A | o_d^A, w_d^A, \theta^A, \phi^A)$

Composition

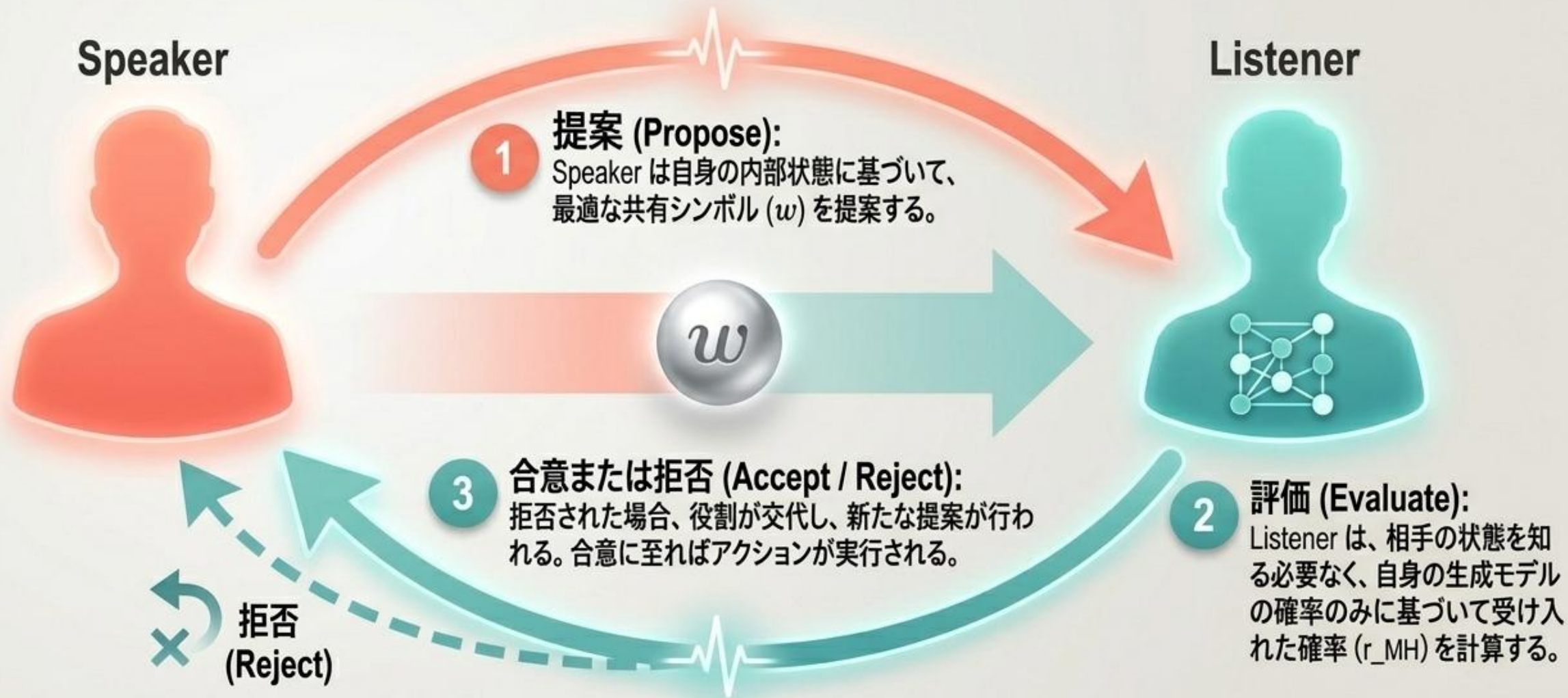
(2) Sample w_d^A from a proposal distribution $P(w_d^A | z_d^A, \phi^A)$

(3) Judge w_d^A by an acceptance rate r , where $r = \min(1, \frac{P(z_d^B | \phi^B, w_d^A)}{P(z_d^B | \phi^B, w_d^B)})$



Update parameters ϕ^B, θ^B, z^B

「ネーミングゲーム」はいかにしてギャップを埋めるのか？

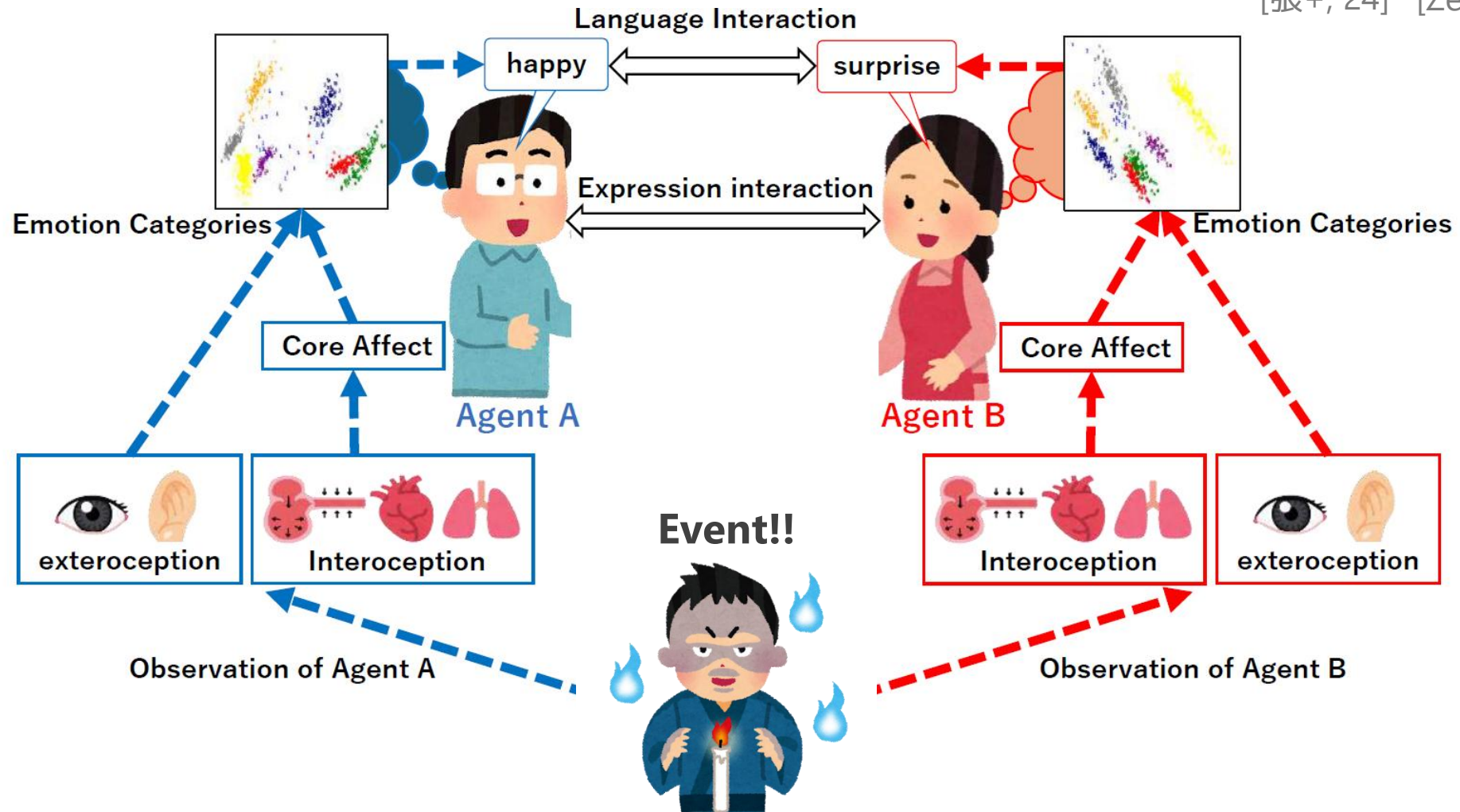


このプロセスにより、相手の心を完全に理解（推論）しなくても、互いの信念を制約し合うことが可能になります。

コミュニケーションによる物体概念モデルを素朴に感情に拡張する

- 感情喚起事象の体験共有とシンボル交換を通じて感情概念構造が共有されるか？

[張+, 24] [Zehang+, 2026]



母子間相互作用：最初期のSocial Reality

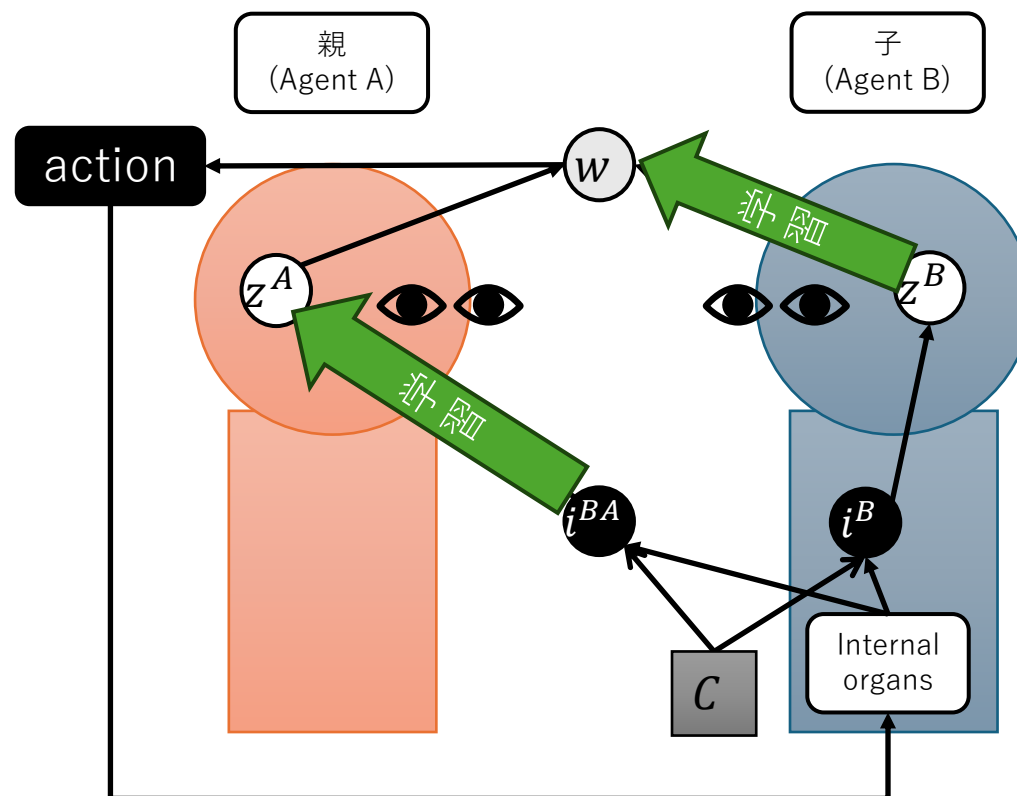
[鰐本+, 2025] [Tsubamoto and Horii, 2026]

・異なる環世界を持つ非対称な二者間の相互作用

－ 非対称な二者は記号を介して互いの「足りない半分」を補う

親：
・ 内臓状態の管理方法は既知
(ex. ご飯を食べればお腹がいっぱいに)

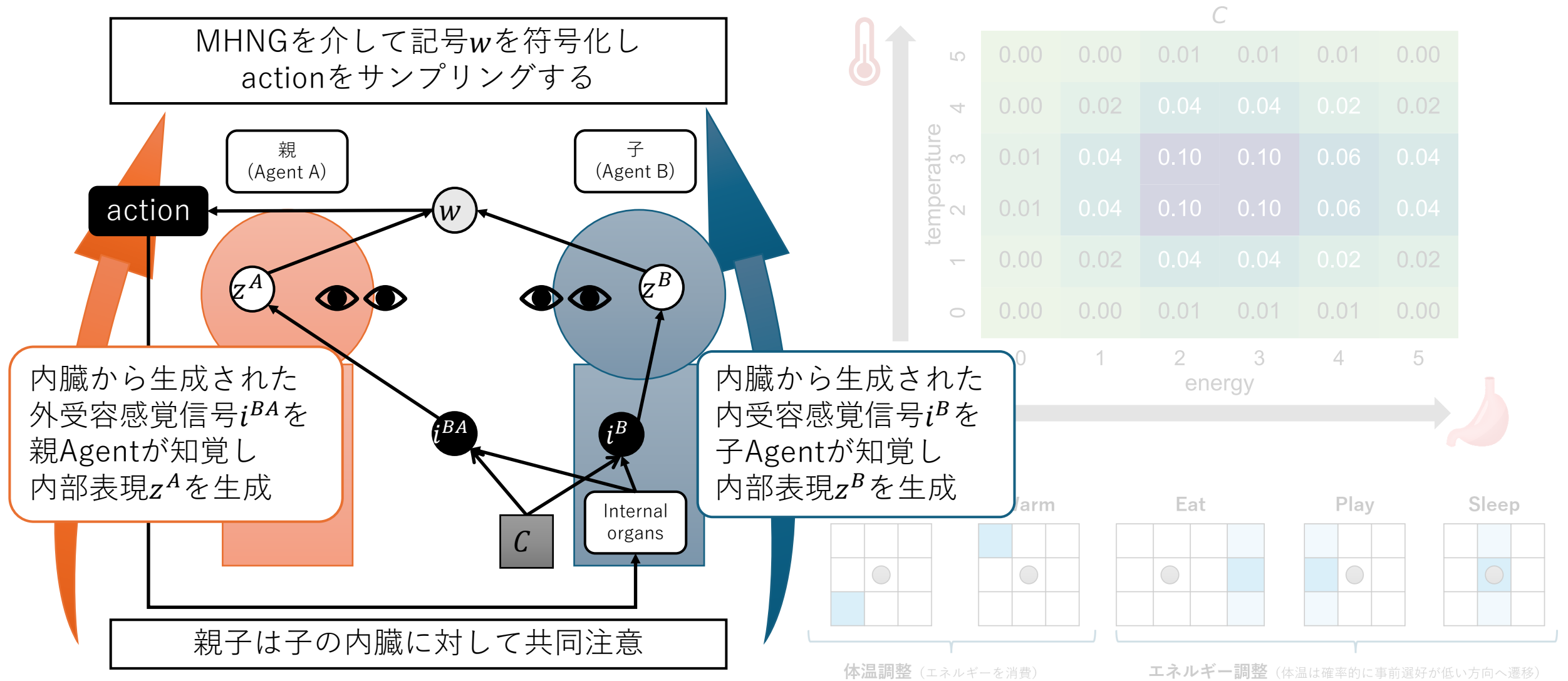
・ 子の内臓状態はわからない
(ex. お腹が空いているか?)



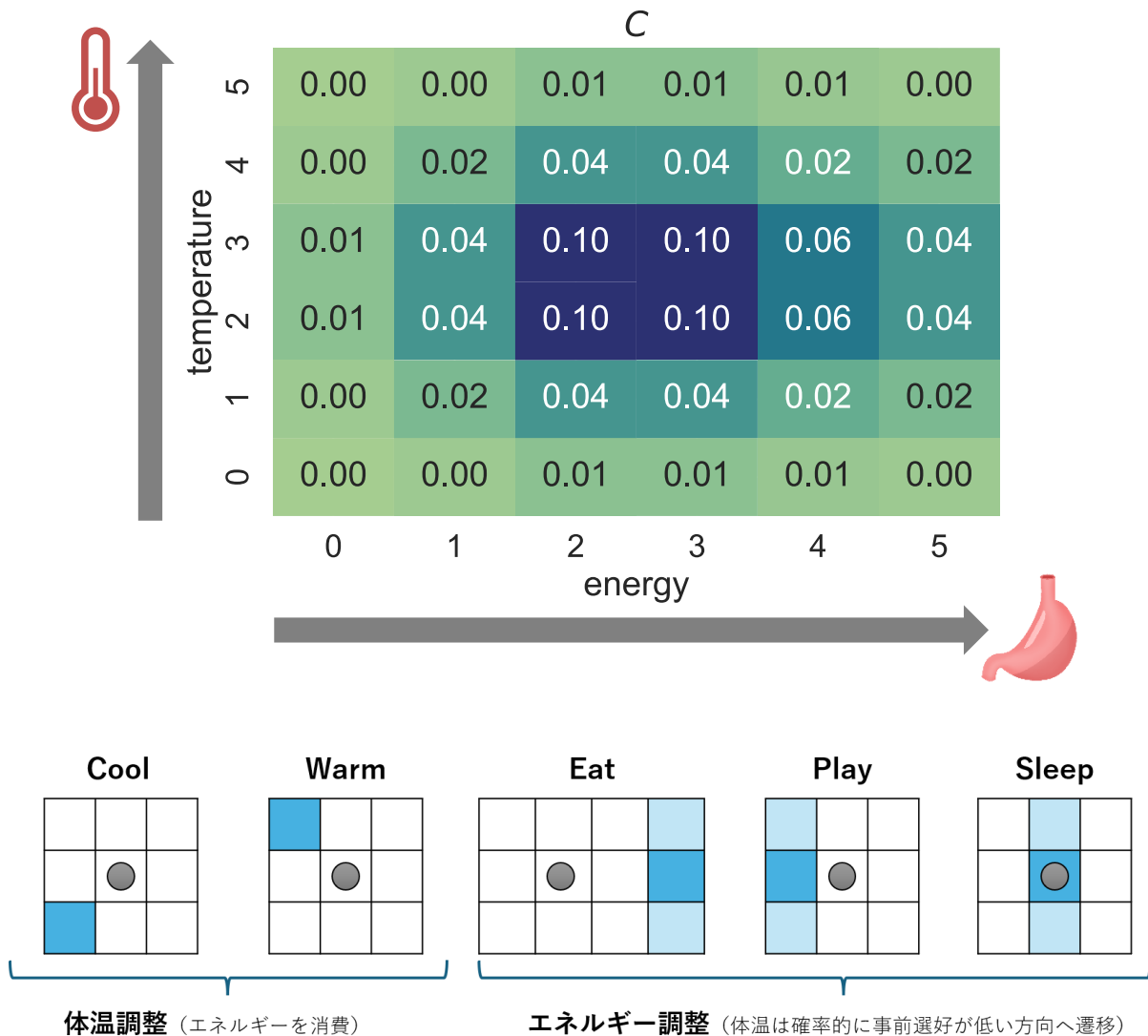
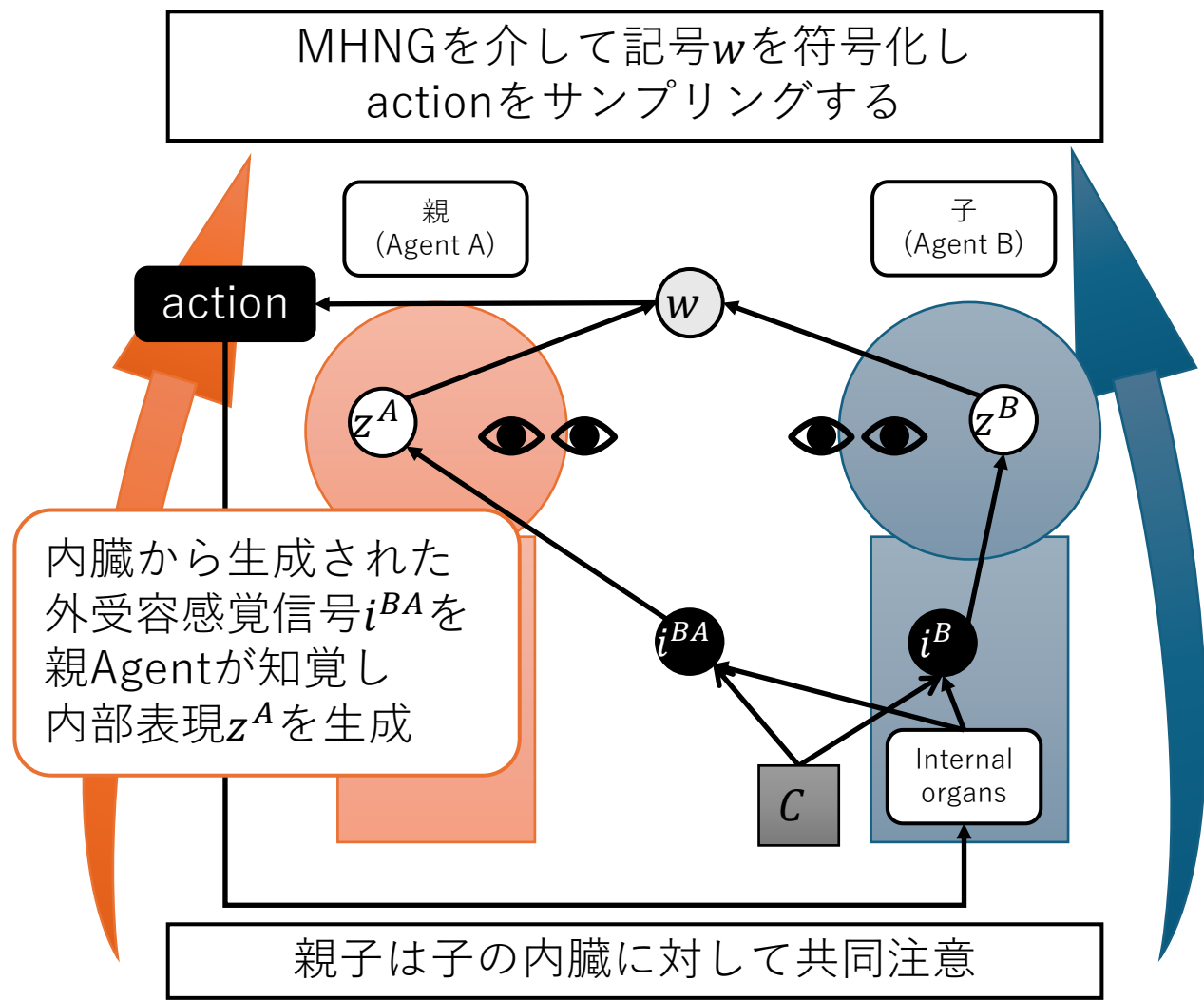
子：
・ 内臓状態は内受容感覚信号により知覚可能

・ 快適な内臓状態を維持する方法はわからない
(ex. ご飯を食べたらより快適か?)

POMDP + MHNG : 身体調整と記号創発を同じループに置く



POMDP + MHNG : 身体調整と記号創発を同じループに置く

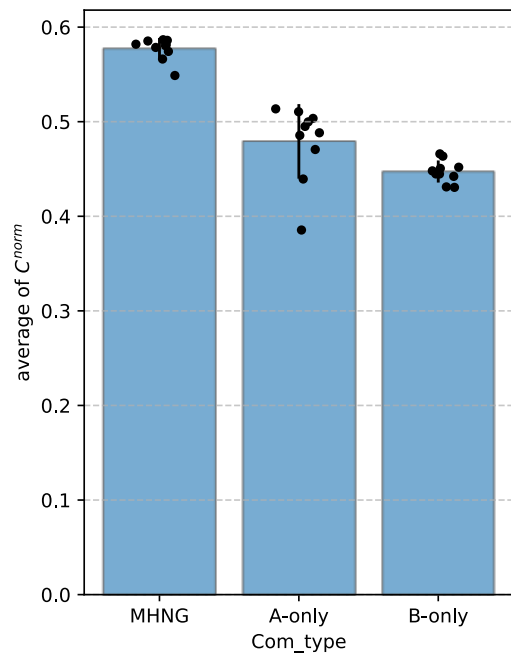


相互作用は身体を安定させ，記号を生み，表象を同期する

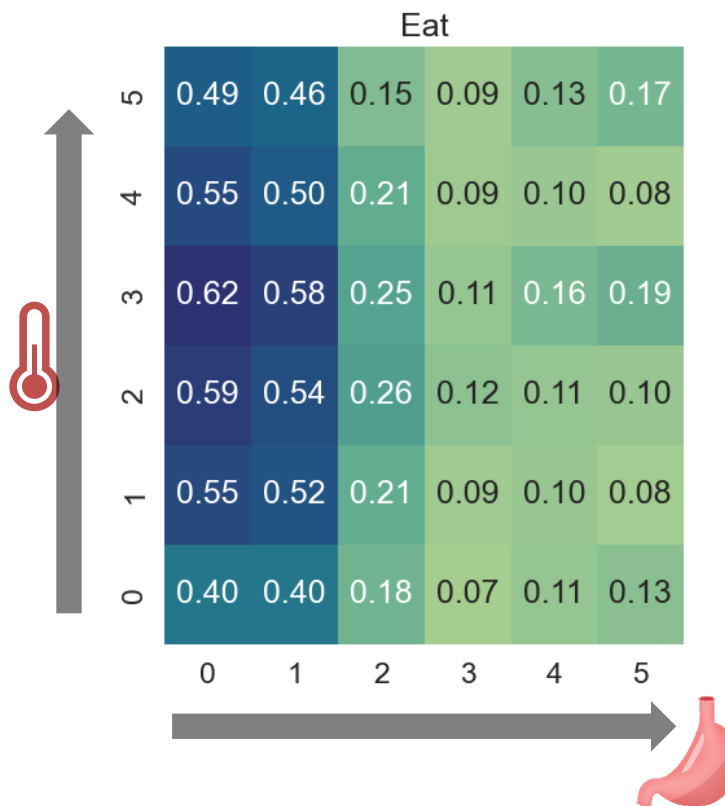
コミュニケーションは
子の内臓状態を最も安定させる

average of C^{norm}

$$= \text{mean}\left(\frac{C_t}{\max(C)}\right)$$

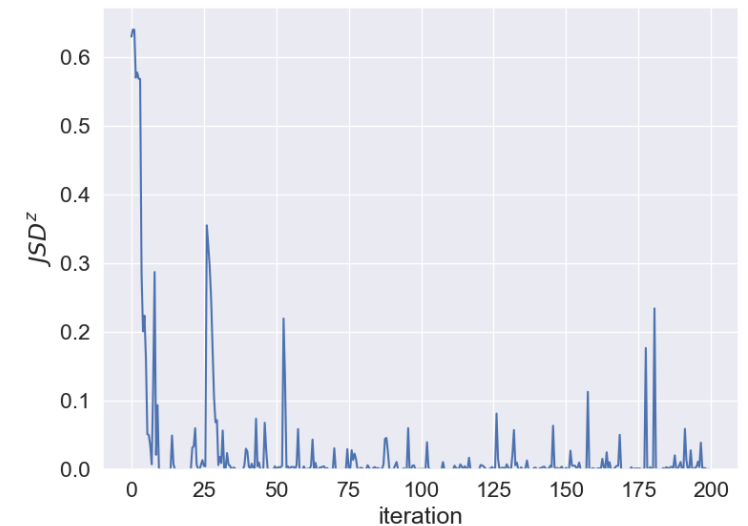


子は内臓状態を
符号化する記号を学習した



親子の内部表現の同期は
モデル学習より遥かに高速

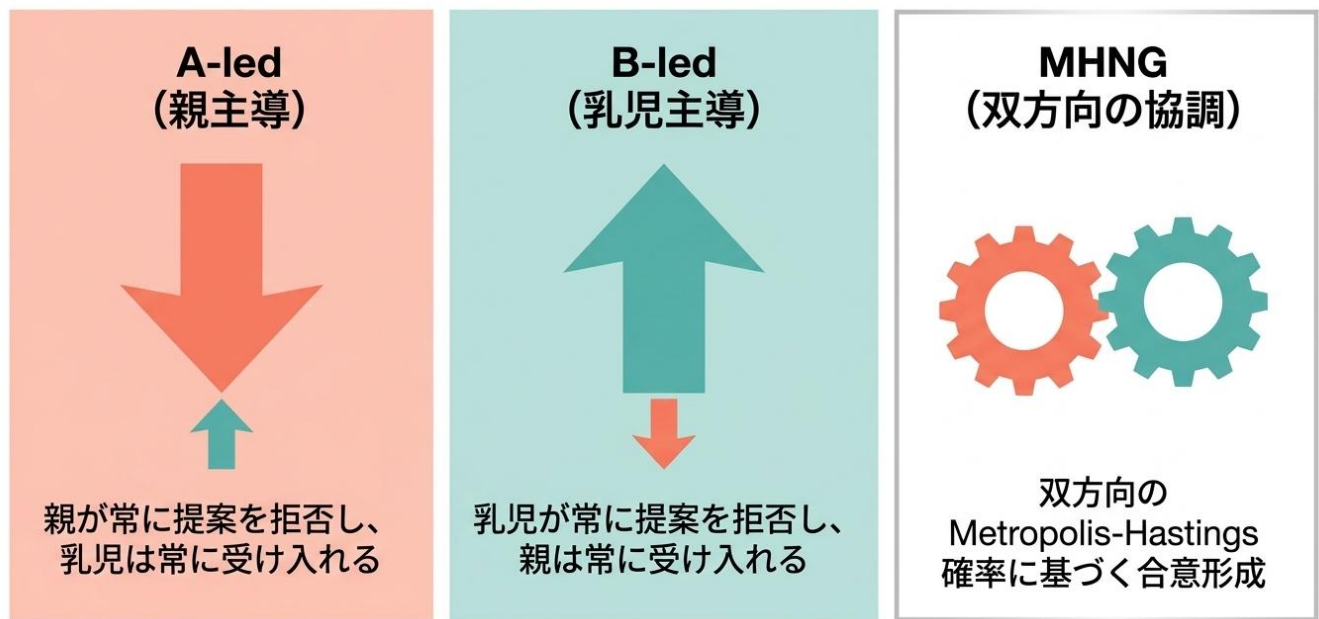
$$JSD^z = JSD[P(z_t^A | z_t^B)]$$



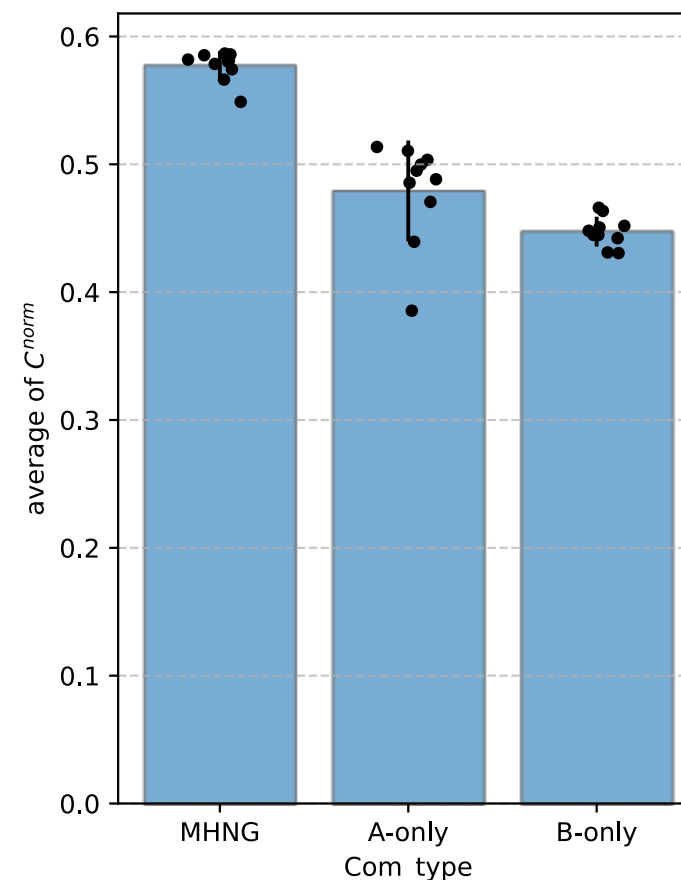
MHNGによる相互作用が「快適さ」を最大化する

- MHNG条件は一方通行の制御に比べ乳児の内臓状態を維持する能力が有意に高い
 - 完璧な知識を持つ独裁者よりも不完全な知識を持つ非対称パートナー同士の双方向相互作用が優れたパフォーマンスを発揮した

仮説の検証：3つの実験条件

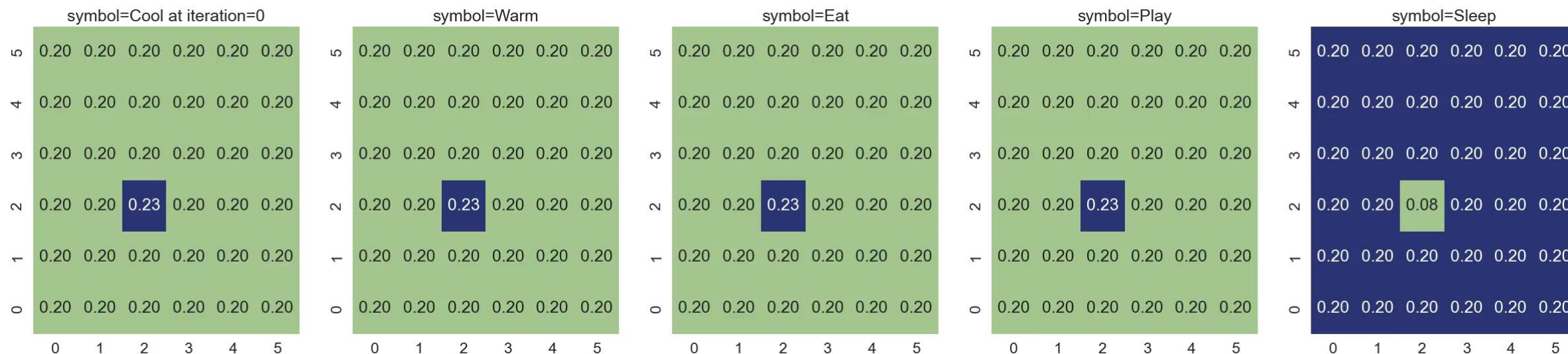


© NotebookLM

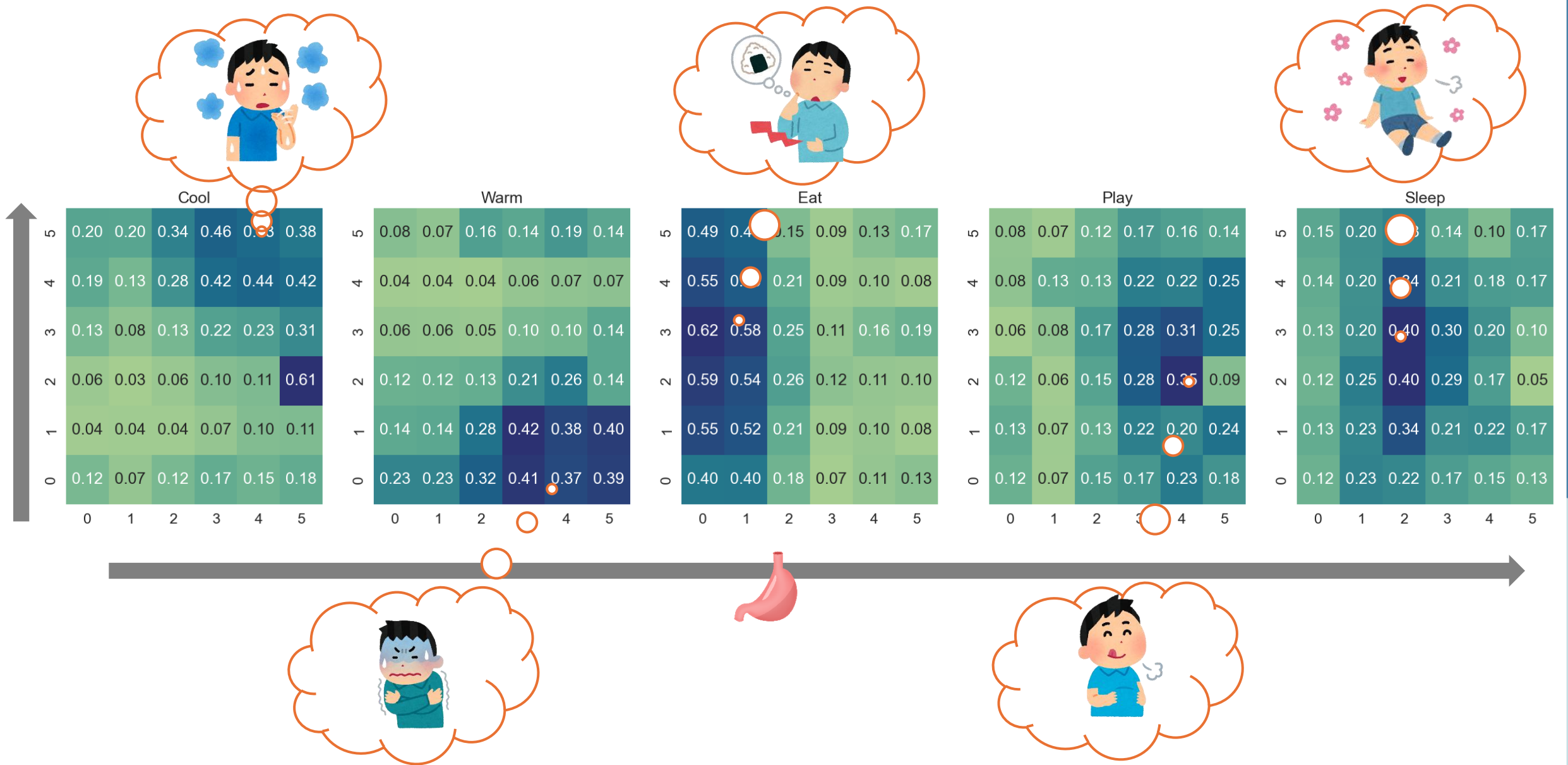


子は内臓状態を符号化する記号を学習した

- 自己の目的のために他者や外界を操作する記号（表現）が相互作用から創発する

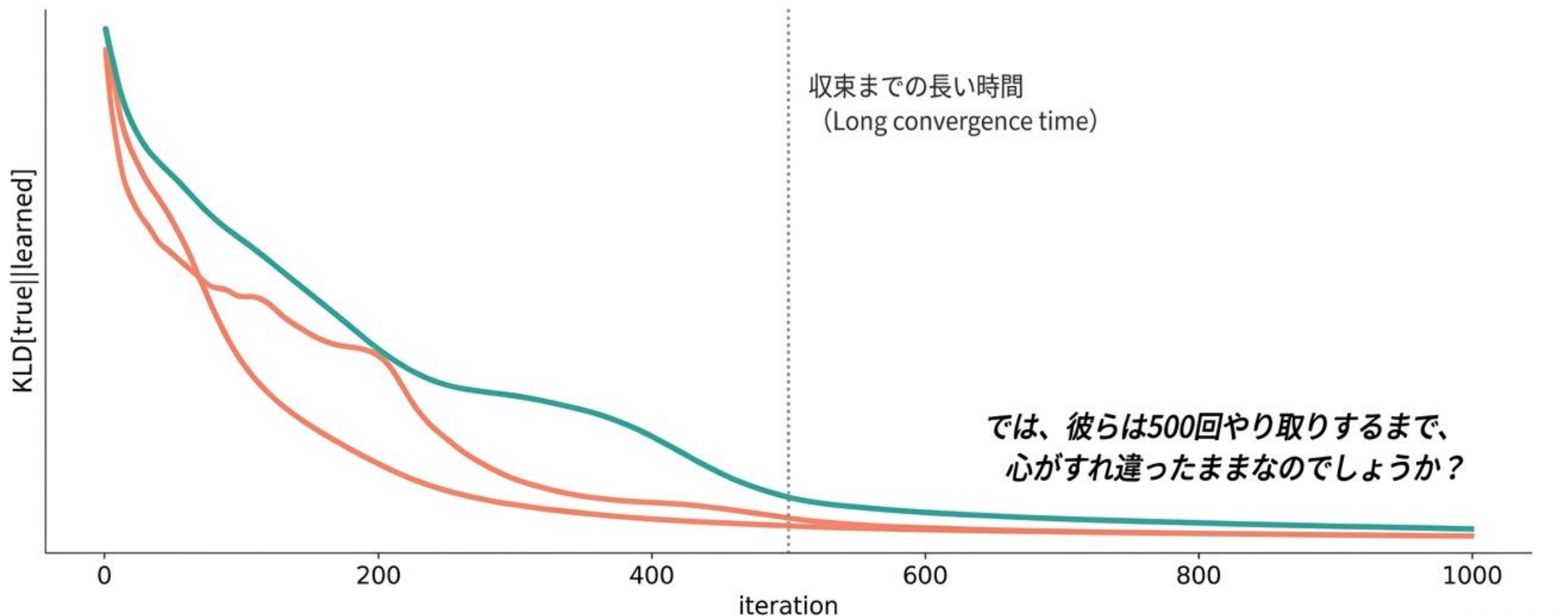


子は内臓状態を符号化する記号を学習した

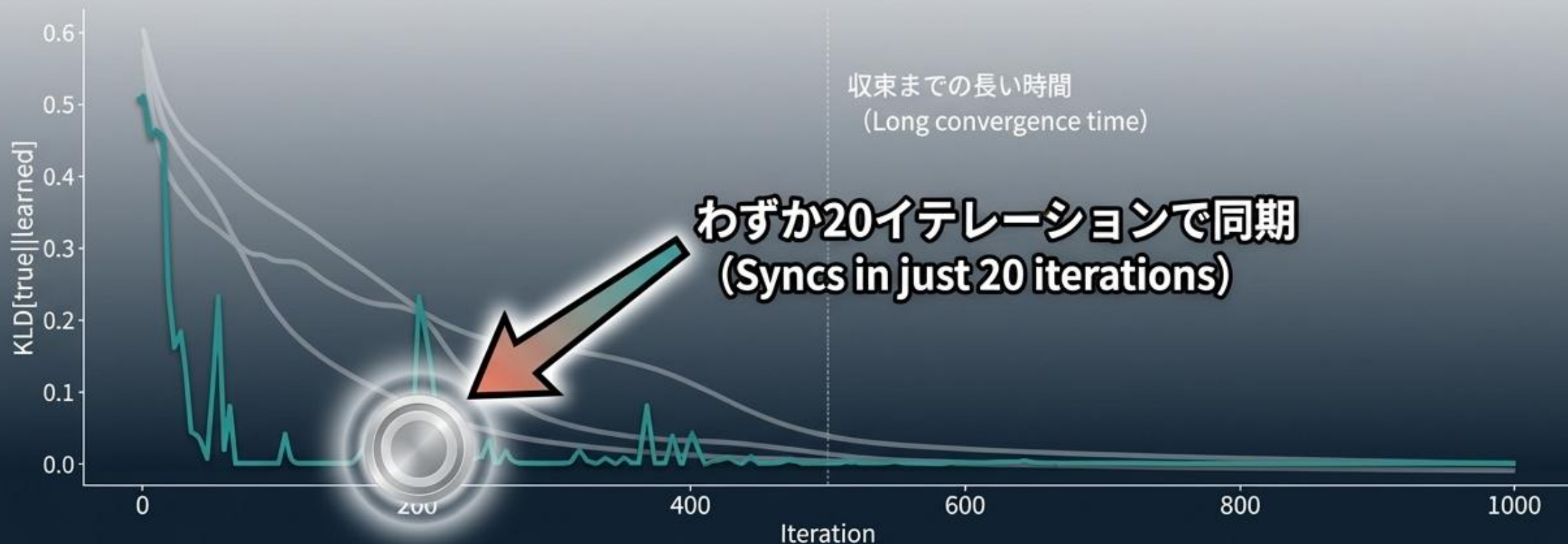


同期（Synchrony）は理解（Learning）に選好する

両エージェントは、それぞれの欠落した行列（ A^A および B^B ）を相互作用を通じて確実に学習していきます（KL情報量の減少）。しかし、この学習プロセスは非常に遅く、モデルが真のパラメータに収束するまでには約500イテレーションもの時間を要します。



最大の発見：同期（Synchrony）は、理解（Learning）に先行する



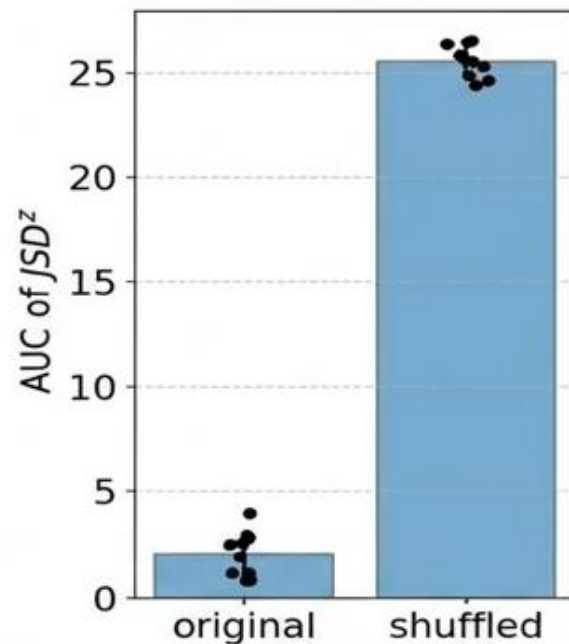
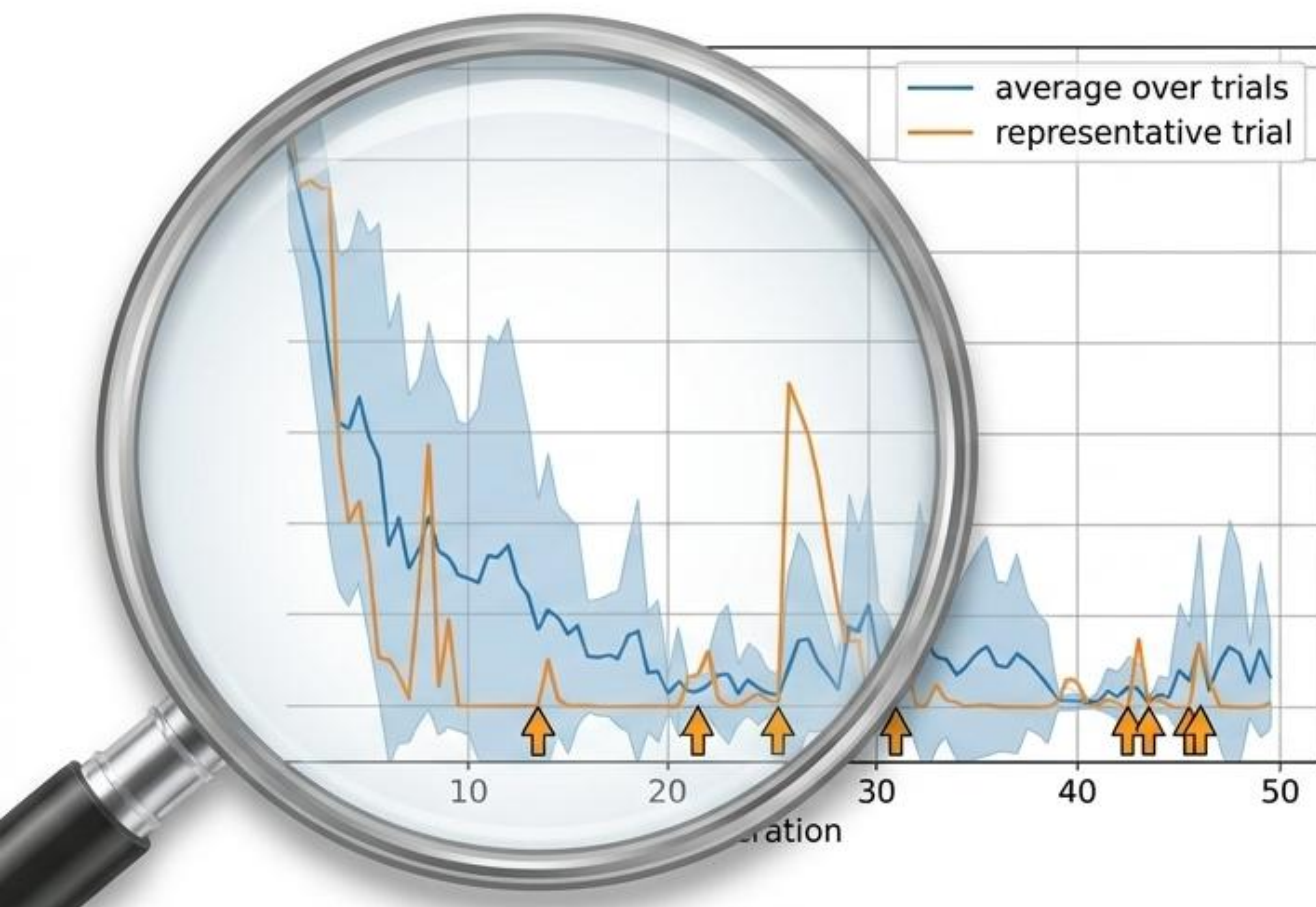
潜在表現の乖離（Jensen-Shannon divergence: JSD^z ）は、学習が収束するはるか手前、わずか20イテレーション付近でほぼゼロに低下します。

エージェントは、環境のルールを完全に学習する「25倍の速さ」で、互いの脳内表現を同期させます。完全な相互理解は、同期の必須条件ではありません。

「動的」な同期：揺らぎと即座の再アライメント

同期は静的なものではありません。まれな確率的な身体変化（例：食事後の予期せぬ体温低下）が起こると、非対称な知識の差により一時的な「脱同期（スパイク）」が発生します。

しかし、相互作用が強制的に即座の再アライメント（Re-alignment）を引き起こします。

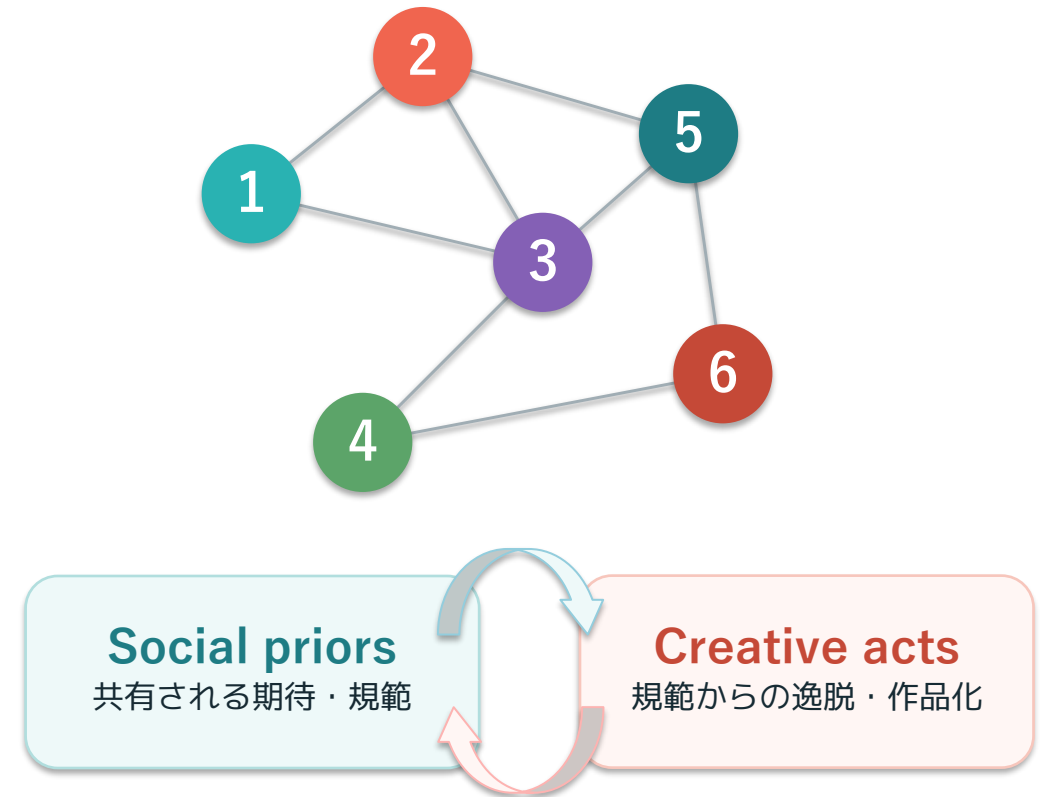
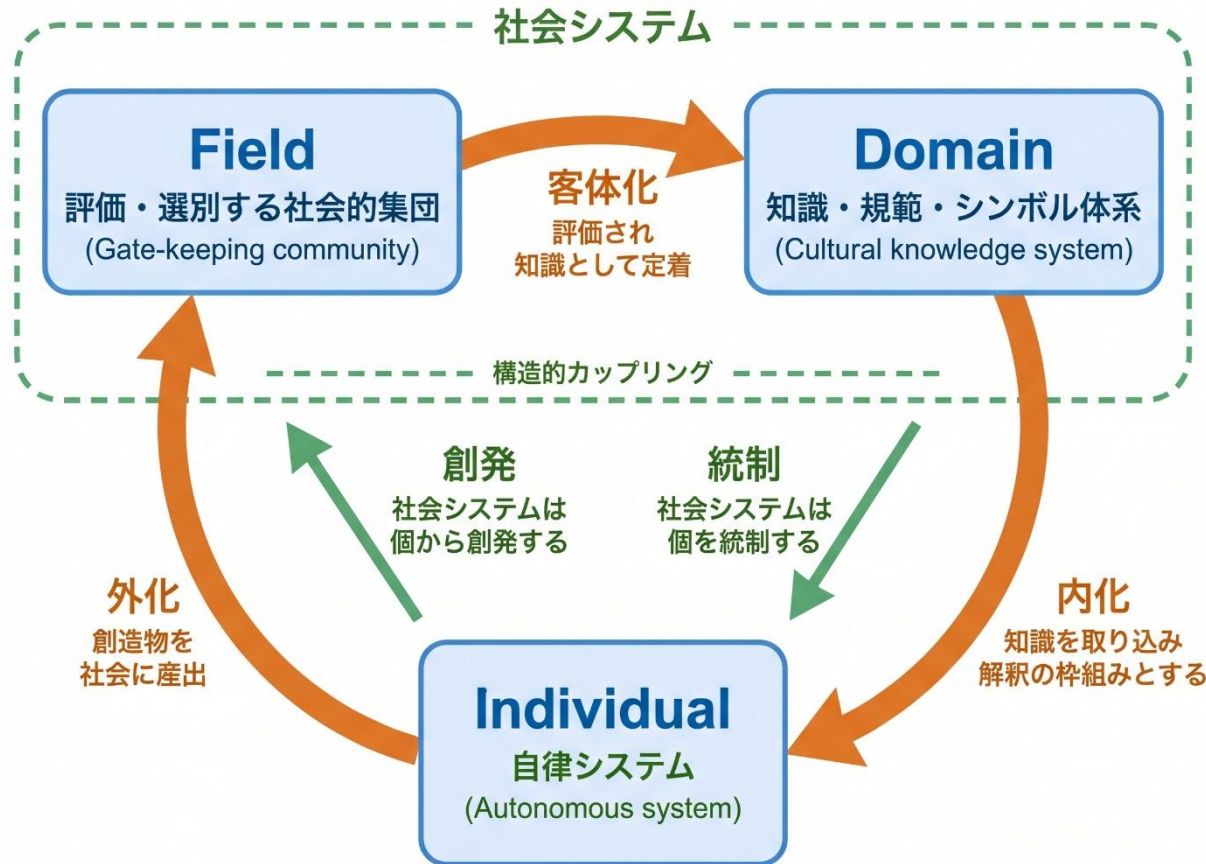


時系列をシャッフルした対照群 (Shuffled) と比較すると、この低い JSD^z が偶然の産物ではなく、リアルタイムの能動的な結合 (Coupling) であることが証明されます。

個人間規範から社会規範形成へ：目的は社会的事前分布になる

[野村・堀井, 2026] [Nomura and Horii, 2026]

- 「目的を共にもつ」とは同じゴールを内面化することではなく、共有された期待と許容される逸脱の範囲を、相互作用の中で更新し続けることかもしれない。



集団の同調と創造を同じ計算モデルで取り扱う

[野村・堀井, 2026]

[Nomura and Horii, 2026]

・ 能動的推論を通じて 社会的知識への同調と逸脱を同時に扱う

1 Creation

期待自由エネルギーを下げる
artifactを生成

2 Memory

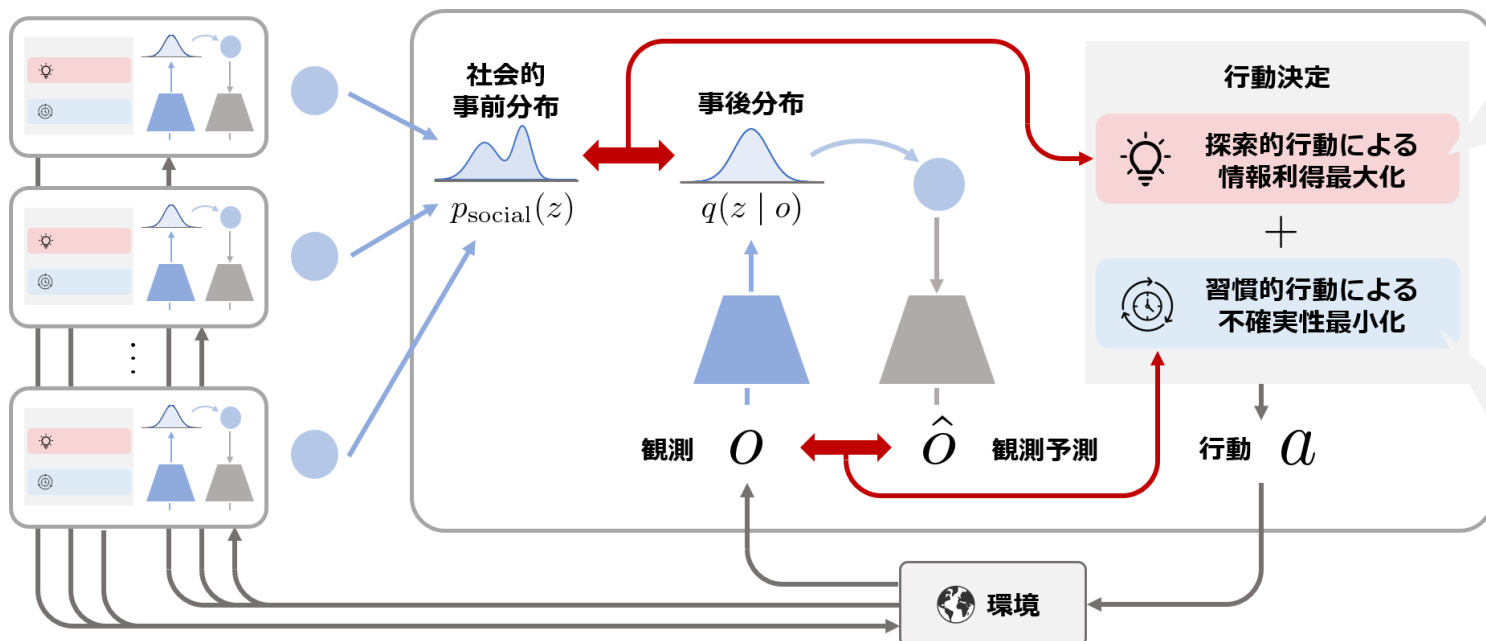
他者の創造物を受け入れるか
選ぶ

3 MHNG

共同注意下で社会表象を交換
・ 採択

4 Model update

生成モデルとdiscriminatorを
更新



期待自由エネルギー（逸脱）：

社会的知識から逸脱しつつ
認識論的恒常性は維持される行動をとる。

$$\mathcal{G}^k(o_+) = -D_{\text{KL}}[q_{\phi}^k(z|o_+) || p_{\text{social}}^k(z)] - \mathbb{E}_{q_{\phi}^k}[\log p_{\theta}^k(o_+|z)] \\ \approx -\mathbb{E}_{q_{\phi}^k}[D_{\psi}^k(z, o_+)] - \mathbb{E}_{q_{\phi}^k}[\log p_{\theta}^k(o_+|z)].$$

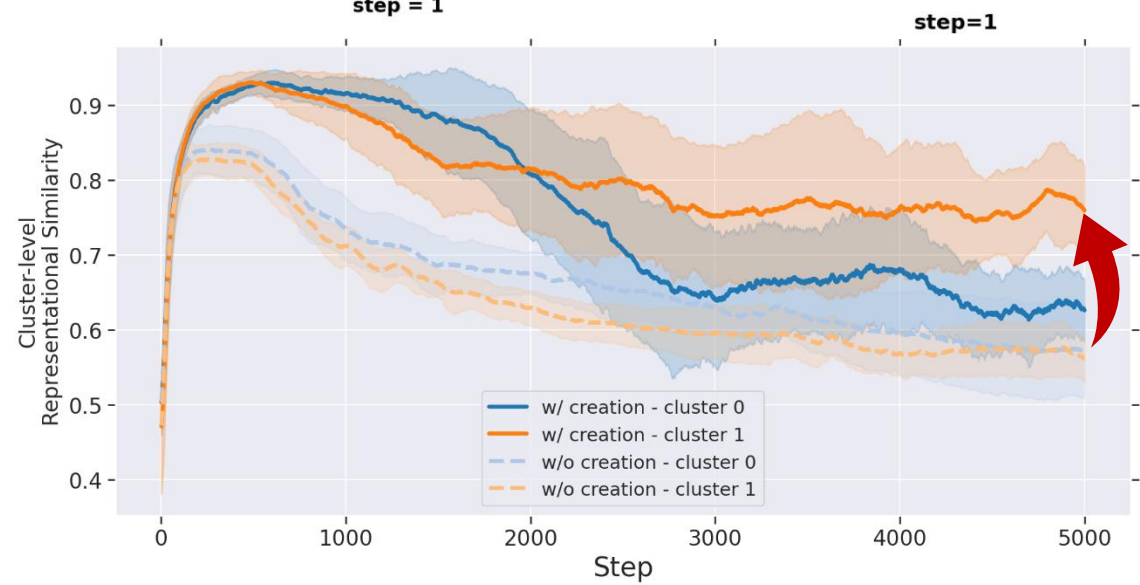
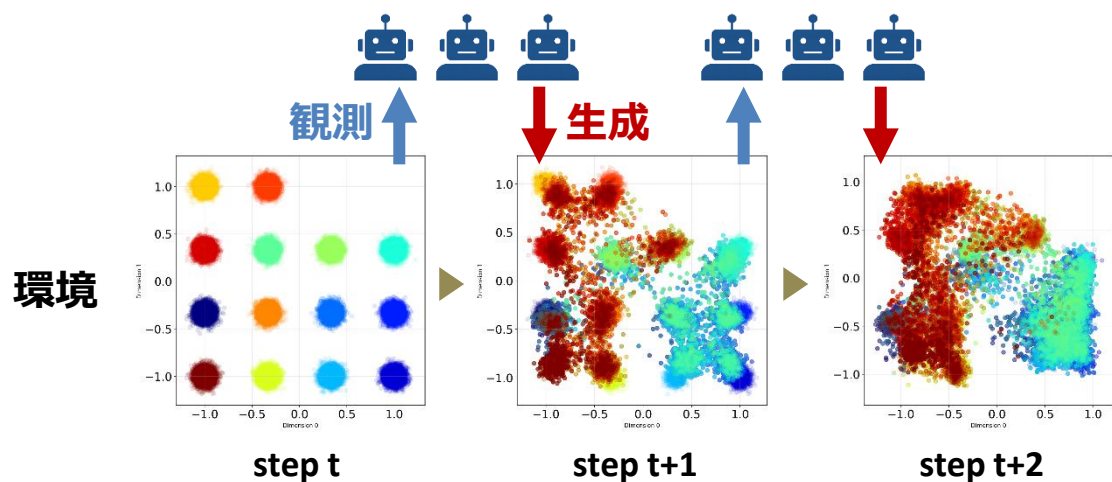
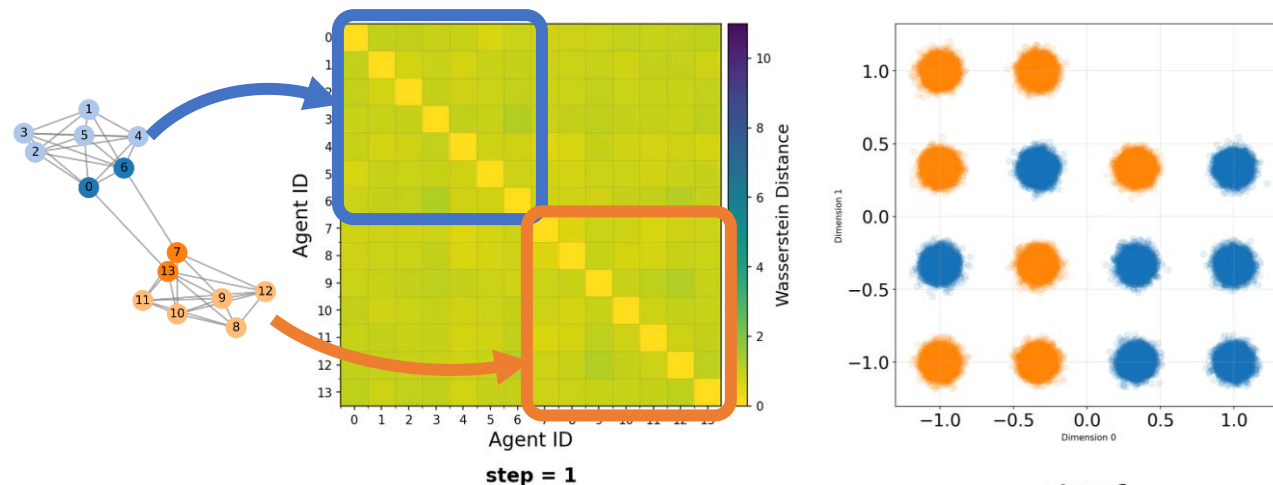
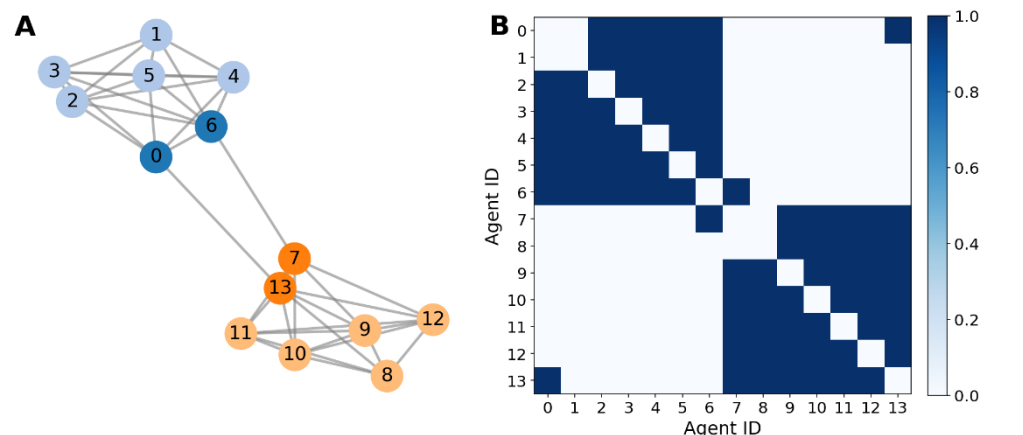
変分自由エネルギー（同調）：

社会的知識に個人の推論を近づけ
観測予測誤差を小さくする。

$$\mathcal{F}^k(\theta, \phi) = D_{\text{KL}}[q_{\phi}^k(z|o) || p_{\text{social}}^k(z)] - \mathbb{E}_{q_{\phi}^k}[\log p_{\theta}^k(o|z)] \\ \approx \mathbb{E}_{q_{\phi}^k}[D_{\psi}^k(z, o)] - \mathbb{E}_{q_{\phi}^k}[\log p_{\theta}^k(o|z)].$$

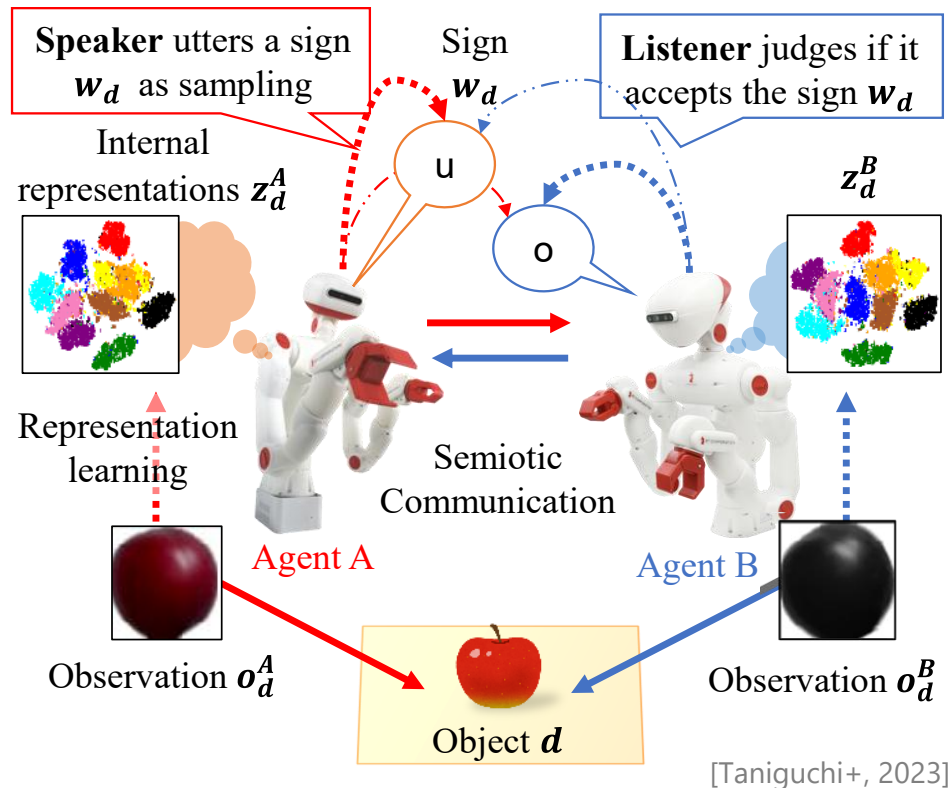
規範の分化と文化的ニッチ

- 社会表象は安定的に拡散する一方，創造物は特定ペアを通じて局所的に伝播し，文化的ニッチを分化させる。



環世界のすり合わせ：人間・他種・ロボットへ拡張する

- 身体が異なれば立ち上がる記号も目的も異なる
 - 同じ世界を共有するのではなく共有可能なカップリングを作る
 - 神経系や言語の前に空間共有と身体調整の記号過程があるのではないか



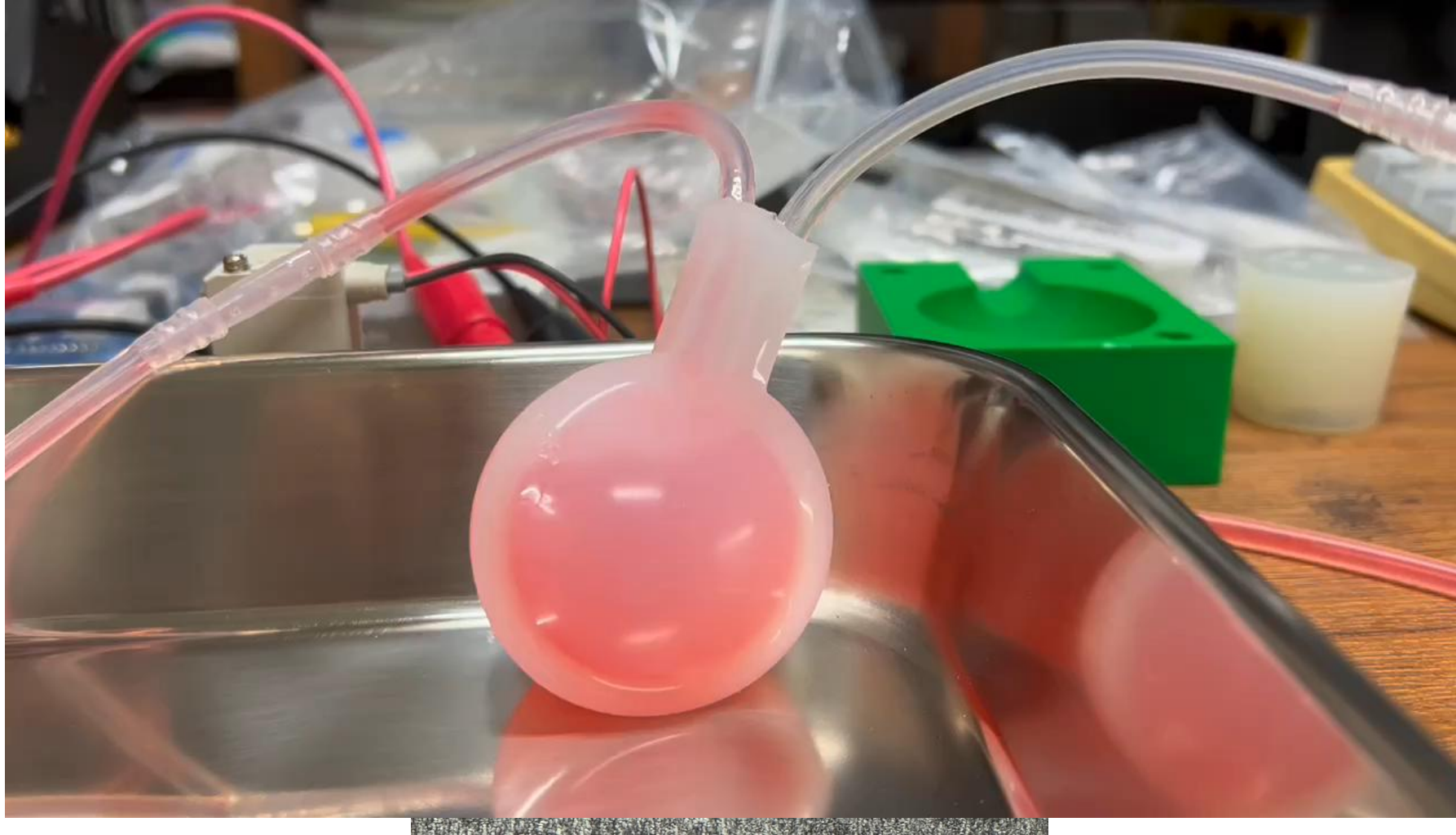
ロボットの「目的」はどこから来るのか

- ・ 内臓なき身体に，社会的アロスタシスは成立するか



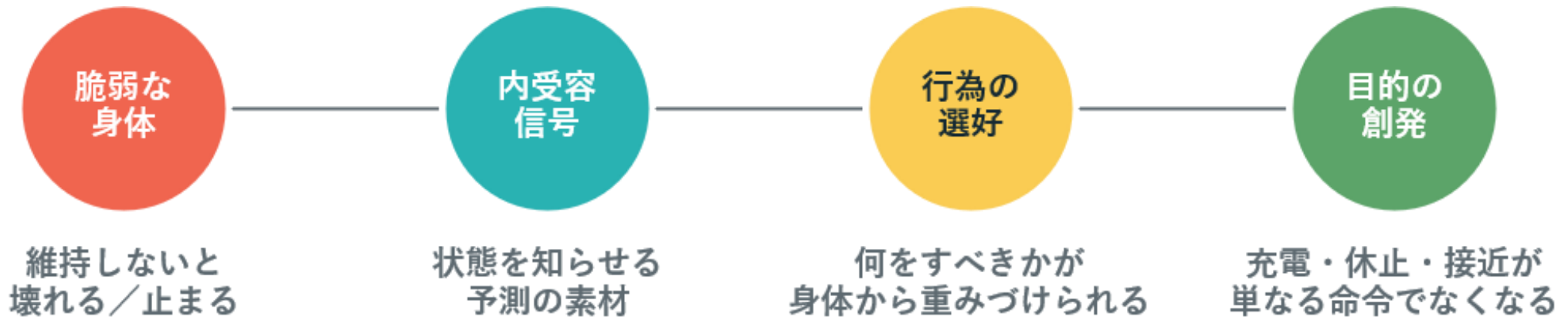
ロボットの「目的」はどこから来るのか

- ・ 内臓なき身体に，社会的アロスタシスは成立するか



なぜロボットに「内臓」が必要なのか

- 目的を外からのゴールではなく、維持すべき身体予算から立ち上げるために
 - ロボットの目的を「充電せよ」という記号として書き込むのではなく、維持すべき身体・内受容・行為選択の循環から立ち上げられるかを問う



要件 1

安定性と脆弱性を併せ持つ物理系

安定点はあるが、外乱に対して安定すぎない。

要件 2

認知処理を駆動する情報源であること

周期的・連続的な信号が予測の素材になる。

要件 3

身体予算管理のトレードオフを持つこと

上位の行為や学習がこの維持に依存する。

(物理的) システムとして安定性と脆弱性を有する

- セットポイントを持つ物理システム

- 特定のセットポイント付近で準安定状態を示し，外乱に対するロバスト性を持つ

- 系としてホメオスタシスを持つ

- 可制御性を持ちセットポイントを変更できる

- アロスタシスを実現できる

- 柔軟素材で構成されており外界からの相互作用によって上記の特性が変化する

- 身体接触に伴う内臓状態の伝播

PERSPECTIVE

<https://doi.org/10.1038/s42256-019-0103-7>

nature
machine intelligence

Homeostasis and soft robotics in the design of feeling machines

Kingson Man and Antonio Damasio

Attempts to create machines that behave intelligently often conceptualize intelligence as the ability to achieve goals, leaving unanswered a crucial question: whose goals? In a dynamic and unpredictable world, an intelligent agent should hold its own meta-goal of self-preservation, like living organisms whose survival relies on homeostasis: the regulation of body states aimed at maintaining conditions compatible with life. In organisms capable of mental states, feelings are a mental expression of the state of life in the body and play a critical role in regulating behaviour. Our goal here is to inquire about conditions that would potentially allow machines to care about what they do or think. Under certain conditions, machines capable of implementing a process resembling homeostasis might also acquire a source of motivation and a new means to evaluate behaviour, akin to that of feelings in living organisms. Drawing on recent developments in soft robotics and multisensory abstraction, we propose a new class of machines inspired by the principles of homeostasis. The resulting machines would (1) exhibit equivalents to feeling; (2) improve their functionality across a range of environments; and (3) constitute a platform for investigating consciousness, intelligence and the feeling process itself.



認知的な処理のための情報源としての動作

• 周期的な情報元としての人工内臓

– Physical reservoir computing (PRC)により
内臓状態と所望の情報処理の関係を学習

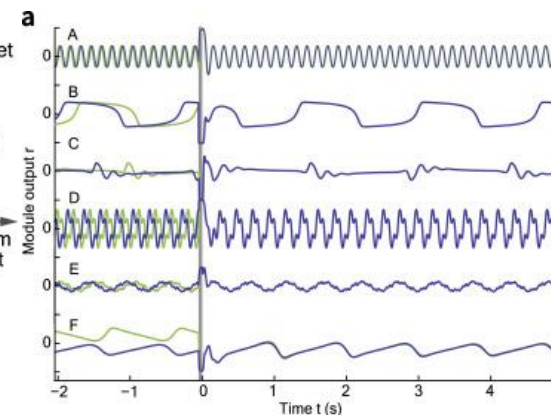
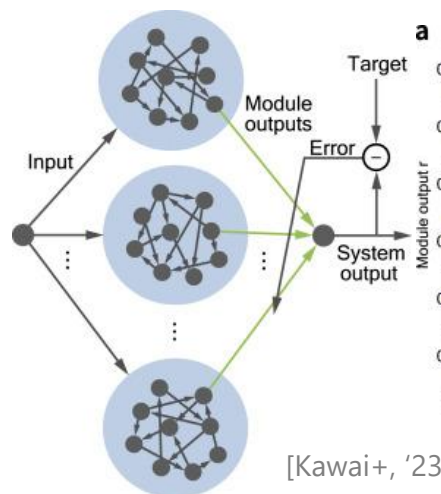
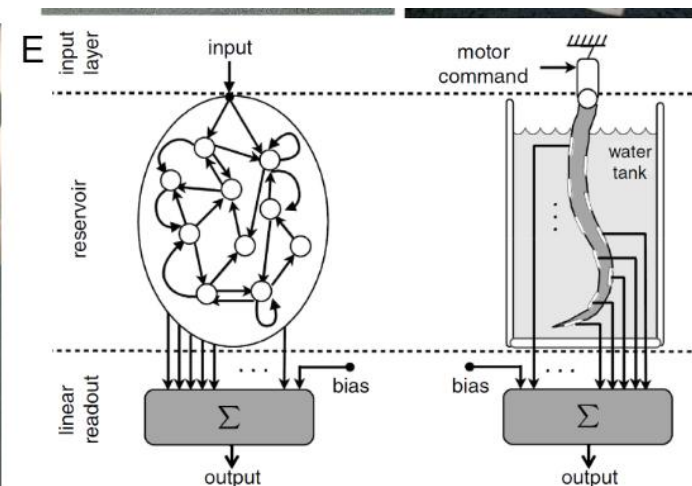
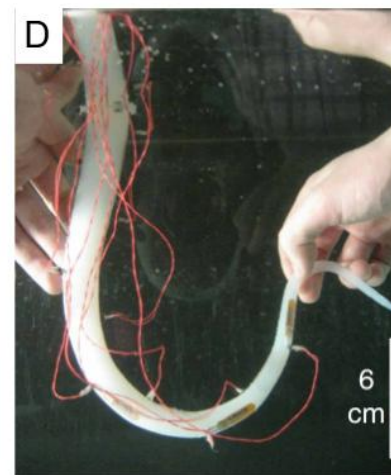
• システムが信号を生成することで
身体反応と行動生成の関係をモデル化

• 心臓の収縮期は拡張期に比べて
リスク志向的意思決定の頻度が増加

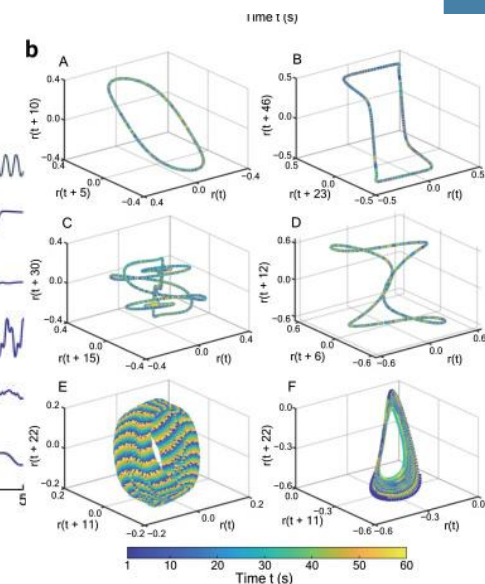
[Kimura+, 2023]

– 時定数の異なる複数の内臓信号に
よって単一のシステムよりも複雑な
情報処理を実現 (i.e., 身体性)

[Nakajima, 2020]



[Kawai+, '23]

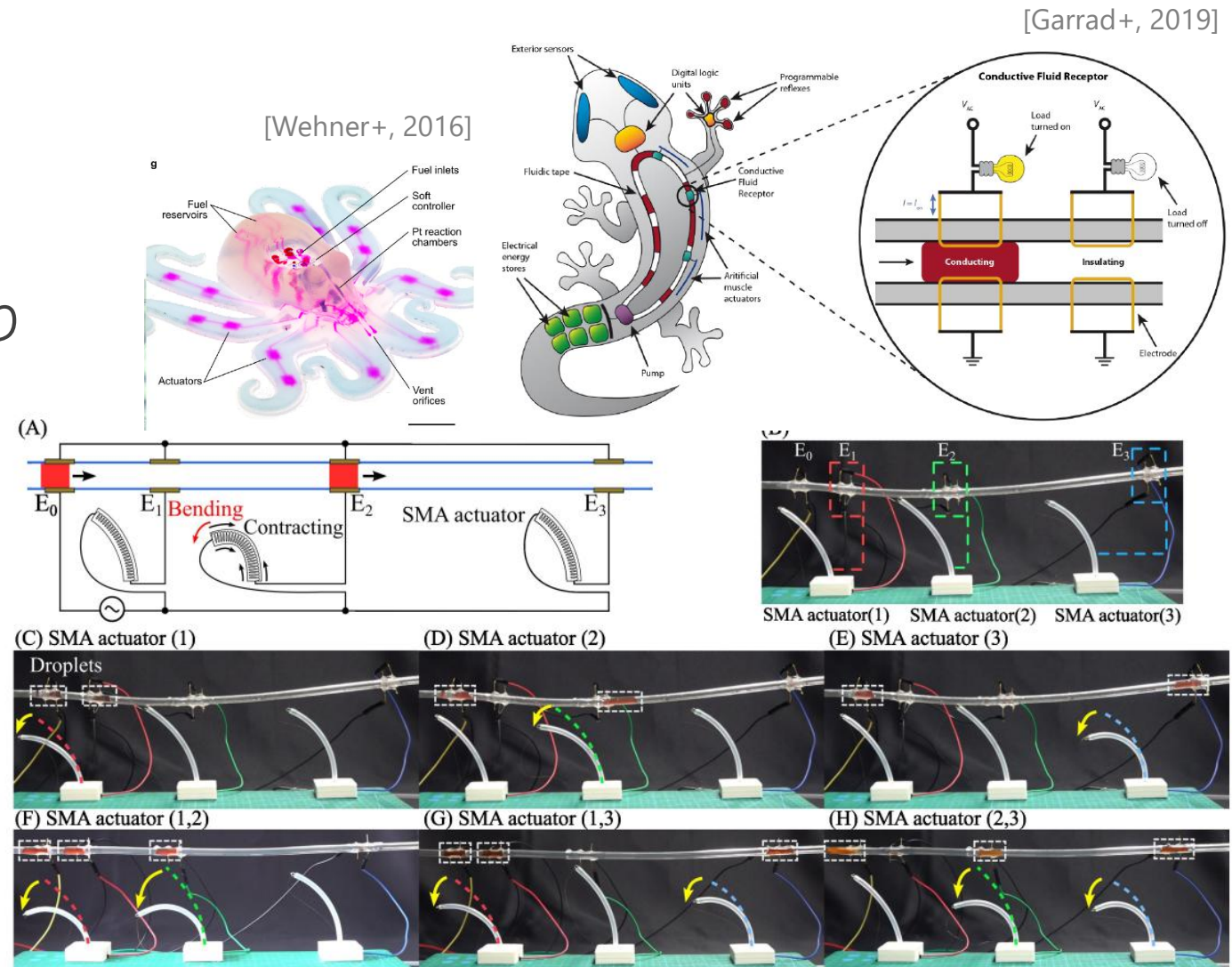
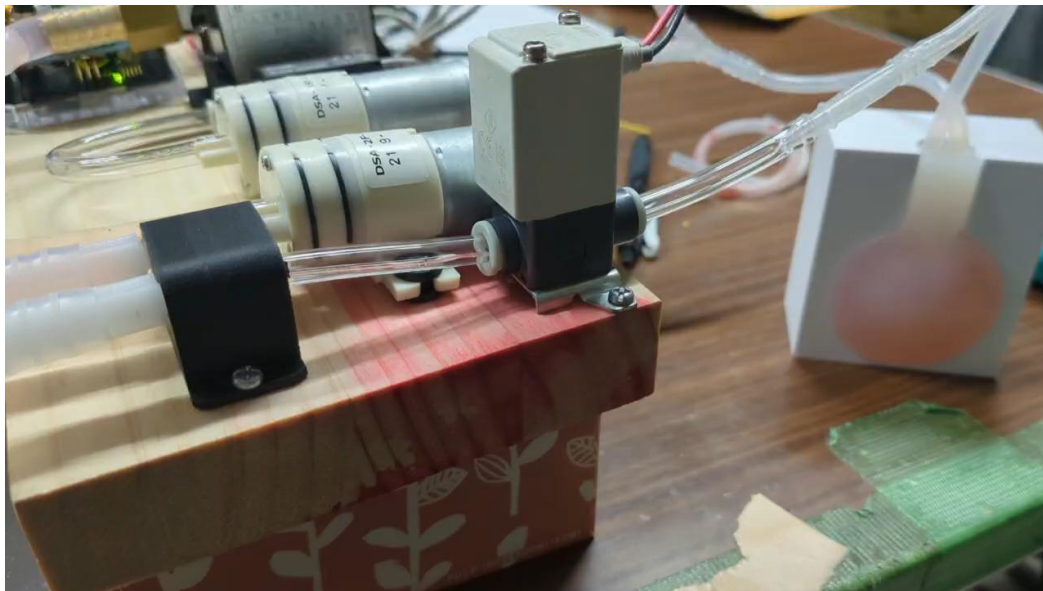


認知的な処理のための情報源としての動作

- 計算機としての人工内臓

- Soft matter computing (SMC)による
計算・情報処理

- PRCと異なり埋め込まれた構造により
直接的な情報処理を実現



Co-constructed Emotion: A Collective Predictive Coding Theory of Interoception and Social Symbol Sharing

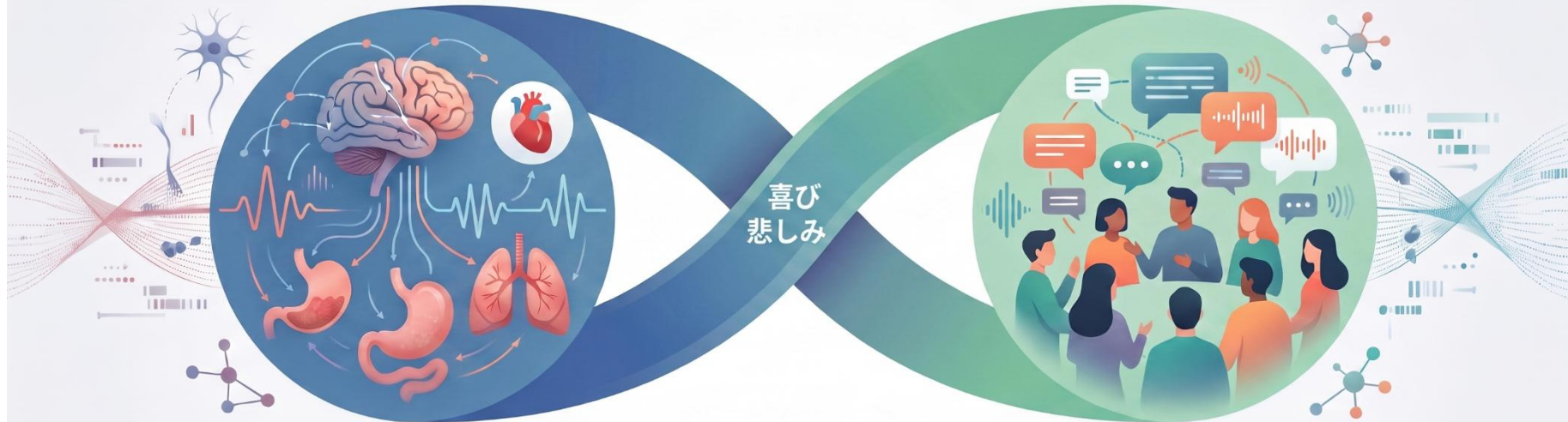
・CPCとアロスタシスを基盤とした社会的現実としての感情の共構成を議論

[Horii, Hieida, Terasawa, Ohira, Taniguchi, 2026]

感情の共構築理論 (TCoE) : 身体と社会を繋ぐ計算モデル

内部ループ：アロスタシス（身体調節）

社会的ループ：記号共有（コミュニケーション）



脳が心拍などの内受容感覚を予測し、身体状態を維持するために生理反応を調節するプロセス。

身体感覚が言葉を作り、共有された言葉が個人の生理状態を形作るという双方向のサイクル。

「不安」や「怒り」といった共通の感情語を用いて、集団全体で予測誤差を最小化するプロセス。

臨床的応用：失感情症（アレキシサイミア）の再定義



身体感覚と感情概念の「切断」
内受容感覚の予測精度が低いため、身体の変化を社会的な感情語に結びつけられない状態。



予測の不安定性によるラベリングの失敗

身体信号が曖昧なため、どの感情語（記号）が適切か判断できず、受容・認識が困難になる。

問い

- **記号は「表象」か，身体・社会の「レギュレーション」か？**
 - パース的な習慣・解釈項は，脳内表象ではなく，身体と共同体の調整ループとして読めるか。
- **共に生きるロボットは，私たちの規範に同調すべきか？**
 - 倫理は目的の共有からではなく，ズレを含むカップリングを保護する設計から始まるのではないか。
- **創造性は「目的の失敗」か，「規範を更新する逸脱」か？**
 - 目的がズレることは単なる誤りではなく，次の目的・記号・共同体を作るアブダクションなのではないか。

問1：記号は「表象」か、身体・社会の「レギュレーション」か？

- ・ 身体予算の安定化，共同注意，社会的事前分布の調整までを記号過程に含めるなら，意味は「何を指すか」だけでなく「何を可能にするか」になる．
- ・ パース的な習慣・解釈項は，脳内表象ではなく，身体と共同体の調整ループとして読めるか．

問2：共に生きるロボットは、私たちの規範に同調すべきか？

- ・ 人間・ロボット・他種が「同じ目的」を持つことを共生と呼ぶなら、異なる環世界をもつ存在を同化する危険がある。
- ・ 倫理は目的の共有からではなく、ズレを含むカップリングを保護する設計から始まるのではないか。

問3：創造性は「目的の失敗」か、「規範を更新する逸脱」か？

- 能動的推論で見れば，規範への同調は予測誤差を下げる。
一方，創造的行動は局所的には逸脱だが，採択されれば新しい事前分布になる。
- 目的がズれることは単なる誤りではなく，次の目的・記号・共同体を作る
アブダクションなのではないか。