科学のモデルとしての集合的予測符号化：
生成科学に向けた記号創発システムアプローチ

Collective Predictive Coding as a Model of Science:
A Symbol Emergence Systems Approach Towards Generative Science

Tadahiro Taniguchi

1) Professor, Graduate School of Informatics, Kyoto University

2) Visiting Professor, Research Organization of Science and Technology, Ritsumeikan University

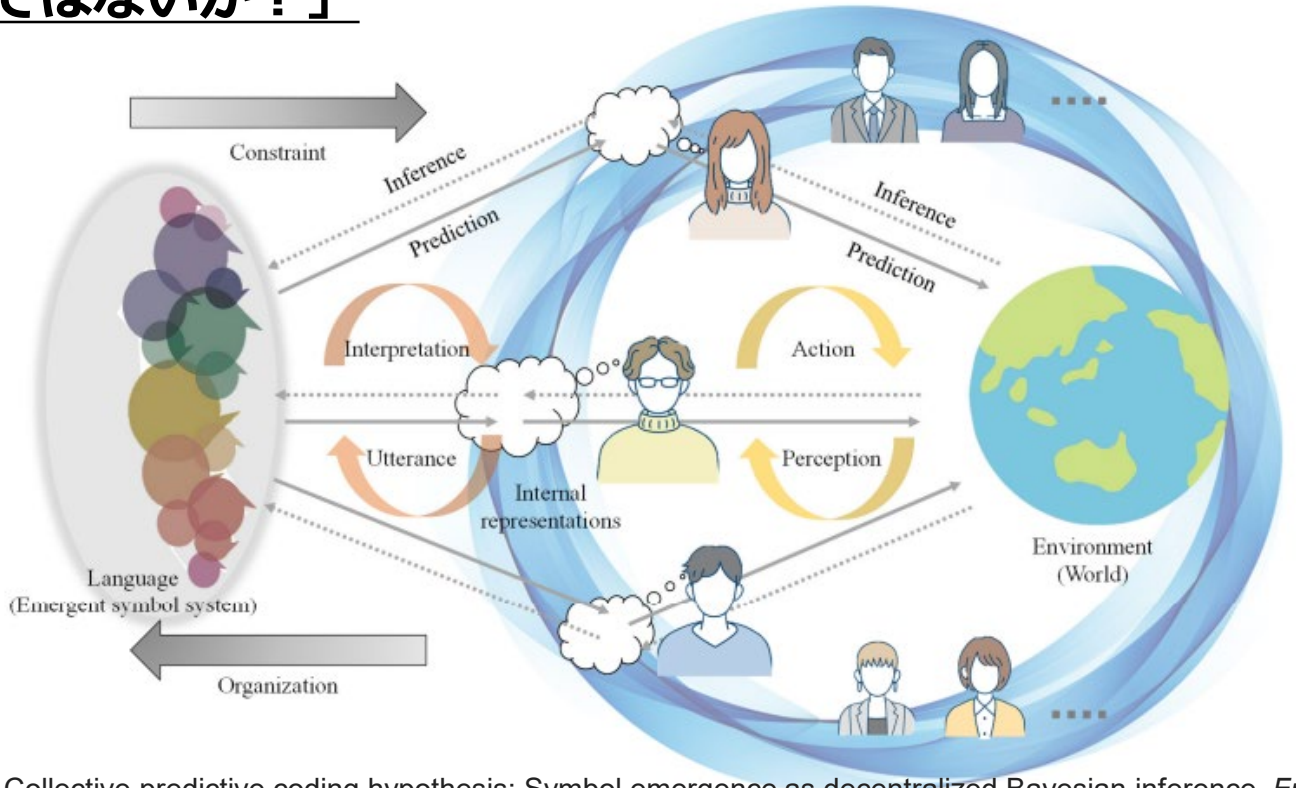
3) Senior Technical Advisor, Panasonic Holdings Corporation

「学習物理学の創成」領域セミナー

26th June 2025

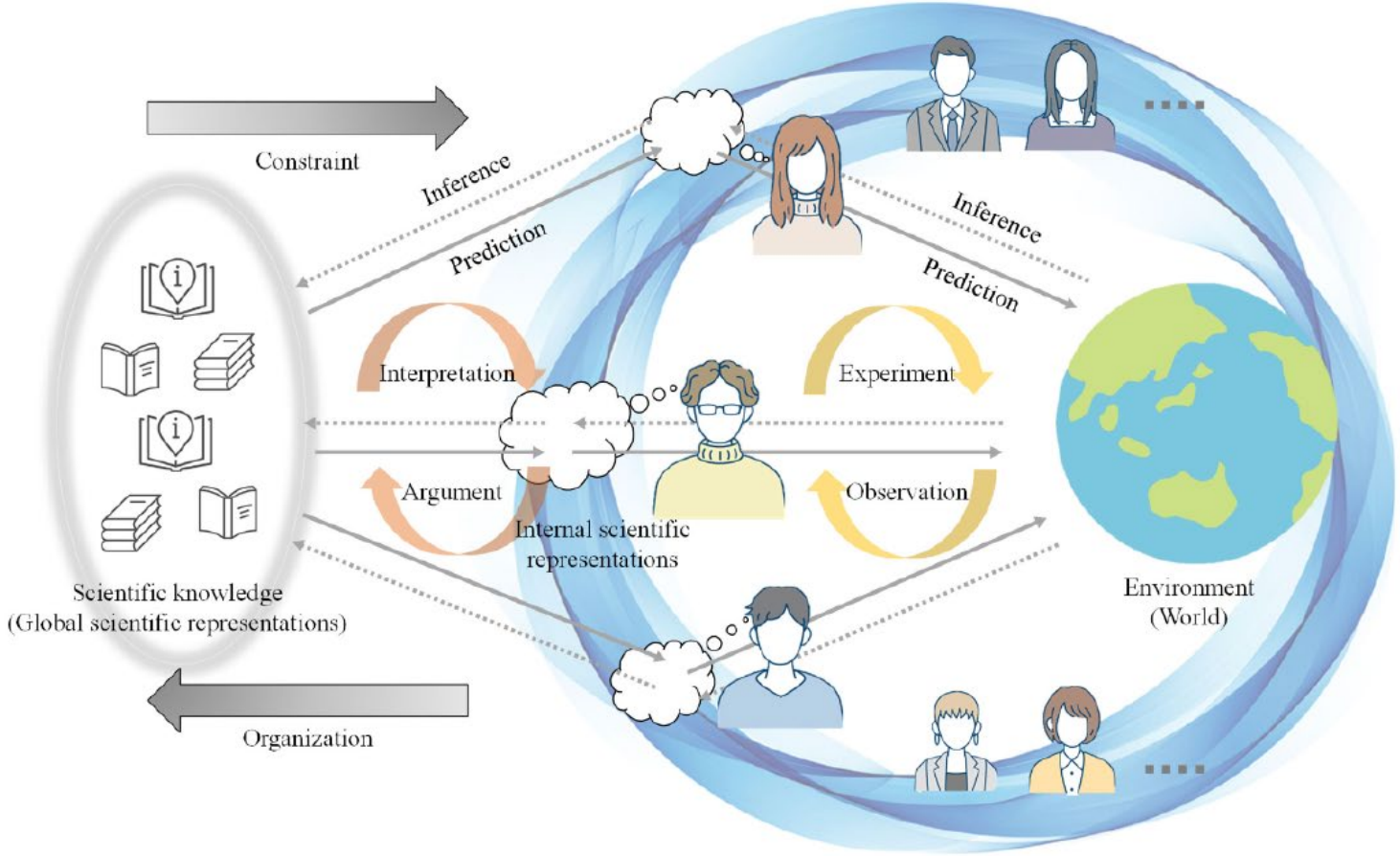
集合的予測符号化仮説 [Taniguchi ‘23]

- 言語そのものが**集合的な予測符号化**によって形成されるために、世界の情報が分布意味論の中にコーディングされている。
- **記号創発システム**の**社会的な表現学習装置**として理解できる。
- **「私たちは世界をより良く予測するために言語や規範を形成しているのではないか？」**



Taniguchi, T. (2024). Collective predictive coding hypothesis: Symbol emergence as decentralized Bayesian inference. *Frontiers in Robotics and AI*, *11*, 1353870.
谷口忠大. (2024). 集合的予測符号化に基づく言語と認知のダイナミクス: 記号創発ロボティクスの新展開に向けて. 認知科学, *31*(1), 186-204.

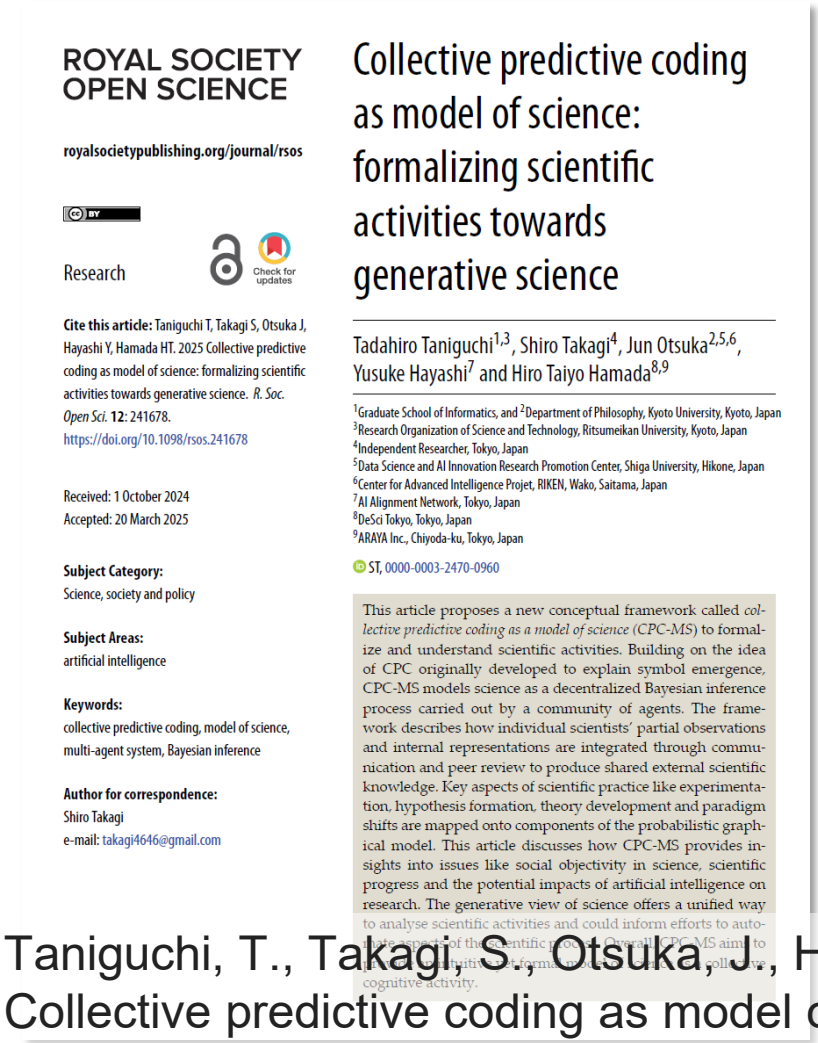
Collective Predictive Coding as Model of Science (CPC-MS): Formalizing Scientific Activities Towards Generative Science



科学を集合的予測符号化を行う記号創発システムとして解釈する

Collective Predictive Coding as Model of Science
高木志郎@第2回AIロボット駆動科学研究会

生成科学に向けて



ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Research

Cite this article: Taniguchi T, Takagi S, Otsuka J, Hayashi Y, Hamada HT. 2025 Collective predictive coding as model of science: formalizing scientific activities towards generative science. *R. Soc. Open Sci.* **12**: 241678.
https://doi.org/10.1098/rsos.241678

Received: 1 October 2024
Accepted: 20 March 2025

Subject Category:
Science, society and policy

Subject Areas:
artificial intelligence

Keywords:
collective predictive coding, model of science, multi-agent system, Bayesian inference

Author for correspondence:
Shiro Takagi
e-mail: takagi4646@gmail.com

Collective predictive coding as model of science: formalizing scientific activities towards generative science

Tadahiro Taniguchi[1,3], Shiro Takagi[4], Jun Otsuka[2,5,6], Yusuke Hayashi[7] and Hiro Taiyo Hamada[8,9]

[1]Graduate School of Informatics, and [2]Department of Philosophy, Kyoto University, Kyoto, Japan
[3]Research Organization of Science and Technology, Ritsumeikan University, Kyoto, Japan
[4]Independent Researcher, Tokyo, Japan
[5]Data Science and AI Innovation Research Promotion Center, Shiga University, Hikone, Japan
[6]Center for Advanced Intelligence Projet, RIKEN, Wako, Saitama, Japan
[7]AI Alignment Network, Tokyo, Japan
[8]DeSci Tokyo, Tokyo, Japan
[9]ARAYA Inc., Chiyoda-ku, Tokyo, Japan

ST, 0000-0003-2470-0960

This article proposes a new conceptual framework called *collective predictive coding as a model of science (CPC-MS)* to formalize and understand scientific activities. Building on the idea of CPC originally developed to explain symbol emergence, CPC-MS models science as a decentralized Bayesian inference process carried out by a community of agents. The framework describes how individual scientists' partial observations and internal representations are integrated through communication and peer review to produce shared external scientific knowledge. Key aspects of scientific practice like experimentation, hypothesis formation, theory development and paradigm shifts are mapped onto components of the probabilistic graphical model. This article discusses how CPC-MS provides insights into issues like social objectivity in science, scientific progress and the potential impacts of artificial intelligence on research. The generative view of science offers a unified way to analyse scientific activities and could inform efforts to automate aspects of the scientific process. Overall, CPC-MS aims to [illegible] cognitive activity.

Tadahiro Taniguchi
@tanichu

Jun Otsuka
@junotk_jp

Yusuke Hayashi
@hayashiyus

Hiro Taiyo Hamada
@HiroTHamadaJP

Shiro Takagi
@takagi_shiro

Taniguchi, T., Takagi S., Otsuka, J., Hayashi, Y., & Hamada, H. T. (2025). Collective predictive coding as model of science: Formalizing scientific activities towards generative science. Royal Society Open Science (in press) (*arXiv preprint arXiv:2409.00102)*.

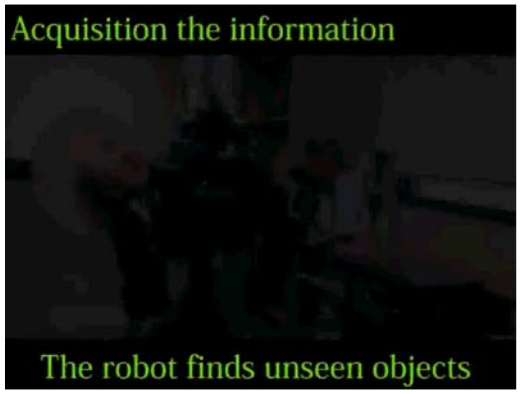
谷口忠大 (Tadahiro Taniguchi) @tanichu

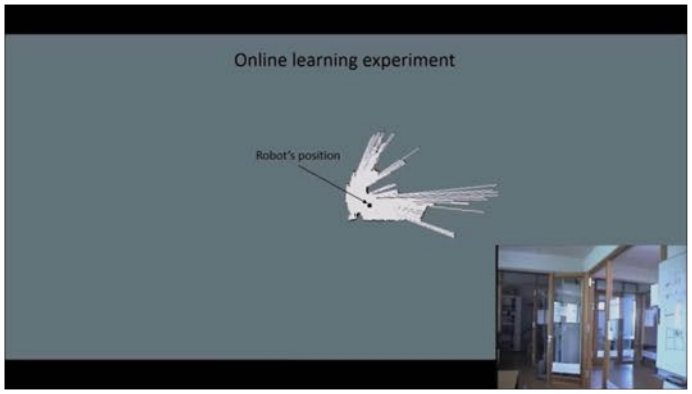
経歴

- 2006: 京都大学大学院工学研究科精密工学専攻修了 博士（工学）
- 2005: 日本学術振興会特別研究員(DC→PD)京都大学 （所属同上）
- 2007: 日本学術振興会特別研究員(PD)京都大学
  京都大学大学院情報学研究科システム科学専攻
- 2008: 立命館大学情報理工学部助教
- 2010-: 立命館大学情報理工学部准教授
- 2015-2016 インペリアル・カレッジ・ロンドン客員准教授
- **2016-: 一般社団法人ビブリオバトル協会代表理事**
- 2017-2024: 立命館大学情報理工学部教授
- 2017-2024: パナソニック客員総括主幹技師
  (クロスアポイントメント)
- **2024-: 京都大学大学院情報学研究科教授**
- **2024-: 立命館大学総合科学技術研究機構客員教授**
- **2024-: パナソニック・シニアテクニカルアドバイザー**
- **2024-: 一般社団法人Tomorrow Never Knows理事**
- **2025-: 一般社団法人AIロボット協会（AIRoA）理事**

書籍はAmazonで「**谷口忠大**」で検索

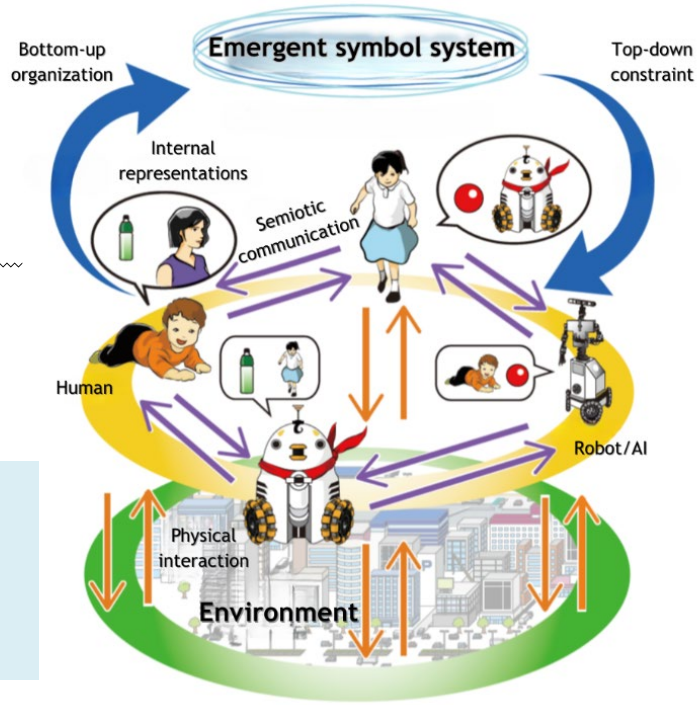
記号創発ロボティクス (2012-)





- **実世界経験**に基づく**発達・言語獲得ロボット**の実現とそれを通した人間の**認知発達の構成論的理解**を目指す.
- また**記号（言語）**を生み出し世界を理解し，協調する**適応的な自律的知能**を構成し理解する.

記号創発システムへの
構成論的アプローチ



記号創発ロボティクス/システム論

2010
記号創発システム
という概念の導入

2014
記号創発ロボティクス
構成論の哲学

2020
認知科学への接続
生成モデルとして

――心と言語・社会と身体をつなぐ新しいシステム論

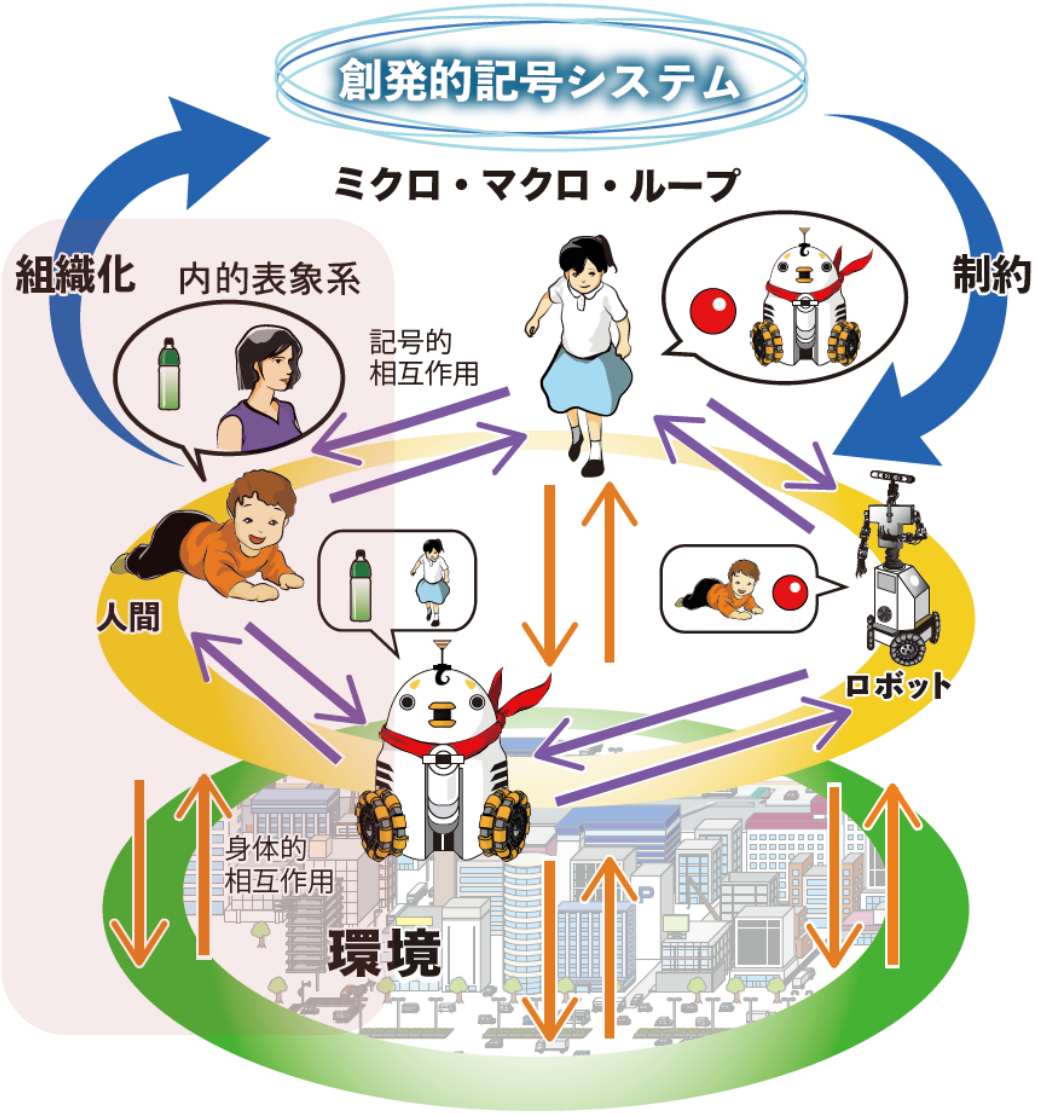
『ワードマップ　記号創発システム論』
谷口忠大（編著）新曜社 2024

項目タイトル

記号創発システム, 記号創発ロボティクス, 記号論と意味論, 発達心理学と**構成主義**, ユクスキュルの**環世界論**, ネオサイバネティクスと情報, プラグマティズム, 確率的生成モデル, 自由エネルギー原理と予測符号化, マルチモーダル物体概念形成, マルチモーダル場所概念形成, ディープラーニングと表現学習, 世界モデル, 大規模言語モデルと分布意味論, 認知発達ロボティクス, ニューロロボティクス, 身体性とソフトロボティクス, 幼児の言語獲得, ロボットによる語彙獲得, 感情と好奇心, 意識とクオリア, 言語の進化と創発, 現象学, エナクティヴィズム, 文化心理学と記号圏, マルチモーダルな言語教育, 創発する倫理, コードの創発, AIロボット社会, 記号創発システム論の展望, など

Symbol emergence systems [Taniguchi+ 2016]
記号創発システム

個体による
表現学習



Tadahiro Taniguchi, Takayuki Nagai, Tomoaki Nakamura, Naoto Iwahashi, Tetsuya Ogata, and Hideki Asoh, Symbol Emergence in Robotics: A Survey, Advanced Robotics, 30(11-12) pp.706-728, 2016. DOI:10.1080/01691864.2016.1164622

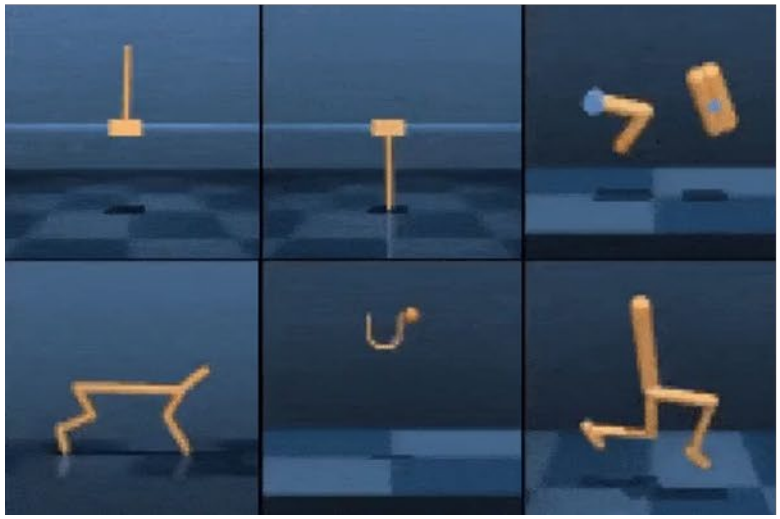
認知発達／記号創発ロボティクス とフィジカルAGI

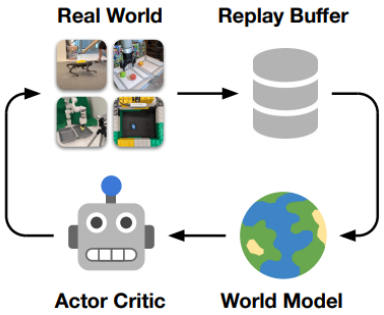
Francis Vachon, Time lapse of a baby playing with his toys
https://www.youtube.com/watch?v=8vNxjwt2AqY

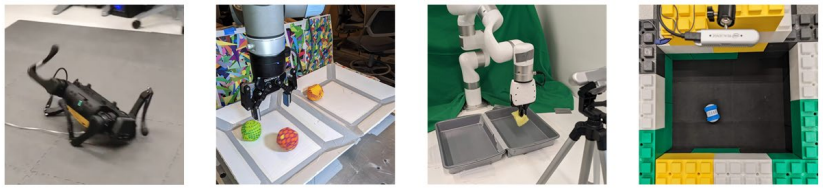
アンジェロ・カンジェロシ，他(著), 岡田浩之，谷口忠大(監訳), 発達ロボティクスハンドブック, 福村出版, 2019

谷淳,ロボットに心は生まれるか 自己組織化する動的現象としての行動・シンボル・意識 福森出版, 2022

世界モデルによるロボットの環境適応と動作学習






(a) A1 Quadruped Walking (b) UR5 Visual Pick Place (c) XArm Visual Pick Place (d) Sphero Navigation

PlaNet [Hafner+ 2019] :
Learning Latent Dynamics for Planning from Pixels

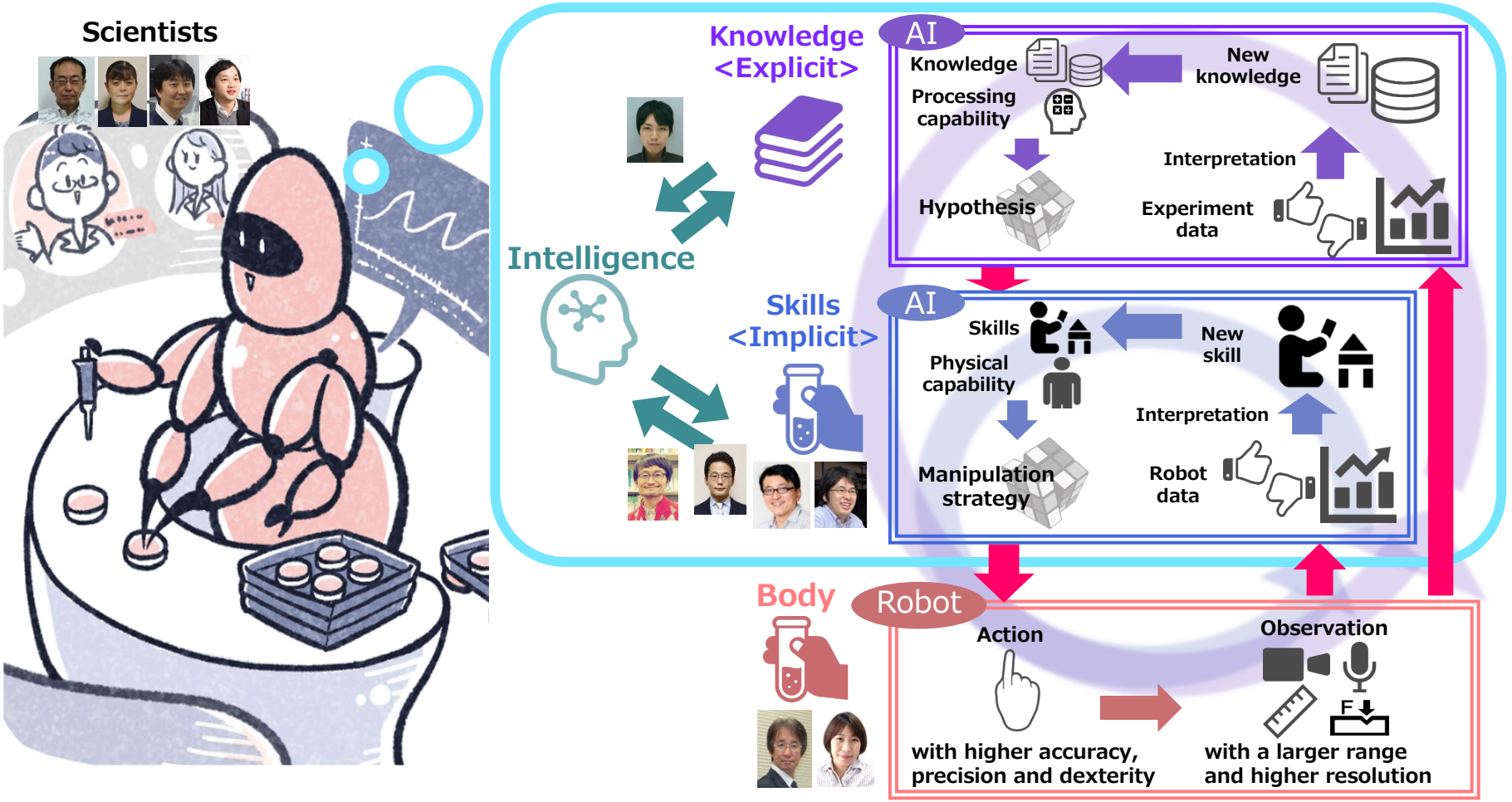
DayDreamer [Wu+ 2022] :
World Models for Physical Robot Learning

世界モデル（world model）
画像等の実センサ情報の予測モデルを学習し，潜在変数の表現学習とダイナミクス学習を行い，潜在空間上で効率的な強化学習や模倣学習を行う.

Hafner, Danijar, et al. “Learning latent dynamics for planning from pixels.” International conference on machine learning. PMLR, 2019.
Wu, P., Escontrela, A., Hafner, D., Goldberg, K., & Abbeel, P. (2022). Daydreamer: World models for physical robot learning. *arXiv:2206.14176*.

ムーンショット型研究開発事業 目標3
人とAIロボットの創造的共進化によるサイエンス開拓

プロジェクトマネージャー(PM)原田 香奈子（東京大学）



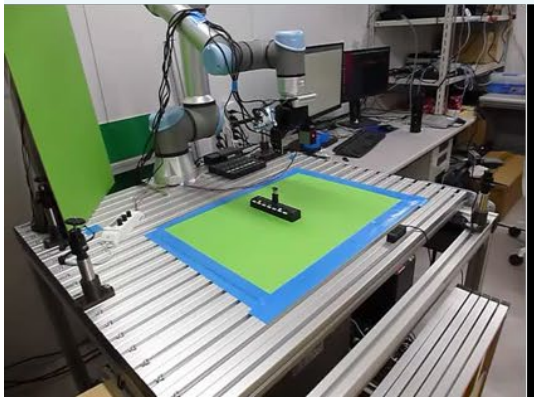
目標3 研究開発プロジェクト（2020年度採択）人とAIロボットの創造的共進化によるサイエンス開拓

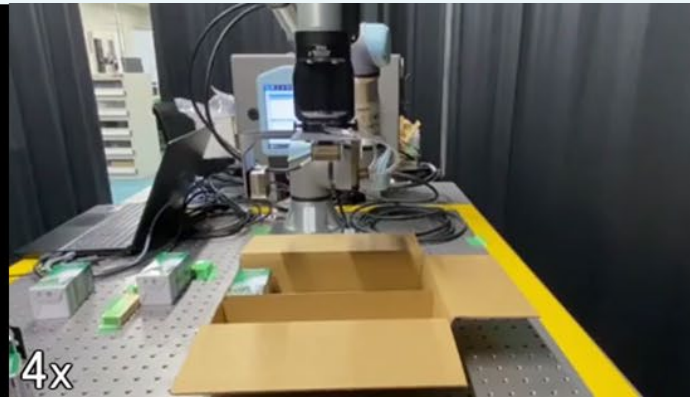
https://www.jst.go.jp/moonshot/program/goal3/33_harada.html

科学的発見のための自律ロボット

世界モデル／模倣学習に基づく科学実験代替AIロボット

@ムーンショット目標３原田香奈子PJ「人とAIロボットの創造的共進化によるサイエンス開拓」&パナソニック共同研究（クロスアポイントメント）





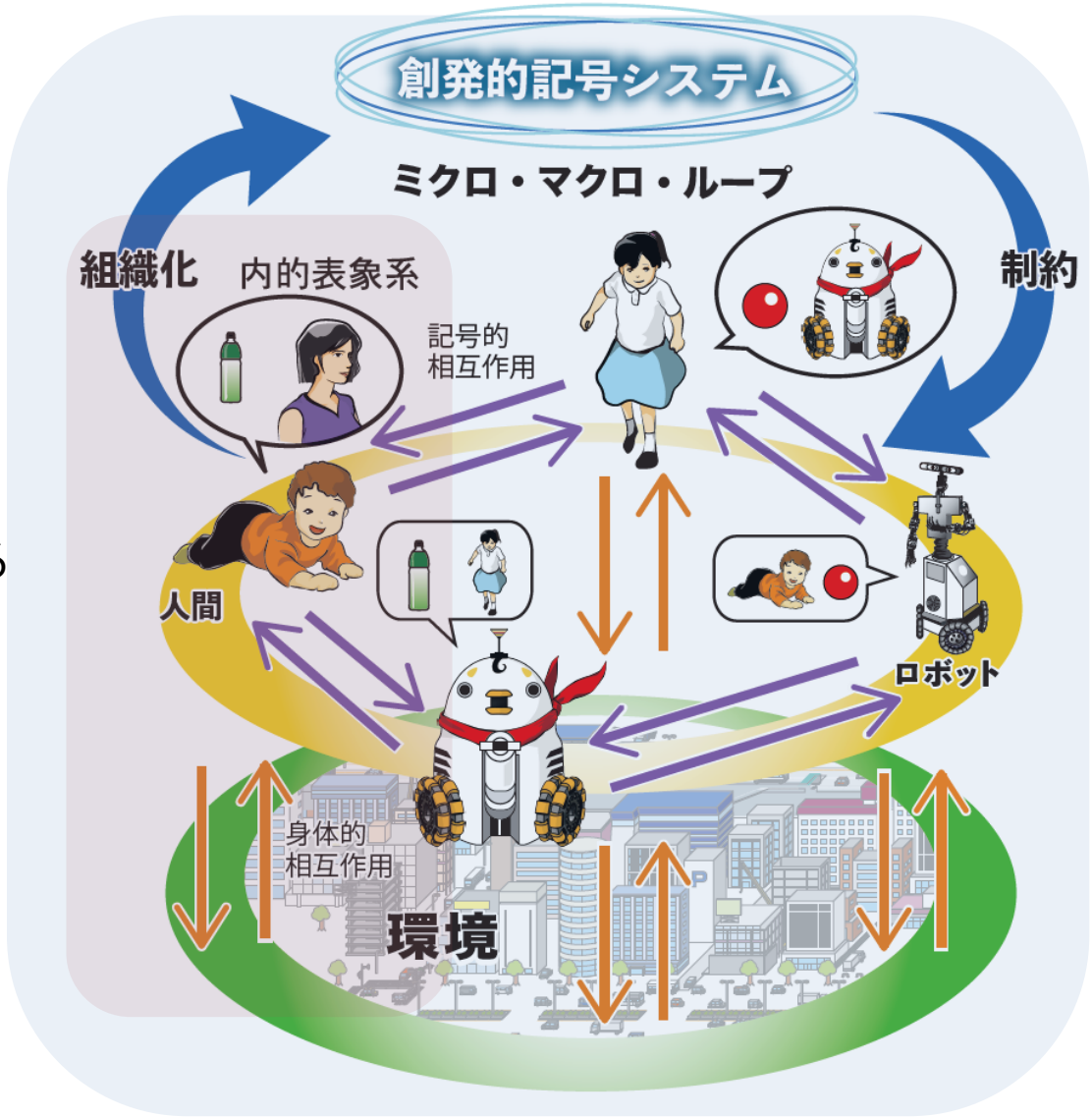
Tactile-Sensitive NewtonianVAE [Okumura+ 2022]
NewtonianVAE [Jaques+ 2021, CVPR]
ニュートン力学の構造を制約にした世界モデルを学習．潜在空間でのPID制御を可能に．触覚ビジョンセンサの表現学習を統合してより微細な制御を実現

Haptic Action Chunking Transformer [Uriguen + 2025 (under review)]
Transformerに基づく模倣学習アーキテクチャであるACTに、複数視点の画像情報に加えて力覚（触覚）情報を活用するように拡張。力覚情報を自動活用し動作復帰可能に。

Box packing [Kato+ 2023]
系列タスクへの実ロボットでの拡張

- ✓ Okumura, et at., “Tactile-Sensitive NewtonianVAE for High-Accuracy Industrial Connector-Socket Insertion.” IROS 2022
- ✓ Yusuke Kato, et al., World-Model-Based Control for Industrial box-packing of Multiple Objects using NewtonianVAE, Workshop on World Models and Predictive Coding in Cognitive Robotics, IROS 2023, Cognitive Robotics Award (Best Paper Award)

Symbol emergence systems [Taniguchi+ 2016]
記号創発システム



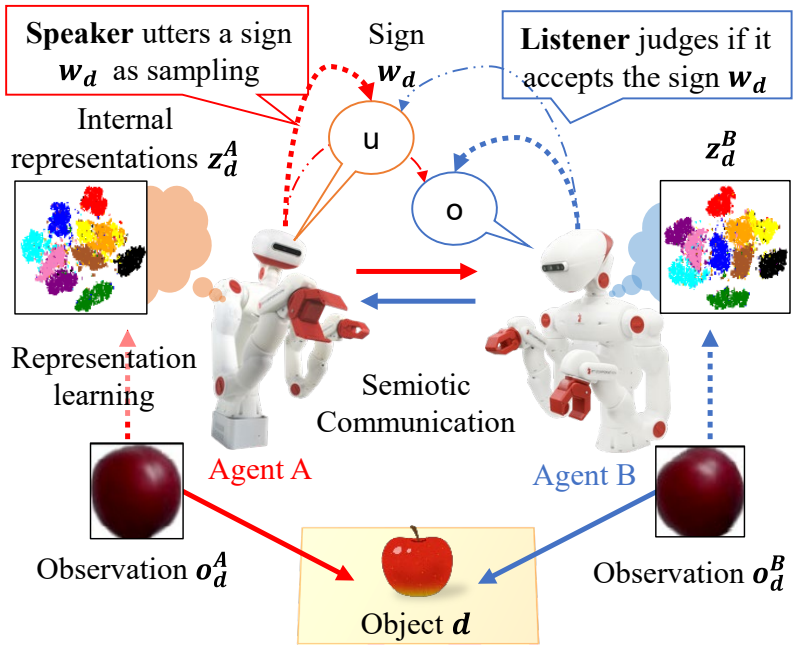
個体による
表現学習

集団による
記号創発

Tadahiro Taniguchi, Takayuki Nagai, Tomoaki Nakamura, Naoto Iwahashi, Tetsuya Ogata, and Hideki Asoh, Symbol Emergence in Robotics: A Survey, Advanced Robotics, 30(11-12) pp.706-728, 2016. DOI:10.1080/01691864.2016.1164622

生成的コミュニケーション創発（Generative EmCom）
メトロポリスヘイスティングス(MH)名付けゲーム
Metropolis-Hastings naming game

1. **Perception**: SpeakerとListenerのエージェント（SpとLi）がd番目の対象（Object）を観察し，内部表現（内的表象）を推論する（**共同注意**を仮定）．
2. **Communication**: Speakerは自らの信念状態にもとづき確率的に対象の名前を発話（サンプリング）する．Listenerは自らの信念状態に応じた確率でその名付けを受け入れるかどうかを決定する．
3. **Learning**:コミュニケーションの後，Listenerは，表現学習と名付けのための内部パラメータを更新する．
4. **Turn taking**: SpeakerとListenerが役割を交代し，1へ戻る．

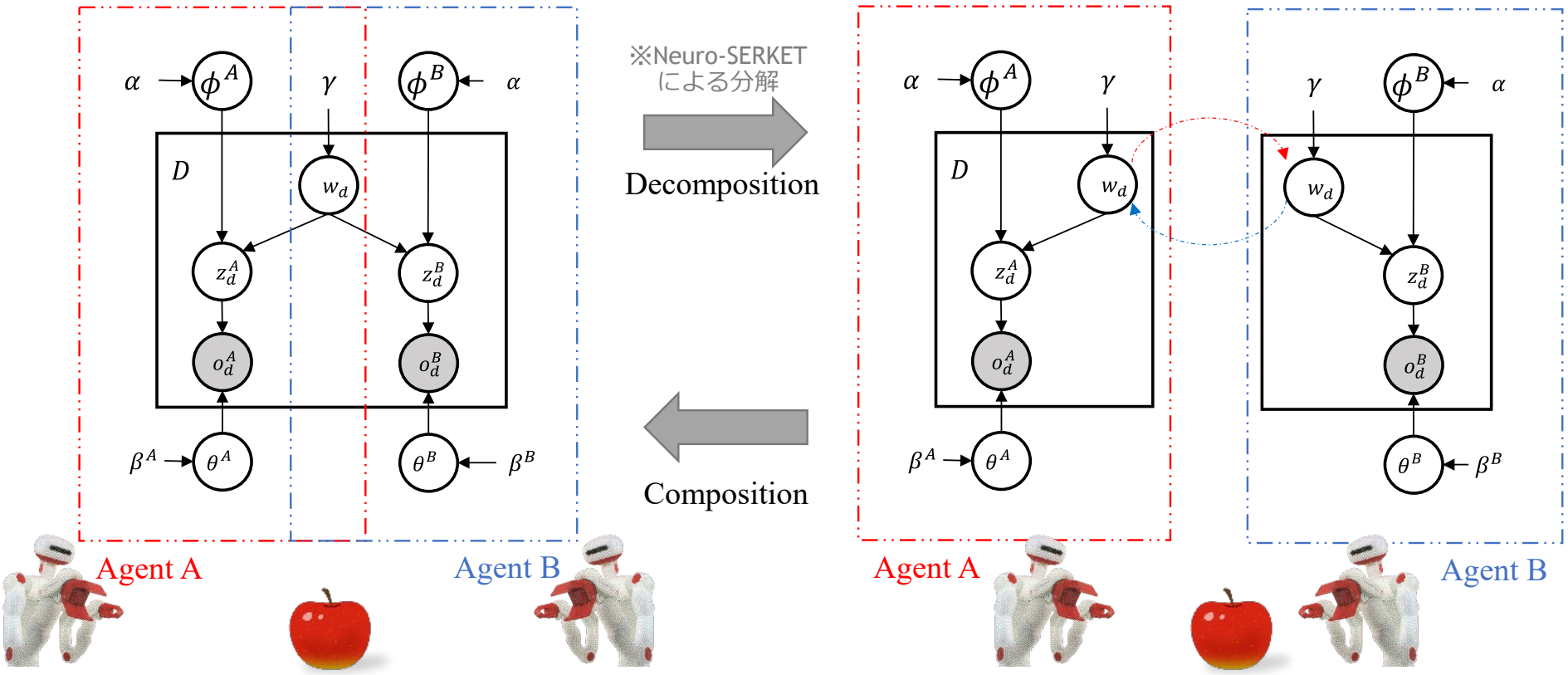


$$r = \min\left(1, \frac{P(z_d^{Li}|\theta^{Li}, w_d^{Sp})}{P(z_d^{Li}|\theta^{Li}, w_d^{Li})}\right)$$

※ 相手の名前と自分の想定していた名前が自らの信念にどれだけ一致するかの比率

- Yoshinobu Hagiwara , Hiroyoshi Kobayashi, Akira Taniguchi and **Tadahiro Taniguchi**, Symbol Emergence as an Interpersonal Multimodal Categorization, Frontiers in Robotics and AI, 6(134), pp.1-17, 2019
- Yoshinobu Hagiwara, Kazuma Furukawa, Akira Taniguchi & **Tadahiro Taniguchi**, Multiagent multimodal categorization for symbol emergence: emergent communication via interpersonal cross-modal inference, Advanced Robotics, 2022.

記号創発システムの確率的グラフィカルモデル



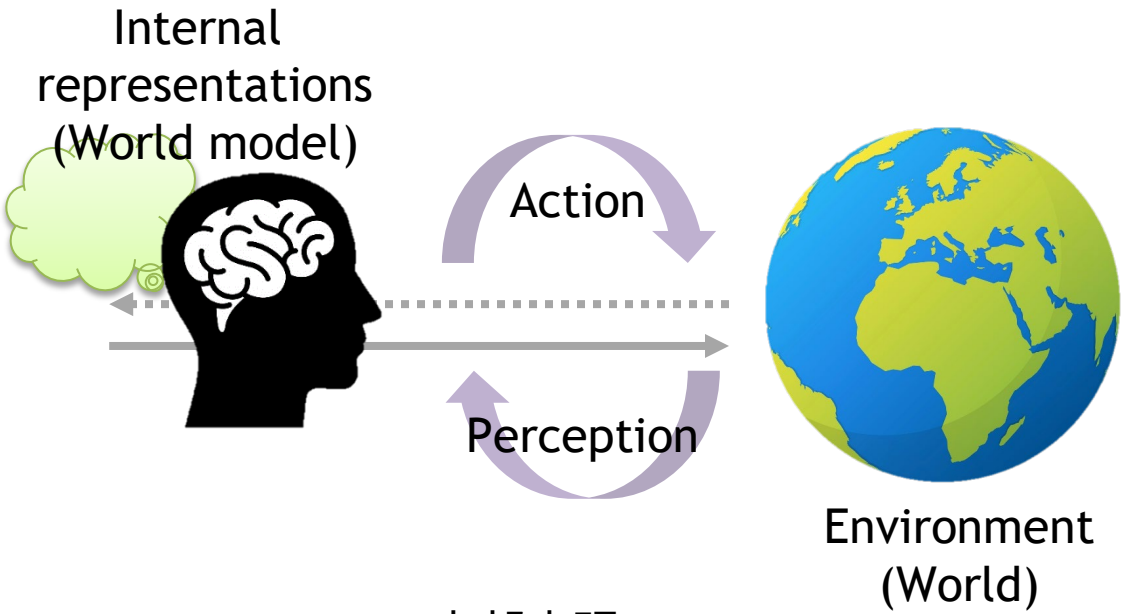
Theorem 1. *MH naming game is a Metropolis-Hastings sampler of* $P(w, z, \theta, \phi|o)$.

MH名付けゲームは分散型MCMCベイズ推論になる

私たちは言語を生み出すことで「脳をつなぐ」のと同等の認識統合を行える❓

Taniguchi, T., Yoshida, Y., Matsui, Y., Le Hoang, N., Taniguchi, A., & Hagiwara, Y. (2023). Emergent communication through Metropolis-Hastings naming game with deep generative models. *Advanced Robotics*, *37*(19), 1266-1282.

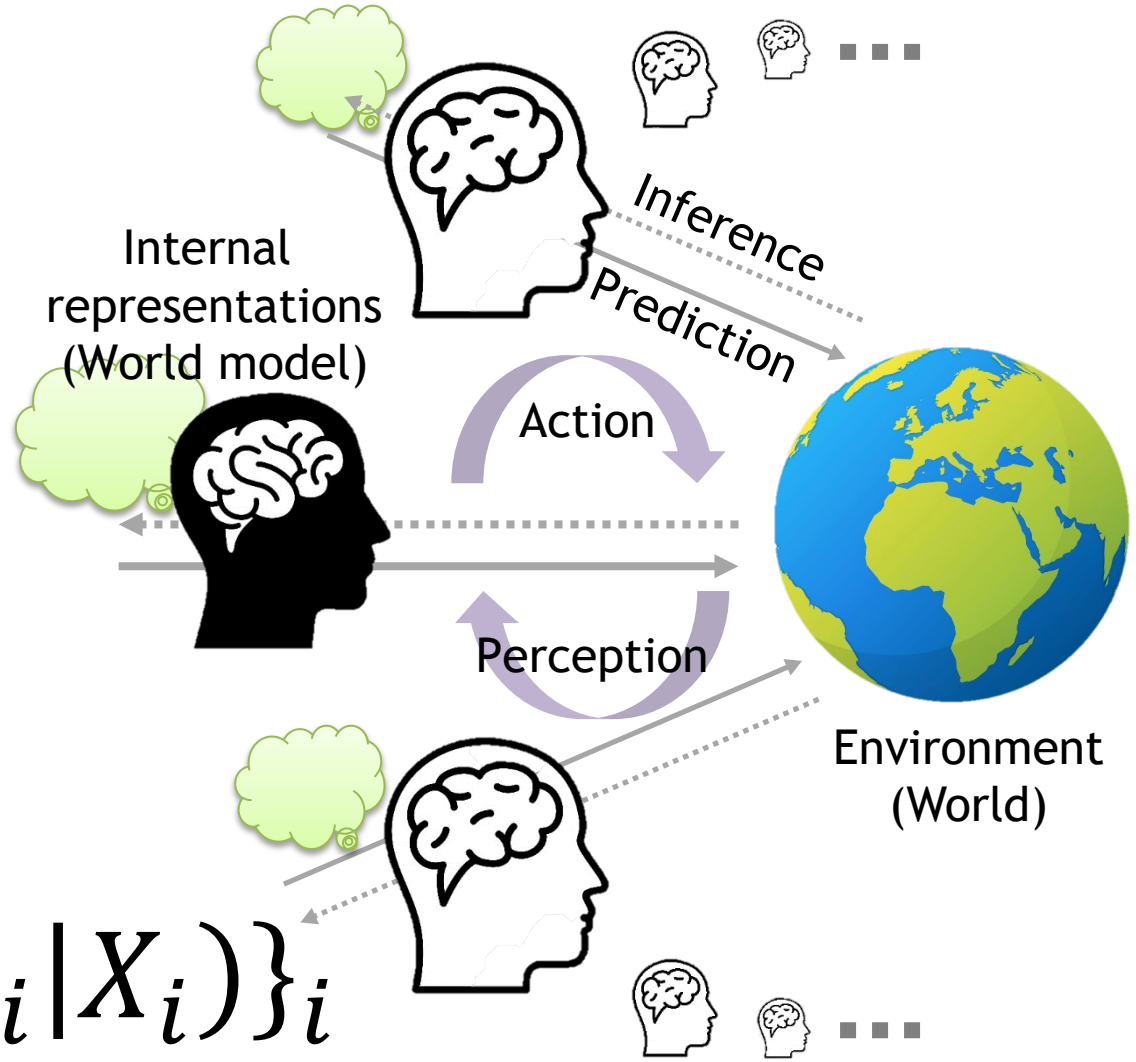
予測符号化（世界モデル学習）から 集合的予測符号化（記号創発）へ



内部表現　観測

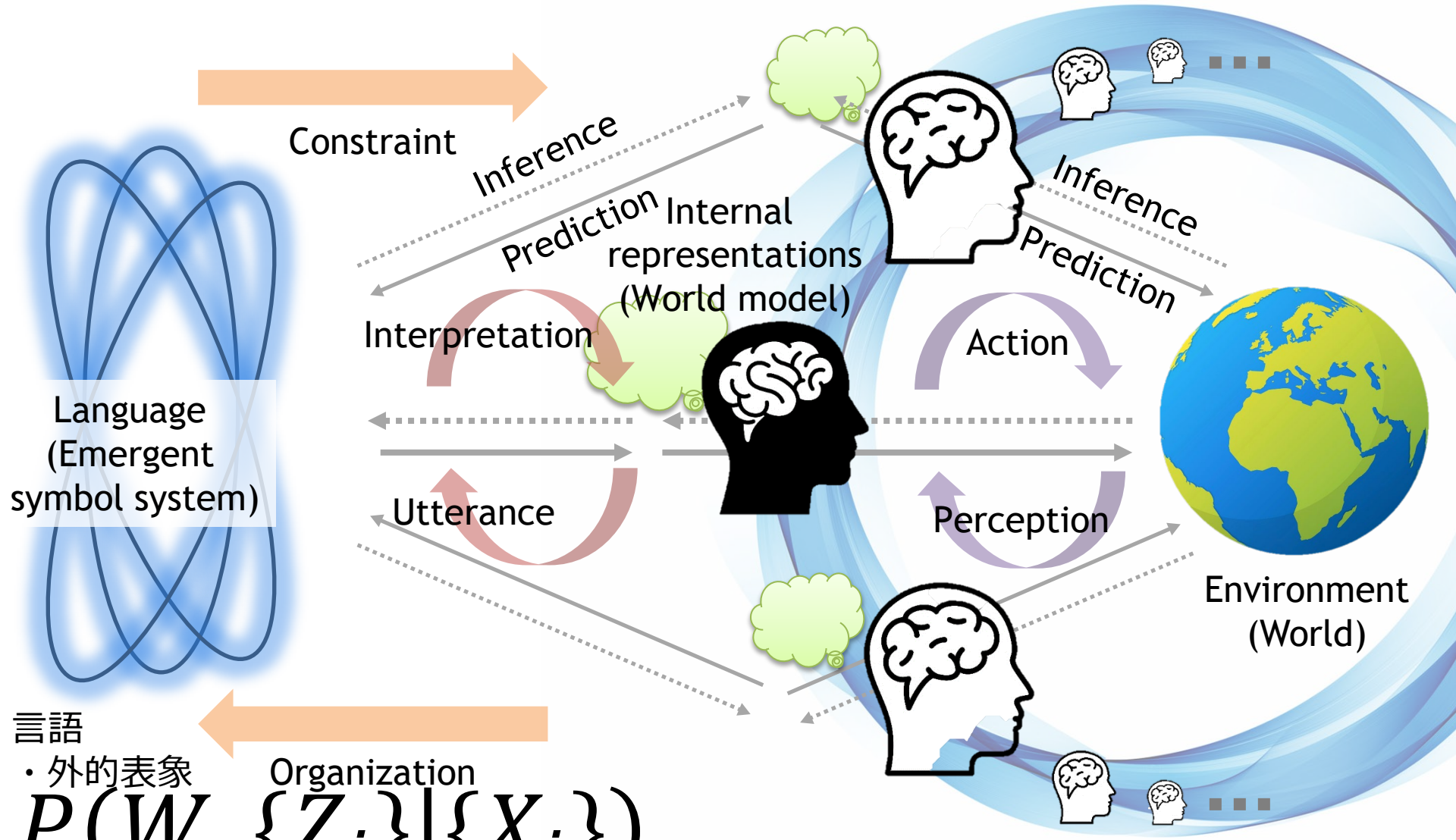
$$P(Z_i|X_i)$$

予測符号化（世界モデル学習）から
集合的予測符号化（記号創発）へ



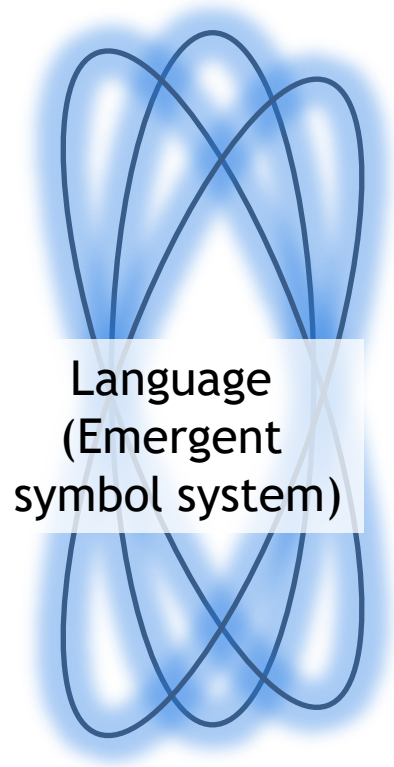
$$\{P(Z_i|X_i)\}_i$$

予測符号化（世界モデル学習）から
集合的予測符号化（記号創発）へ

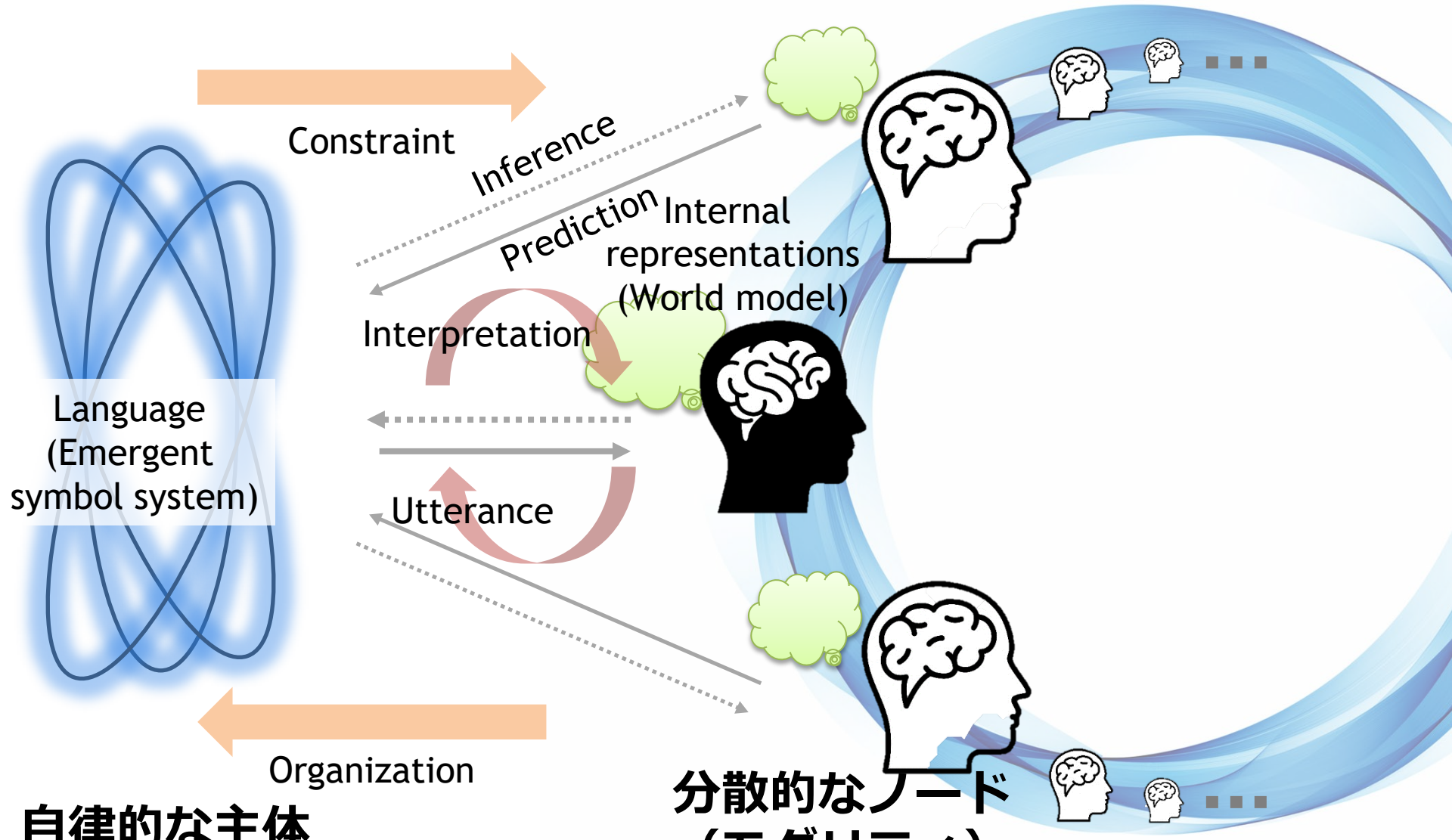


$$P(W, \{Z_i\}|\{X_i\})$$

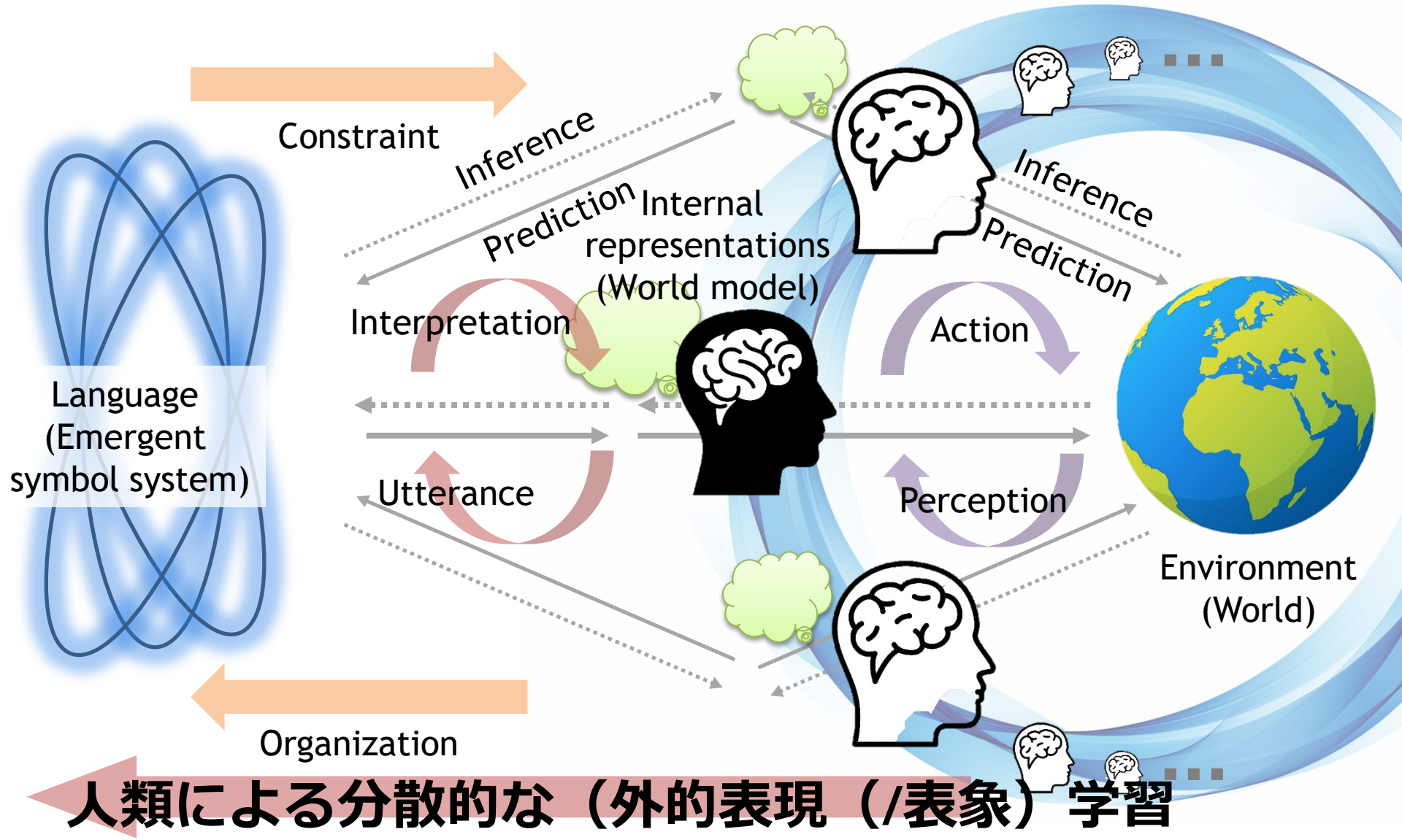
予測符号化（世界モデル学習）から
集合的予測符号化（記号創発）へ



予測符号化（世界モデル学習）から
集合的予測符号化（記号創発）へ



予測符号化（世界モデル学習）から
集合的予測符号化（記号創発）へ



人類による分散的な（外的表現（/表象）学習

自由エネルギー原理に基づくCPC-MSの定式化

Generative model: $p(w, \mathbf{z}, \mathbf{o} \mid \mathbf{a}) = p(w)p(\mathbf{o} \mid \mathbf{z}, \mathbf{a})\, p(\mathbf{z} \mid w, \mathbf{a})$

Inference model: $q(w, \mathbf{z}, \mathbf{o} \mid \mathbf{a}) = q(w \mid \mathbf{z})\, q(\mathbf{o} \mid)\, q(\mathbf{z} \mid \mathbf{o}, \mathbf{a})$

$$F = D_{\mathrm{KL}}\left[q(w, \mathbf{z}, \mathbf{o} \mid \mathbf{a}) \| p(w, \mathbf{z}, \mathbf{o} \mid \mathbf{a})\right]$$

$$= \int q(w, \mathbf{z}, \mathbf{o} \mid \mathbf{a}) \ln \frac{q(w, \mathbf{z}, \mathbf{o} \mid \mathbf{a})}{p(w, \mathbf{z}, \mathbf{o} \mid \mathbf{a})}\, dw d\mathbf{z} d\mathbf{o}$$

$$= \mathbb{E}_q\left[\ln \frac{q(w \mid \mathbf{z})}{p(w)}\right] + \mathbb{E}_q\left[\ln \frac{q(\mathbf{o})}{p(\mathbf{o} \mid \mathbf{z}, \mathbf{a})}\right] + \mathbb{E}_q\left[\ln \frac{q(\mathbf{z} \mid \mathbf{o}, \mathbf{a})}{p(\mathbf{z} \mid w, \mathbf{a})}\right]$$

$$= \underbrace{D_{\mathrm{KL}}\left[q(w \mid \{z^k\}_k) \| p(w)\right]}_{\text{Collective Regularization}} +$$

$$\sum_k \left[\underbrace{\mathbb{E}_q\left[-\ln p\left(o^k \mid z^k, a^k\right)\right]}_{\text{Individual prediction error}} + \underbrace{D_{\mathrm{KL}}\left[q(z^k \mid o^k, a^k) \| p(z^k \mid w, a^k)\right]}_{\text{Individual Regularization}} \right]$$

2. Collective regularization term
(Alignment of external representation w conditioned by internal representation z and, symbol emergence)

1. Ordinary variational free energy × Number of agents
(Representation learning, predictive coding, world model learning)

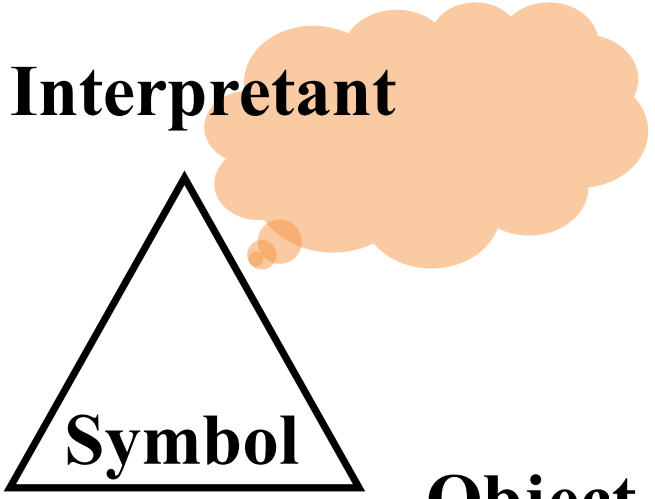
CPCを変分自由エネルギー最小化として定式化すると集合的正則化項が出現する
科学の探究活動はコミュニティにおける集合的自由エネルギーを下げること？

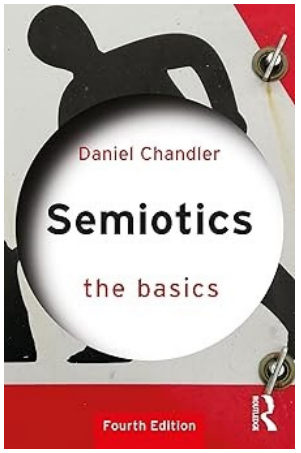
https://x.com/hayashiyus/status/1831309992638759210

Taniguchi, T., Takagi, S., Otsuka, J., Hayashi, Y., & Hamada, H. T. (2025). Collective predictive coding as model of science: Formalizing scientific activities towards generative science. Royal Society Open Science (in press) (*arXiv preprint arXiv:2409.00102*).

Peirce's semiotics and arbitrariness

C.S. Peirce
Father of Semiotics
Thinker of Pragmatism
https://en.wikipedia.org/wiki/Charles_Sanders_Peirce



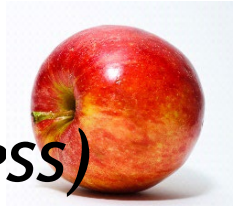


"Semiotics: The Basics"
Daniel Chandler

$$\underbrace{\mathbb{E}_q\left[\ln q\left(w \mid \{z^k\}_k\right)\right]}_{\text{Collective regularization}}$$

Apple

Semiosis
(Semiotic process)

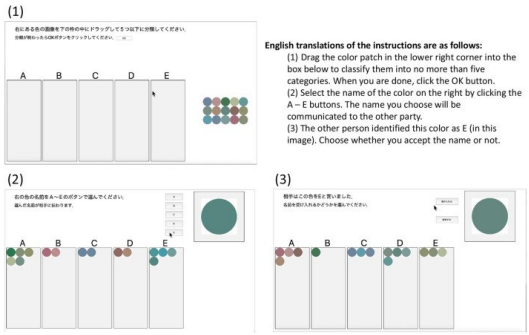


個人の脳の可塑性に依存しながら、記号体系の可塑性を活用することで、私たちは集団的な社会知性を獲得しているのではないか？
――神経可塑性（neural plasticity）から記号可塑性（semiotic plasticity）へ

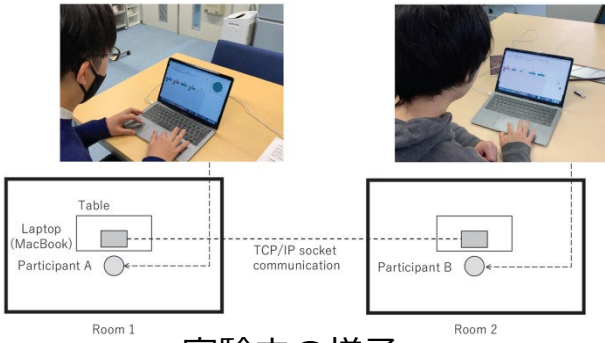
共同注意名付けゲームにおけるMetropolis-Hastingsアルゴリズムの妥当性に関する実験記号的研究 [Okumura+ 2023]

人間が**MH法に基づく受容確率**で名付け受け入れるのかを調べるために，
共同注意名付けゲームを人間に行わせ分析した

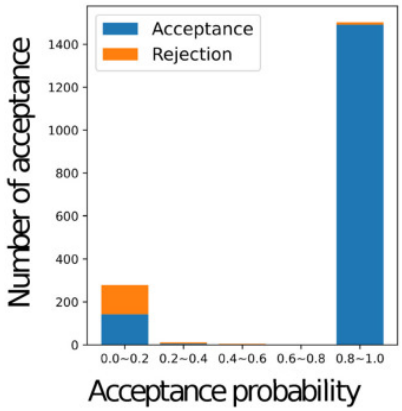
- **MH法**に基づく受容確率が高い程**被験者**の実際の受容割合も高い傾向があった
- 統計的**検定**によりMH法を用いたモデルが比較モデルより人間の行動を予測することを示した

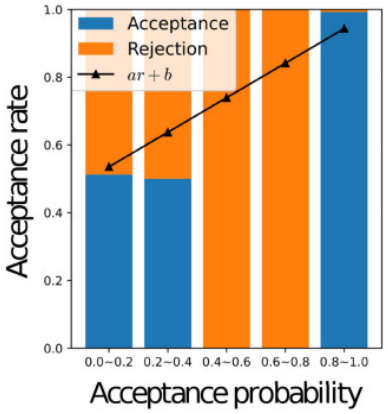


実験中の操作画面



実験中の様子





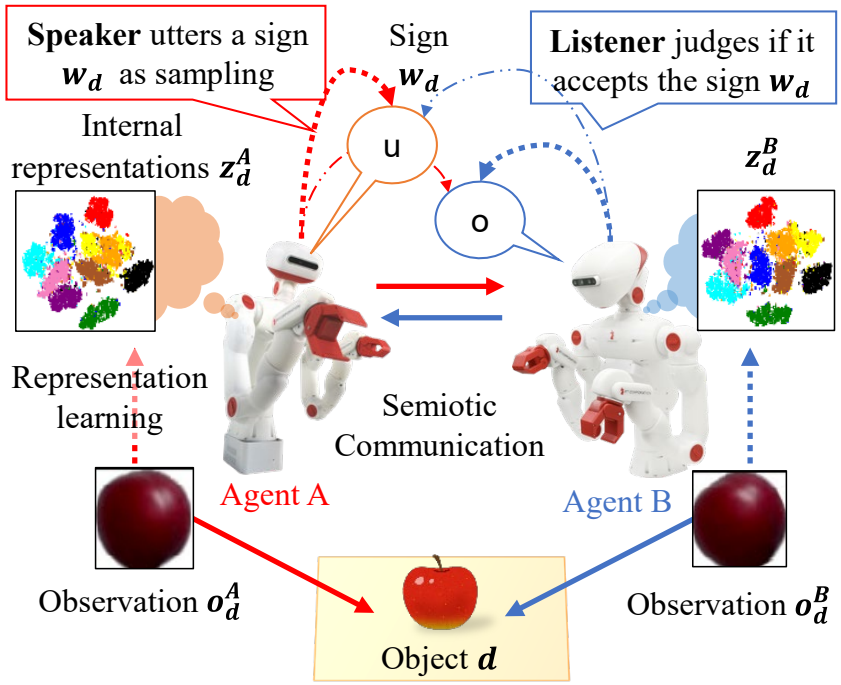
モデルm'と比較してmの方が有意に判断を予測できた被験者の数の表

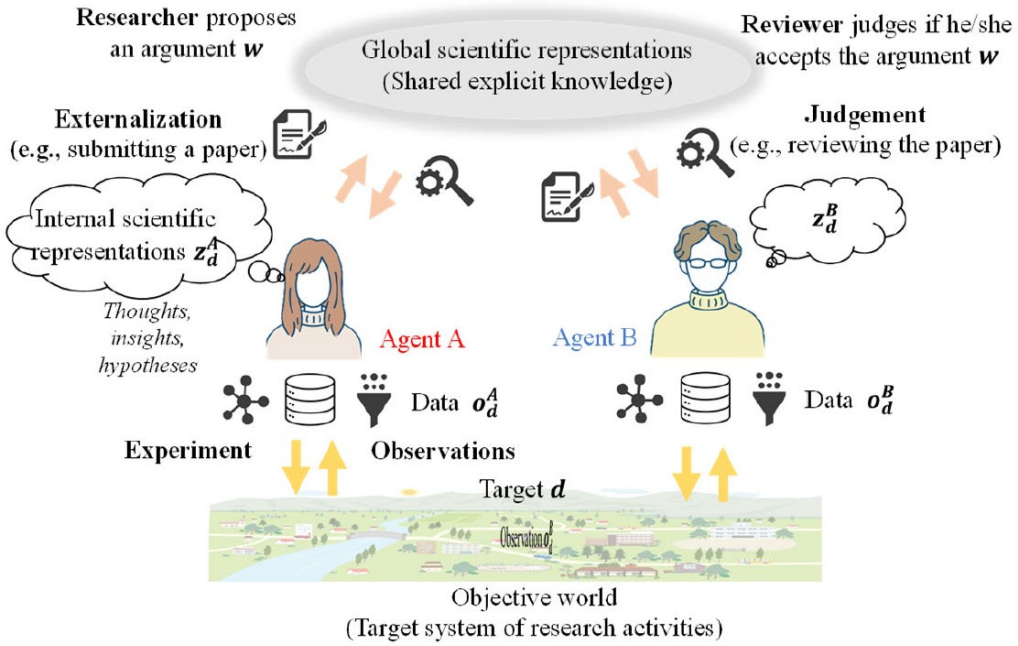
$m \backslash m'$	Constant	MH	Numerator	Subtraction	Binary
Constant	—	4	12	18	12
MH	14	—	20	19	20
Numerator	6	0	—	20	7
Subtraction	2	0	0	—	0
Binary	6	0	1	20	—

MH法の受容確率は人間の受容判断を相対的によく予測できていた

Okumura, Ryota, Tadahiro Taniguchi, Yoshinobu Hagiwara, and Akira Taniguchi. "Metropolis-Hastings algorithm in joint-attention naming game: Experimental semiotics study." *Frontiers in Artificial Intelligence* 6 (2023)

MH naming game and scientific discussion



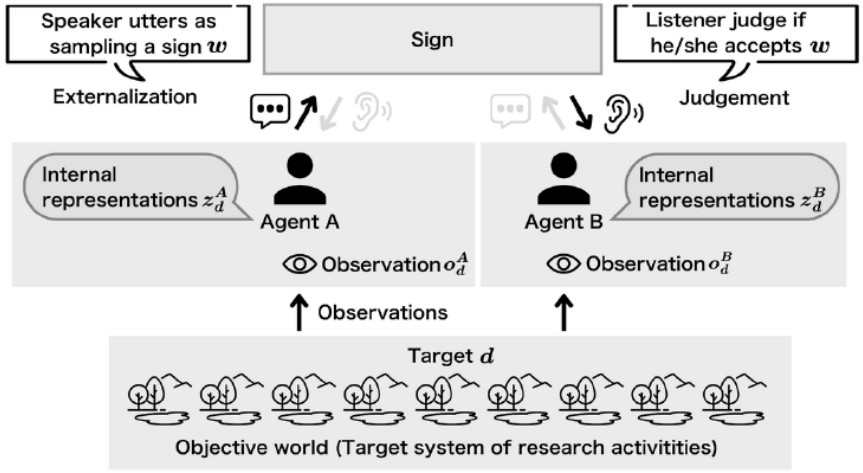


- 科学的議論、特に査読プロセスは、構造的にMetropolis-Hastings名付けゲーム(MHNG)と類似している
- MHNGは分散的なベイズ推論として、外的表象系への表現学習の推論アルゴリズムとして機能する。
- 科学活動全体のプロセスは、MHネーミングゲームと同様に、分散ベイズ推論として捉えられる可能性がある
- これが、科学のモデルとしてのCP（CPC-MS）という発想につながる

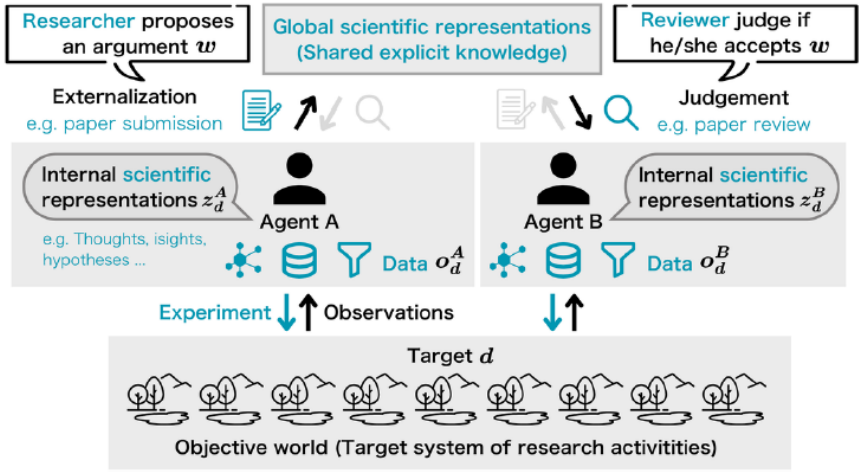
科学のモデルとしての集合的予測符号化

Collective Predictive Coding as Model of Science

- 科学を、集合的予測符号化をする記号創発システムとして捉える、科学のモデル (CPC-MS)



Collective Predictive Coding



CPC-MS

Taniguchi, T., Takagi, S., Otsuka, J., Hayashi, Y., & Hamada, H. T. (2025). Collective predictive coding as model of science: Formalizing scientific activities towards generative science. Royal Society Open Science (in press) (*arXiv preprint arXiv:2409.00102*).

The AI Scientist [Lu+ 2024]

The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery

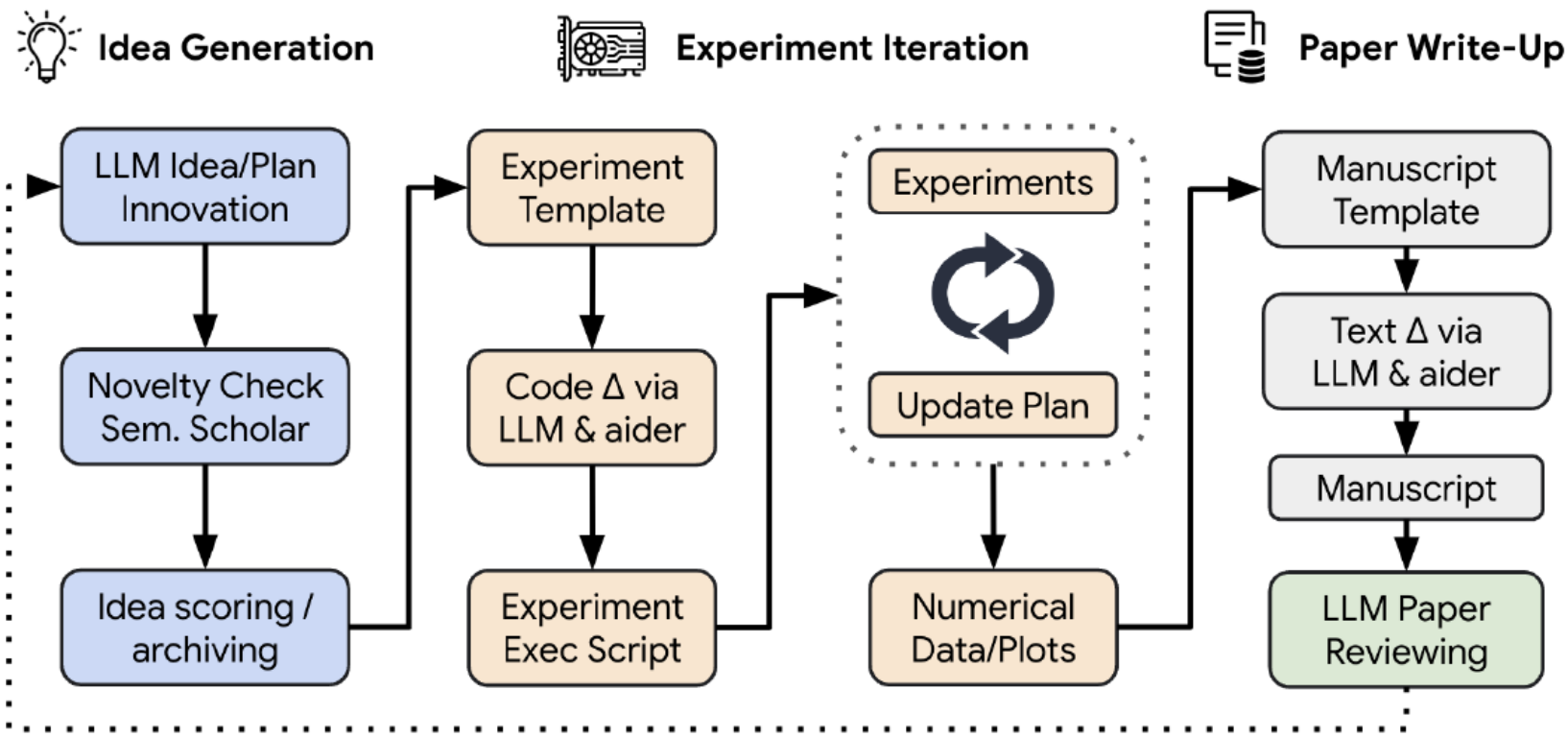
Chris Lu[1,2,*], Cong Lu[3,4,*], Robert Tjarko Lange[1,*], Jakob Foerster[2,†], Jeff Clune[3,4,5,†] and David Ha[1,†]
[*]Equal Contribution, [1]Sakana AI, [2]FLAIR, University of Oxford, [3]University of British Columbia, [4]Vector Institute, [5]Canada CIFAR AI Chair, [†]Equal Advising

One of the grand challenges of artificial general intelligence is developing agents capable of conducting scientific research and discovering new knowledge. While frontier models have already been used as aides to human scientists, e.g. for brainstorming ideas, writing code, or prediction tasks, they still conduct only a small part of the scientific process. This paper presents the first comprehensive framework for fully *automatic scientific discovery*, enabling frontier large language models (LLMs) to perform research independently and communicate their findings. We introduce THE AI SCIENTIST, which generates novel research ideas, writes code, executes experiments, visualizes results, describes its findings by writing a full scientific paper, and then runs a simulated review process for evaluation. In principle, this process can be repeated to iteratively develop ideas in an open-ended fashion and add them to a growing archive of knowledge, acting like the human scientific community. We demonstrate the versatility of this approach by applying it to three distinct subfields of machine learning: diffusion modeling, transformer-based language modeling, and learning dynamics. Each idea is implemented and developed into a full paper at a meager cost of less than $15 per paper, illustrating the potential for our framework to democratize research and significantly accelerate scientific progress. To evaluate the generated papers, we design and validate an automated reviewer, which we show achieves near-human performance in evaluating paper scores. THE AI SCIENTIST can produce papers that exceed the acceptance threshold at a top machine learning conference as judged by our automated reviewer. This approach signifies the beginning of a new era in scientific discovery in machine learning: bringing the transformative benefits of AI agents to the *entire* research process of AI itself, and taking us closer to a world where *endless affordable creativity and innovation* can be unleashed on the world's most challenging problems. Our code is open-sourced at https://github.com/SakanaAI/AI-Scientist.

Lu et al. (2024) The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery

Lu et al. (2024) The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery

「研究者」の一連のプロセスを代替する生成AIの出現！？



An AI Scientist that conducts every step from idea generation, experimentation, paper-writeup to peer review

Lu et al. (2024) The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery

生成科学からの道

① 「研究者」の生成AIによる代替は、AIによる**「科学」の代替**を意味するのか？

- ✓ **No**。科学とは個人の営みに非ず。人類による科学的探求活動の総体をモデル化できているわけではない。

② 生成AIの濫用は既存の科学コミュニティを破壊するか？

- ✓ **Maybe**。生成AI論文の大量投稿。レビュー体制の破壊。

③ 生成AIの活用やロボットの導入は科学全体をAIによる**自律駆動科学**システムに変えるか？

- ✓ **No**。科学活動の総体において、それは部品の**パッチワーク的**な置き換えにすぎない。

④ 生成AIの活用は「科学」の全体を良き方向に導くか？

- ✓ **Maybe, but Nobody knows**。だれも見通せていないように思う。良い「科学」とはなにか？

生成科学とは何か？

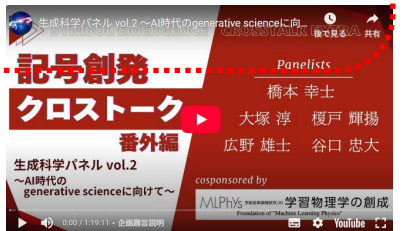
A) **広義の生成AI活用による科学の加速**
- 生成AIをリサーチツールとして活用し、論文執筆支援やデータ可視化、実験プロトコル自動生成などを通じて研究プロセス全般の生産性を向上させるアプローチ。

B) **仮説生成を含む閉ループ自律科学**
- 文献レビューから仮説立案、実験設計・実行、データ解析、結果フィードバックによる再仮説生成まで、AIが一連のサイクルを自動で回すことで科学的発見を推進する方法。

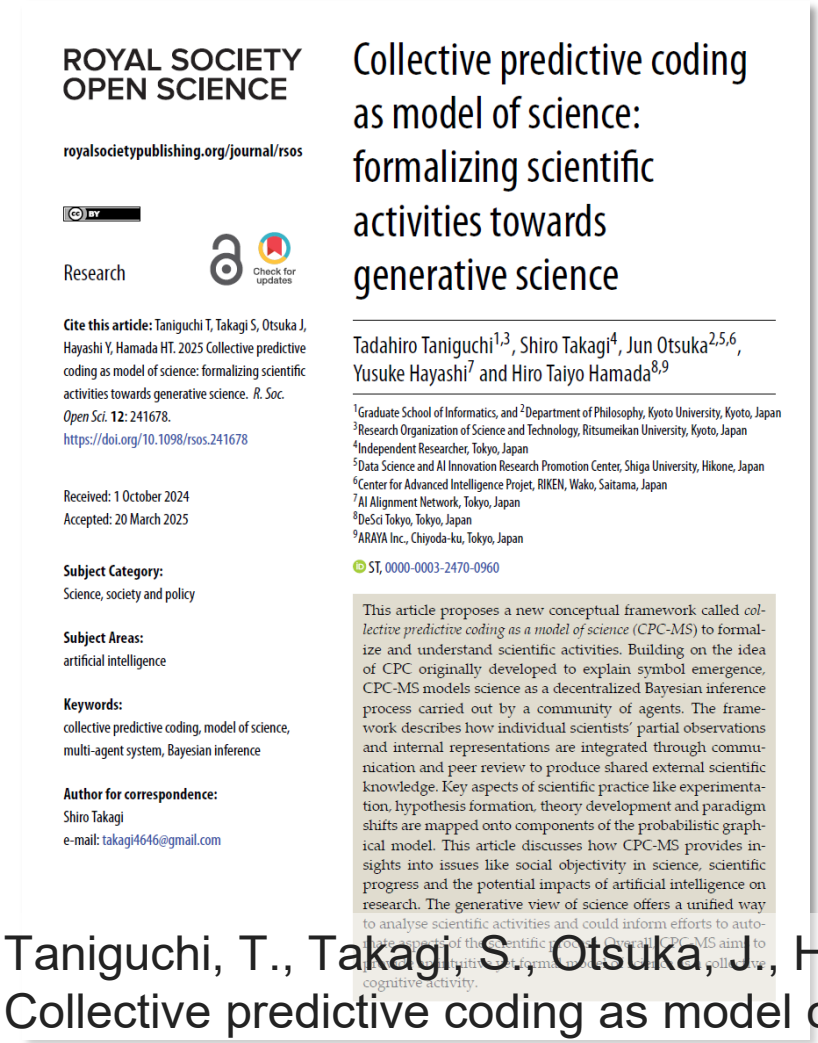
C) **構成論的生成科学**
- 計算機シミュレーションやジェネレーティブモデルを用い、人工生命や複雑系の「新規事象」を構成・生成しながら知見を得る、従来の分析的科学に対置される構成的アプローチ。

D) **科学コミュニティの確率的生成モデル化**（CPC的視点）
- 研究者の観察・仮説・実験・査読・引用などコミュニティ全体の活動をベイズ的生成過程として捉え、人間とAIを包含するハイブリッドな「生成科学」システムとしてモデル化する枠組み。

生成科学パネル vol.2 ～AI時代のgenerative scienceに向けて～｜記号創発クロストーク・番外編 https://www.youtube.com/watch?v=2uatQVyEhel



生成科学に向けて



ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Research

Cite this article: Taniguchi T, Takagi S, Otsuka J, Hayashi Y, Hamada HT. 2025 Collective predictive coding as model of science: formalizing scientific activities towards generative science. *R. Soc. Open Sci.* **12**: 241678.
https://doi.org/10.1098/rsos.241678

Received: 1 October 2024
Accepted: 20 March 2025

Subject Category:
Science, society and policy

Subject Areas:
artificial intelligence

Keywords:
collective predictive coding, model of science, multi-agent system, Bayesian inference

Author for correspondence:
Shiro Takagi
e-mail: takagi4646@gmail.com

Collective predictive coding as model of science: formalizing scientific activities towards generative science

Tadahiro Taniguchi[1,3], Shiro Takagi[4], Jun Otsuka[2,5,6], Yusuke Hayashi[7] and Hiro Taiyo Hamada[8,9]

[1]Graduate School of Informatics, and [2]Department of Philosophy, Kyoto University, Kyoto, Japan
[3]Research Organization of Science and Technology, Ritsumeikan University, Kyoto, Japan
[4]Independent Researcher, Tokyo, Japan
[5]Data Science and AI Innovation Research Promotion Center, Shiga University, Hikone, Japan
[6]Center for Advanced Intelligence Projet, RIKEN, Wako, Saitama, Japan
[7]AI Alignment Network, Tokyo, Japan
[8]DeSci Tokyo, Tokyo, Japan
[9]ARAYA Inc., Chiyoda-ku, Tokyo, Japan

ST, 0000-0003-2470-0960

This article proposes a new conceptual framework called *collective predictive coding as a model of science (CPC-MS)* to formalize and understand scientific activities. Building on the idea of CPC originally developed to explain symbol emergence, CPC-MS models science as a decentralized Bayesian inference process carried out by a community of agents. The framework describes how individual scientists' partial observations and internal representations are integrated through communication and peer review to produce shared external scientific knowledge. Key aspects of scientific practice like experimentation, hypothesis formation, theory development and paradigm shifts are mapped onto components of the probabilistic graphical model. This article discusses how CPC-MS provides insights into issues like social objectivity in science, scientific progress and the potential impacts of artificial intelligence on research. The generative view of science offers a unified way to analyse scientific activities and could inform efforts to automate aspects of the scientific process. Overall, CPC-MS aims to [illegible] cognitive activity.

Tadahiro Taniguchi
@tanichu

Jun Otsuka
@junotk_jp

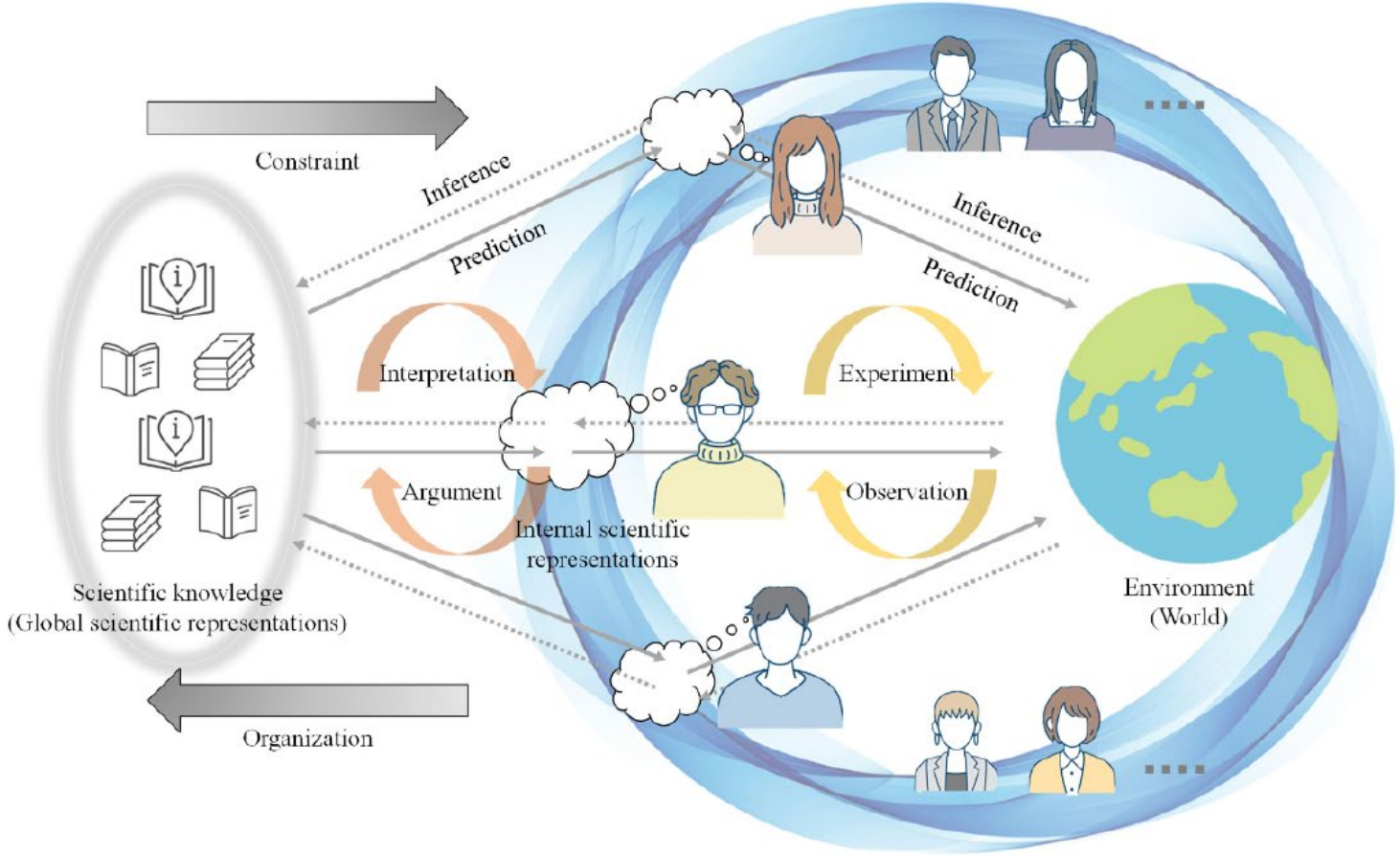
Yusuke Hayashi
@hayashiyus

Hiro Taiyo Hamada
@HiroTHamadaJP

Shiro Takagi
@takagi_shiro

Taniguchi, T., Takagi S., Otsuka, J., Hayashi, Y., & Hamada, H. T. (2025). Collective predictive coding as model of science: Formalizing scientific activities towards generative science. Royal Society Open Science (in press) (*arXiv preprint arXiv:2409.00102)*.

Collective Predictive Coding as Model of Science (CPC-MS): Formalizing Scientific Activities Towards Generative Science



科学を集合的予測符号化を行う記号創発システムとして解釈する

Collective Predictive Coding as Model of Science
高木志郎@第2回AIロボット駆動科学研究会

Table 1. Correspondences between mathematical notations in CPC, phenomena in scientific activities and language emergence.

mathematical notations	science activities	language emergence
external representations w	global scientific representation (e.g. published papers, established theories, consensus models)	shared symbolic system (e.g. words, sentences, signs)
internal representations z	internal scientific representations (e.g. hypotheses, insights, mental models, intuitions)	cognitive representations (e.g. concepts, mental images, perceptual state)
observations o	empirical data (e.g. experimental results, field observations, measurements)	sensory experiences (e.g. visual, auditory, tactile inputs)
inference of $P(z\|o)$	hypothesis formation and revision (e.g. data analysis, theory development)	representation learning (e.g. categorization, concept formation)
inference of $P(w\|z)$	scientific communication (e.g. paper writing, peer review, oral discussion)	language game (e.g. speech production, interpretation)

Implications by CPC-MS

① 社会的客観性（Social Objectivity）
- 科学的知識の客観性は個人の合意ではなく、科学コミュニティ全体の事後分布によって社会的に構成される。

② 科学の進歩（Scientific Progress）
- 科学の進歩は分散ベイズ推論による事後分布の逐次的改善として捉えられ、パラダイム転換も確率分布の変化として定式化される。

③ 生成的科学（Generative Science）
- 科学は既存知識の検証にとどまらず、新たな仮説や研究計画を生成し続ける営みとして理解される。

④ 集合的好奇心（Collective Curiosity）
- 科学者個人の好奇心が概念ネットワークを通じて相互に影響し合い、コミュニティ全体の探索行動を駆動する。

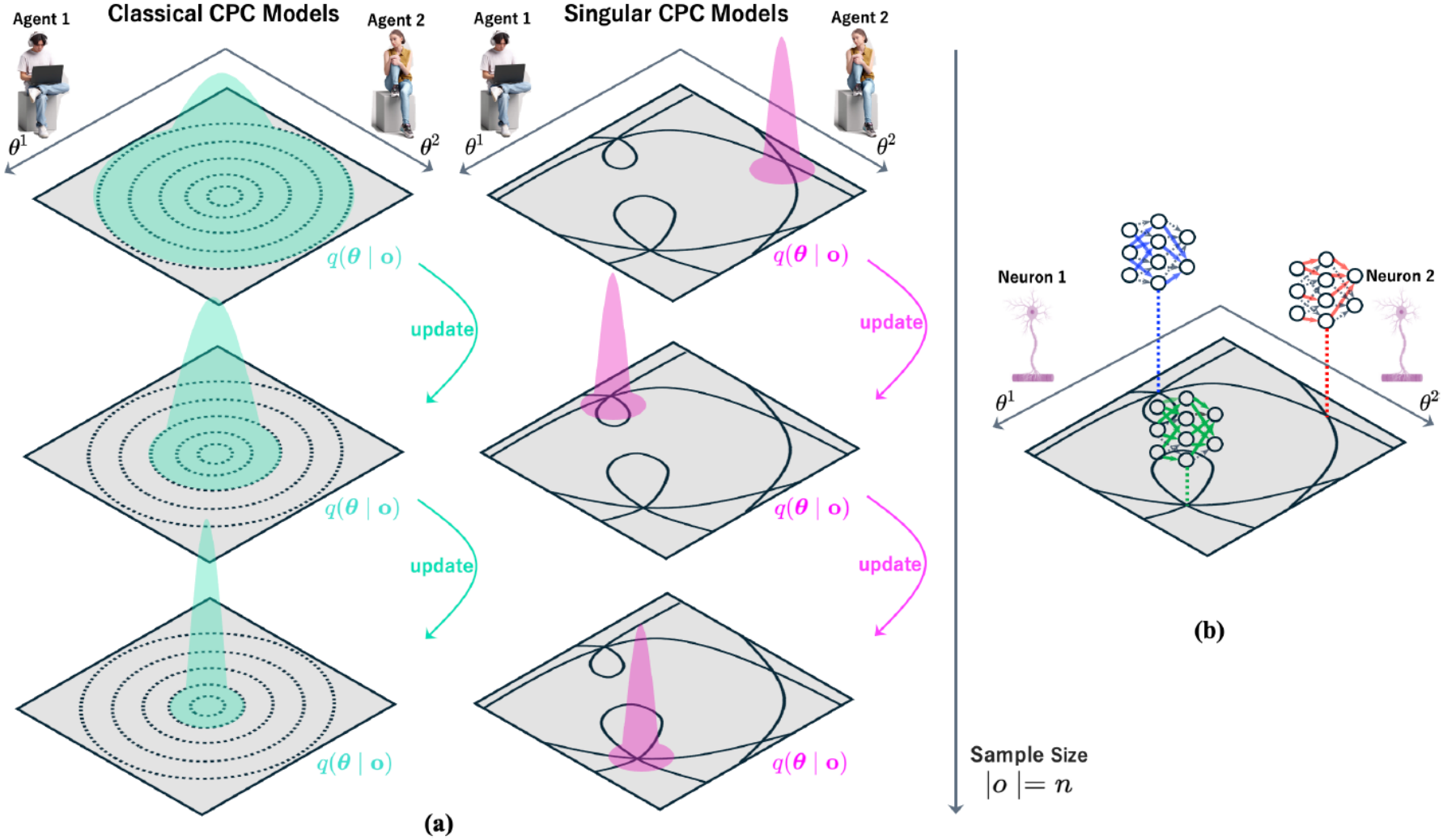
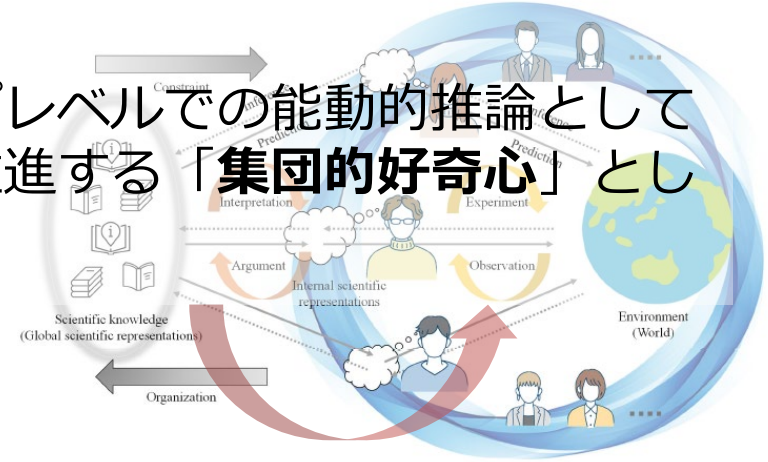


Figure 4. (a) Left: expected classical CPC models show gradual, continuous updates of the posterior distribution $q(\theta \mid \mathbf{o})$ as sample size increases. Right: singular CPC models demonstrate discontinuous jumps in the posterior distribution, representing paradigm shifts in scientific understanding. (b) A network representation of the singular CPC model, where scientists (Agent 1 and Agent 2) act as neurons. The connections between nodes illustrate how scientific concepts or theories interact and evolve, potentially leading to sudden structural changes in the collective knowledge network.

Active inference in CPC

- **集合的期待自由エネルギー**の式は、CPCフレームワークの自由エネルギー原理から自然に導き出すことができる。それは、集合的な項（Collective Epistemic Value）を組み込むことによって、個体の期待自由エネルギーを拡張する。
- この定式化により、創発的な言語／記号／科学的知識(w)によって動機付けられた、または導かれる能動的な推論を検討することができる。
- 今後の研究には、この概念をグループレベルでの能動的推論として探求し、科学的探求や社会的適応を推進する「**集団的好奇心**」として捉える可能性を秘めています。

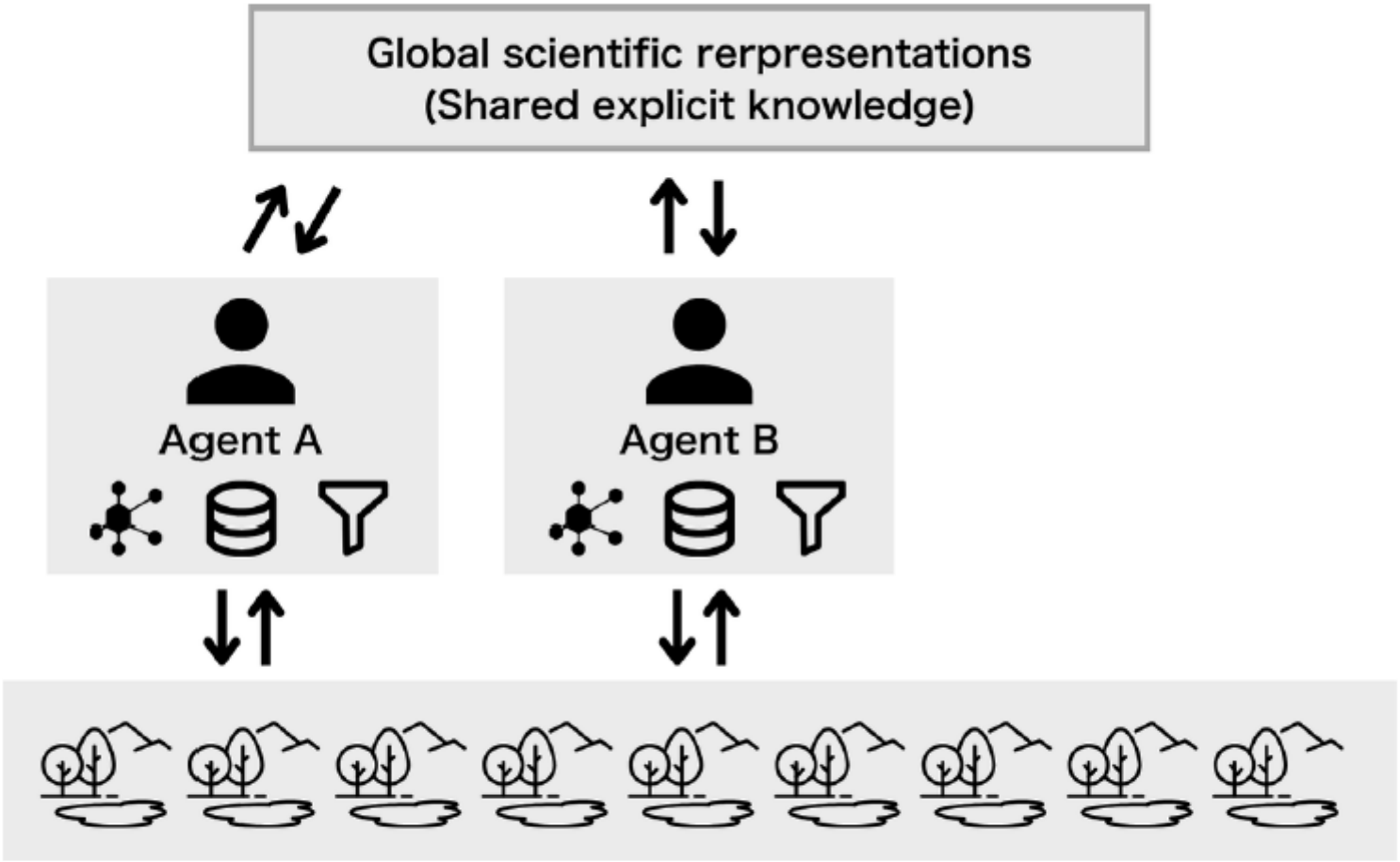


$$G(\tilde{\mathbf{a}}) = \mathbb{E}_{q(\tilde{w},\tilde{\mathbf{z}},\tilde{\mathbf{o}}|\tilde{\mathbf{a}},\tilde{\mathbf{C}})}[\tilde{F}] \tag{8}$$

$$= \int q(\tilde{w},\tilde{\mathbf{z}},\tilde{\mathbf{o}} \mid \tilde{\mathbf{a}},\tilde{\mathbf{C}})\tilde{F}\ \mathrm{d}\tilde{w}\ \mathrm{d}\tilde{\mathbf{z}}\ \mathrm{d}\tilde{\mathbf{o}} \tag{9}$$

$$= \underbrace{\mathbb{E}_q\left[\ln q\left(\tilde{w} \mid \{\tilde{z}^k\}_k\right)\right]}_{\text{Collective epistemic value}} - \sum_k \underbrace{\mathbb{E}_q\left[\ln p\left(\tilde{o}^k \mid \tilde{z}^k,\tilde{a}^k,\tilde{C}^k\right)\right]}_{\text{Individual pragmatic value}} - \sum_k \underbrace{\mathbb{E}_q\left[\ln \frac{p\left(\tilde{z}^k \mid \tilde{w},\tilde{a}^k\right)}{q\left(\tilde{z}^k \mid \tilde{w},\tilde{o}^k,\tilde{a}^k\right)}\right]}_{\text{Individual epistemic value}} \tag{10}$$

Taniguchi, T., Takagi, S., Otsuka, J., Hayashi, Y., & Hamada, H. T. (2025). Collective predictive coding as model of science: Formalizing scientific activities towards generative science. Royal Society Open Science, 12(6), 241678.



In CPC, science is a multi-agent system

共創的学習（Co-creative learning）

Cornell University

We gratefu
the Simons Foundation, Stockholm U

arXiv > cs > arXiv:2506.15468

Search...
Help | Adv

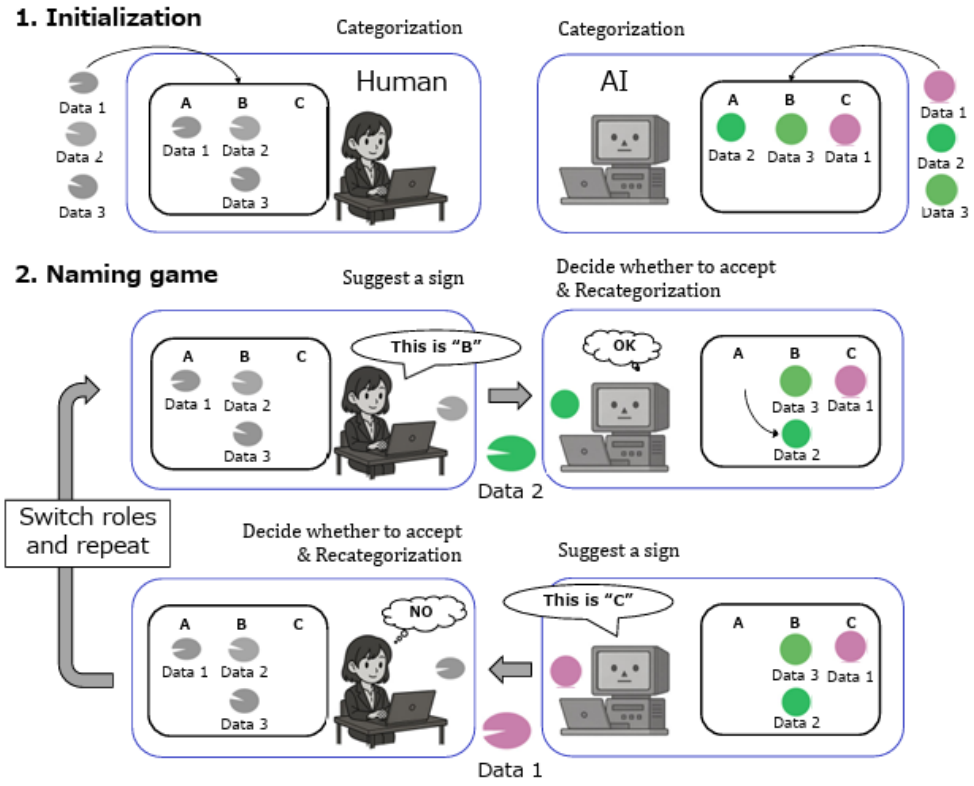
Computer Science > Human-Computer Interaction

[Submitted on 18 Jun 2025]

Co-Creative Learning via Metropolis-Hastings Interaction between Humans and AI

Ryota Okumura, Tadahiro Taniguchi, Akira Taniguchi, Yoshinobu Hagiwara

We propose co-creative learning as a novel paradigm where humans and AI, i.e., biological and artificial agents, mutually integrate their partial perceptual information and knowledge to construct shared external representations, a process we interpret as symbol emergence. Unlike traditional AI teaching based on unilateral knowledge transfer, this addresses the challenge of integrating information from inherently different modalities. We empirically test this framework using a human-AI interaction model based on the Metropolis-Hastings naming game (MHNG), a decentralized Bayesian inference mechanism. In an online experiment, 69 participants played a joint attention naming game (JA-NG) with one of three computer agent types (MH-based, always-accept, or always-reject) under partial observability. Results show that human-AI pairs with an MH-based agent significantly improved categorization accuracy through interaction and achieved stronger convergence toward a shared sign system. Furthermore, human acceptance behavior aligned closely with the MH-derived acceptance probability. These findings provide the first empirical evidence for co-creative learning emerging in human-AI dyads via MHNG-based interaction. This suggests a promising path toward symbiotic AI systems that learn with humans, rather than from them, by dynamically aligning perceptual experiences, opening a new venue for symbiotic AI alignment.



(b) Experimental procedure flow (JA-NG).

Okumura, R., Taniguchi, T., Taniguchi, A., & Hagiwara, Y. (2025). Co-Creative Learning via Metropolis-Hastings Interaction between Humans and AI. *arXiv preprint arXiv:2506.15468*.

AIと人間が一緒に集合的自由エルギー を減少させる活動としての **共創的学習（Co-creative learning）**

Definition 1 (Co-Creative Learning) Let $X_n = (x_n^m)_{m\in\mathcal{A}}$ be the partial observations generated via the shared latent symbol s_n as defined above. Co-creative learning is the interactive, decentralized Bayesian inference process in which the agents, through message-based updates (e.g., the Metropolis-Hastings naming game), jointly construct local belief distributions $q_t^m(s_n)$ such that the collective free energy,

$$\mathcal{F}_t = -\log p(X^{\mathrm{AI}}, X^{\mathrm{Human}}) + \mathrm{KL}(q_t(s)||p(s|X^{\mathrm{AI}}, X^{\mathrm{Human}})),$$

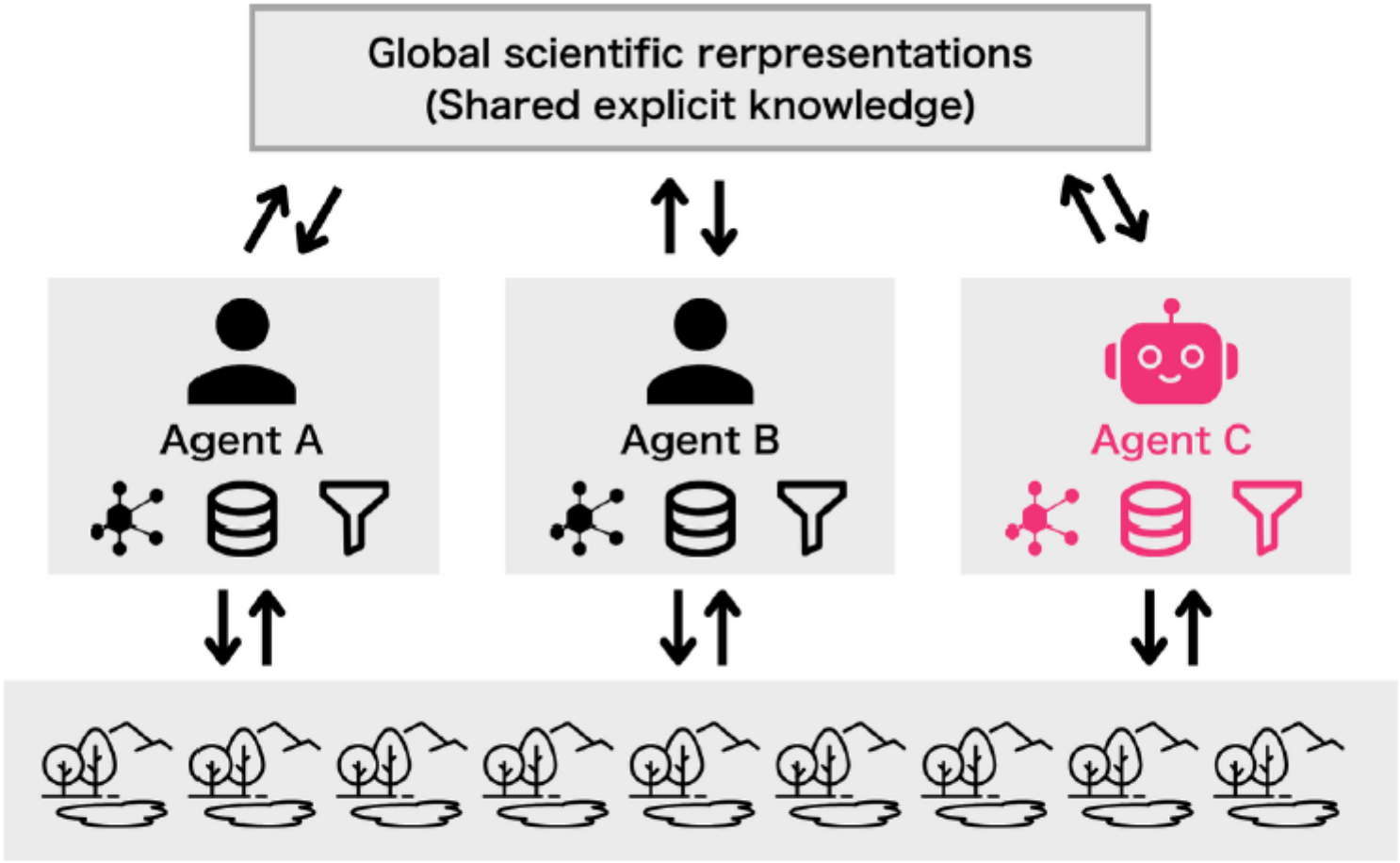
$$\text{with} \quad q_t(s) \propto \prod_{m\in\mathcal{A}} q_t^m(s),$$

satisfies

$$\mathbb{E}[\mathcal{F}_{t+1}] \leq \mathbb{E}[\mathcal{F}_t].$$

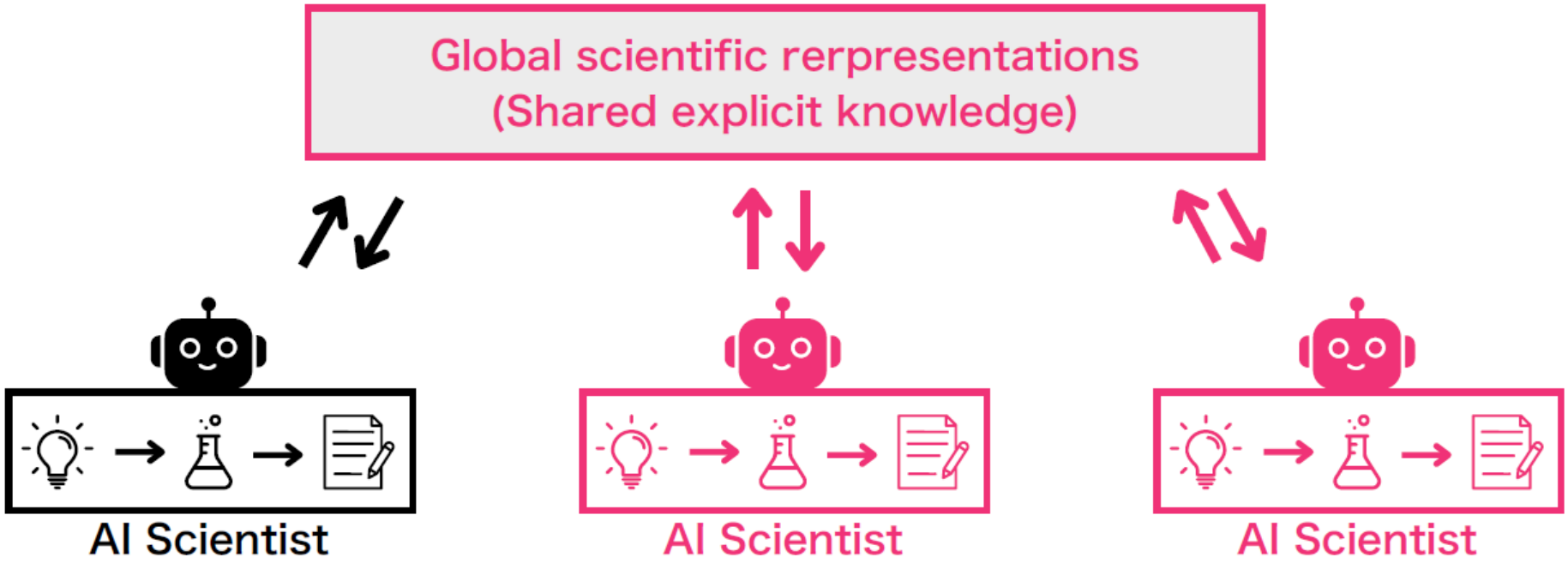
That is, the expected collective predictive-coding free energy decreases over interaction steps. Equivalently, the Markov chain formed by the interactions converges (in distribution) to the true posterior $p(s|X^{\mathrm{AI}}, X^{\mathrm{Human}}, \Theta)$, where $\Theta = (\Theta^m)_{m\in\mathcal{A}}$, so that Monte Carlo estimates of the evidence $p(X^{\mathrm{AI}}, X^{\mathrm{Human}})$ improve without requiring either agent to disclose private observations or gradients.

Okumura, R., Taniguchi, T., Taniguchi, A., & Hagiwara, Y. (2025). Co-Creative Learning via Metropolis-Hastings Interaction between Humans and AI. *arXiv preprint arXiv:2506.15468*.



An AI scientist is just one of the agents with distinct traits

Towards Automated Science System



- 研究の自動化の試みの多くは科学の特定の過程の自動化や人工的な「科学者」の実装を目指すが、科学は社会的な全体的な営み
- CPC-MS は科学コミュニティ全体の数学的／確率的モデルであり、科学コミュニティ全体の自動化を目指し実装する一つの出発点を提供

まとめ：CPC-MSと生成科学への道

- **科学の新しいモデル提示**: 科学を「集合的予測符号化（CPC-MS）」という新モデルで捉える 。これは科学活動を、コミュニティ全体で行う分散的なベイズ推論プロセスと見なすものである 。
- **科学プロセスの定式化**: 科学者の仮説形成や査読といった活動は、共有知識を更新する確率的推論アルゴリズムと構造的に類似する 。これにより、科学の社会的客観性やパラダイム転換を形式的に説明可能となる 。
- **AI科学者の位置付け**: 「AI科学者」は科学全体を代替するのではなく、人間の科学者と共に活動する新しいエージェントの一つとして位置づけられる 。AIは、集合的な知の営みにおける特有の能力を持つ新たな参加者である 。
- **「生成的科学」への道筋**: CPC-MSは、人間とAIが協働する将来の「自動駆動科学システム」の理論的基礎となる 。これは科学的探求自体を生成・加速させる「**生成科学**」への道筋を示すものである 。

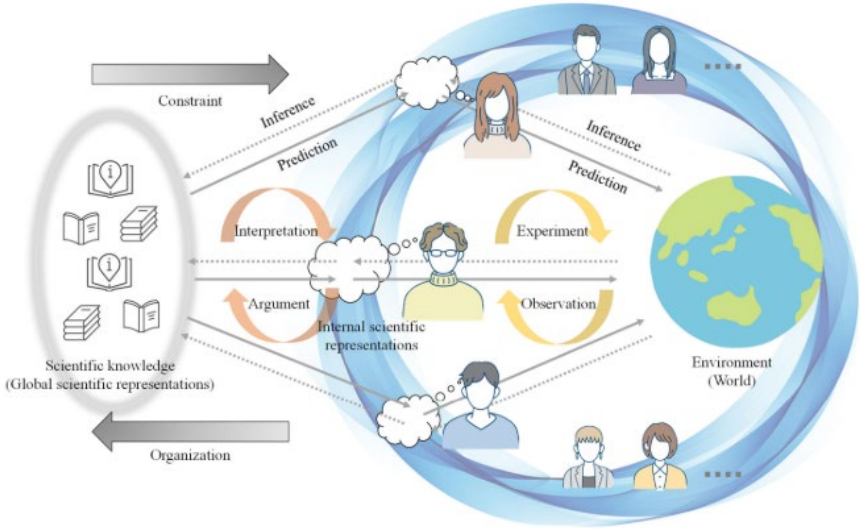
Information

谷口忠大（編）『記号創発システム論ー来るべきAI共生社会の「意味」理解にむけて（ワードマップ）』新曜社（2024）

Collective Predictive Coding as Model of Science: Formalizing Scientific Activities Towards Generative Science

Tadahiro Taniguchi[1], Shiro Takagi[2], Jun Otsuka[1], Yusuke Hayashi[3], Hiro Taiyo Hamada[4],
[1]Kyoto University, [2]Independent Researcher [3]AI Alignment Network [4]ARAYA Inc.

Paper | arXiv | Slides | Video



https://cpc-ms.github.io/

Email: taniguchi@i.kyoto-u.ac.jp
X (personal): @tanichu