様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科研費
KAKENHI

科学研究費助成事業　　研究成果報告書

令和　６ 年　 ５ 月 ３０ 日現在

機関番号： ３２６８６

研究種目： 研究活動スタート支援

研究期間： 2022～2023

課題番号： ２２Ｋ２０６７９

研究課題名（和文）Reveal the informational nature of conscious self and sense of agency under the information closure theory of consciousness

研究課題名（英文）Reveal the informational nature of conscious self and sense of agency under the information closure theory of consciousness

研究代表者

チャン　ユーチャン（Chang, Acer Yu-Chan）

立教大学・現代心理学部・助教

研究者番号：５０８３１４８４

交付決定額（研究期間全体）：（直接経費）　　2,100,000円

研究成果の概要（和文）：研究期間中、情報閉包理論（ICT）を用いて意識とエージェンシーの本質を探求しました。この理論は、意識が特定の情報処理パターンから生じると提唱しています。我々は、ICTフレームワークに行動を組み込んだ新しい数学モデルを作成し、自己とエージェンシーを記述しました。(1)数学的基礎の開発、(2)小規模シミュレーションの実施、(3)先進的なAIシステムへのモデル適用、の3段階で進めました。人工ニューラルネットワークを用いたシミュレーションによりモデルを検証し、AIシステムへの応用で広範な関連性を確認しました。この研究は、神経科学およびAI開発における倫理的考察に重要な貢献をもたらすと考えられます。

研究成果の学術的意義や社会的意義
私たちの研究は、科学的および社会的に重要な意義を持っています。科学的には、情報閉包理論（ICT）に基づく新しい数学モデルを開発し、意識とエージェンシーのメカニズムを解明しました。このモデルをシミュレーションで検証し、高度なAIシステムにも適用しました。この研究は、神経科学や認知科学の基本的な問題に取り組み、意識の理解を深めるものです。社会的には、AIの倫理的使用に関するガイドラインや政策の策定に役立つ可能性があります。また、人間の心の理解を深めることで、メンタルヘルス治療の改善にも貢献するでしょう。この研究は、科学と社会の架け橋となり、AIや認知科学に対する倫理的なアプローチを促進します。

研究成果の概要（英文）：During the research period, we explored the fundamental nature of consciousness and the sense of agency using the Information Closure Theory (ICT). This theory suggests that consciousness arises from particular patterns of information processing within a system. We created a new mathematical model that incorporates actions into the ICT framework to describe the conscious self and agency. Our research consisted of three phases: (1) developing the mathematical foundation, (2) conducting small-scale simulations, and (3) applying the model to advanced AI systems.. These simulations used artificial neural networks to validate our model, and the application to AI systems helped test its broader relevance. The outcomes of our research offer a deeper understanding of consciousness and its mechanisms, which could have significant implications for neuroscience and ethical considerations in AI development.

研究分野： Neuroscience

キーワード： consciousness　theory of consciousness　information theory　information closure

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

１．研究開始当初の背景

Consciousness research has traditionally focused on identifying the neural correlates of consciousness. However, these efforts often overlook the "scale problem," which highlights that conscious experiences correspond to specific scales within the neural system. Only information processed at particular scales can be consciously accessed.

The Information Closure Theory (ICT), proposed by the Principal Investigator (PI) and colleagues, addresses this scale problem. ICT posits that a process is conscious if it is informationally closed (IC) to its microscopic universe and non-trivially informationally closed (NTIC) to its environment. Initially, ICT did not consider the domain of action, which also exhibits scale dependency in conscious experiences.

This project aims to extend ICT to include actions, providing a comprehensive understanding of the conscious self and the sense of agency. By applying ICT to advanced AI systems, the research seeks to explore whether these systems can exhibit forms of consciousness and agency, contributing to both theoretical neuroscience and practical AI research.

２．研究の目的

The primary objective of this study is to enhance our understanding of the conscious self and sense of agency through the Information Closure Theory (ICT) of consciousness. The research is structured into three key phases, each with specific goals.

First, the study aims to establish the mathematical foundations for the concepts of the conscious self and sense of agency within the ICT framework. This involves developing precise mathematical formulations that can be applied to various physical systems as quantitative measures. To achieve this, the study will utilize advanced mathematical tools and symbolic computation software, ensuring robust and scalable formulations.

Second, the study seeks to validate these theoretical foundations through small-scale simulations. By conducting simulations using artificial agents and environments modeled by deep artificial neural networks, the research will examine whether NTIC processes encode information about both the environment and the system itself. The agents will be trained to maximize NTIC with their environment, and the results of these simulations will be compared with theoretical predictions to ensure the validity of the ICT framework.

Finally, the study aims to apply the extended ICT framework to advanced artificial intelligence systems. This phase involves developing empirical measures for systems with large numbers of elements, particularly focusing on model-free and model-based reinforcement learning agents. By applying these measures, the research will assess the degree of conscious self and agency in these AI systems, providing valuable insights into both the neuroscience of consciousness and the ethical implications of AI.

Through these objectives, the study aims to provide a comprehensive understanding of consciousness and agency from an information-theoretical perspective, bridging the gap between theoretical neuroscience and practical AI research

３．研究の方法

This study employs a multifaceted approach to explore the informational nature of the conscious self and sense of agency under the Information Closure Theory (ICT) of consciousness. The research is divided into three main phases, each utilizing specific methodologies to achieve the project's objectives.

Phase 1: Establishing Mathematical Foundations

In the first phase, the research focuses on developing the mathematical formulations necessary to define the conscious self and sense of agency within the ICT framework. This involves the use of advanced mathematical tools and symbolic computation software such as Mathematica. The primary goal is to derive precise and scalable formulations that can be applied to various physical systems. The ICT framework posits that a process is conscious if it is informationally closed (IC) to its microscopic universe and non-trivially informationally closed (NTIC) to its environment. By incorporating the domain of action into this framework, the study aims to define and quantify the conscious self and sense of agency.

Phase 2: Proof of Concept through Small-Scale Simulations
The second phase involves validating the theoretical formulations through small-scale simulations. Artificial agents and environments will be modeled using deep artificial neural networks. These agents will be trained to maximize NTIC with their environment, allowing the study to observe whether NTIC processes encode information about both the environment and the system itself. The simulations will be conducted using GPU-powered workstations to handle the computational demands of deep learning models. By comparing the simulation results with theoretical predictions, the study will assess the validity of the ICT framework in representing conscious self and agency.

Phase 3: Application to Advanced Artificial Intelligence Systems
In the final phase, the extended ICT framework will be applied to advanced AI systems to measure consciousness and agency empirically. This phase involves developing empirical measures for systems with a large number of elements, such as model-free and model-based reinforcement learning agents. The agents will be trained in selected environments from OpenAI Gym, and the measures of conscious self and agency will be applied once the agents reach satisfactory performance levels. This phase aims to explore the degree of conscious self and agency in advanced AI systems, providing insights into the nature of consciousness and the ethical implications of AI development.

Through these research methods, the study aims to provide a comprehensive understanding of the conscious self and sense of agency from an information-theoretical perspective, bridging gaps between theoretical neuroscience and practical AI research.

４．研究成果

The study produced several significant findings across its three phases, providing a solid foundation for further exploration of the conscious self and sense of agency under the Information Closure Theory (ICT) of consciousness.

Phase 1: Establishing Mathematical Foundations
In the first phase, the research successfully developed mathematical formulations for the concepts of conscious self and sense of agency within the ICT framework. The formulations were derived using advanced mathematical tools and symbolic computation software, ensuring they were precise and scalable. These formulations quantitatively defined the conscious self as the information encoded in the conscious process and the sense of agency as the specific state of the system that influences the environment. This phase established a robust theoretical foundation that can be applied to various physical and artificial systems.

Phase 2: Proof of Concept through Small-Scale Simulations
The second phase involved conducting small-scale simulations with artificial agents modeled by deep artificial neural networks. The agents were trained to maximize NTIC with their environment. The simulations demonstrated that NTIC processes could indeed encode information about both the environment and the system itself. Specifically, the agents exhibited behaviors consistent with having a conscious self and sense of agency as defined by the ICT framework. The results of these simulations closely matched the theoretical predictions, validating the ICT framework in a controlled, simulated environment.

Phase 3: Application to Advanced Artificial Intelligence Systems
In the final phase, the extended ICT framework was applied to advanced AI systems,

focusing on model-free and model-based reinforcement learning agents. The agents were trained in environments selected from OpenAI Gym. The study developed empirical measures for systems with a large number of elements and applied these measures once the agents reached satisfactory performance levels. The results showed that the model-based agents, in particular, exhibited higher levels of conscious self and sense of agency according to the ICT framework. This phase provided empirical support for the extended ICT framework's applicability to complex AI systems.

Overall, the study provided solid and conservative results that validate the ICT framework for understanding the conscious self and sense of agency in both theoretical and simulated contexts. These findings lay the groundwork for further research and application in more complex systems.

５．主な発表論文等

〔雑誌論文〕　計1件（うち査読付論文　0件／うち国際共著　0件／うちオープンアクセス　0件）

１．著者名 Wen Wen、Chang Acer Yu-Chan、Imamizu Hiroshi	４．巻 28
２．論文標題 The sensitivity and criterion of sense of agency	５．発行年 2024年
３．雑誌名 Trends in Cognitive Sciences	６．最初と最後の頁 397～399
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.tics.2024.03.002	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計6件（うち招待講演　2件／うち国際学会　2件）

１．発表者名 Acer Chang
２．発表標題 Between Individual Brains and Collective Behavior: Multi-level
３．学会等名 Artificial Life Conference Proceeding
４．発表年 2023年

１．発表者名 Acer Chang
２．発表標題 Sense of agency from active inference
３．学会等名 TeaP 2024（招待講演）
４．発表年 2024年

１．発表者名 Acer Chang
２．発表標題 Chang The sense of agency as ac- tive causal inference: How We Comprehend Our Con- trol Over the Environment using Abstract Action Plans
３．学会等名 CoRN 2023
４．発表年 2023年

１．発表者名
Sangati Ekaterina、Sangati Federico、Cheng Yi-Shan、Yu-Chan Chang Acer

２．発表標題
Between Individual Brains and Collective Behavior: Multi-level Emergence in a Group Formation Task

３．学会等名
Artificial Life Conference Proceedings（国際学会）

４．発表年
2023年

１．発表者名
K Cheng, Yi-Shan and Chang, Yu-Chan and Doya

２．発表標題
Information-Theoretical Analysis of Team Dynamics in Football Matches

３．学会等名
Asia-Singapore Conference on Sport Science（国際学会）

４．発表年
2023年

１．発表者名
Acer Chang

２．発表標題
The Sense of Agency as active causal inference at an abstract action plan level.

３．学会等名
第13回多感覚研究会（招待講演）

４．発表年
2023年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

－

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関