

# 存在を認識する過程と意識

## - 忘れられ勝ちな種別上の**主キー**系列 -



2022年5月11日

山川宏

# アジェンダ

1. 世界モデルにおいて存在を記述する意義
2. Entification処理： 存在を見出す処理
3. 意識のEntification仮説
4. Entification仮説の位置づけ
5. 脳に学ぶEntification機構の探求に向けて
6. まとめ

# AIにおける存在を得る(Entification)の必要性

---

- 世界のモデル化において、「存在」を記述することは重要。
- 昨今の、Object centricな表現を扱う機械学習AIでは、存在の表現を獲得しつつある(後述)が、まだ十分ではないように思われる。
- この状況を明らかにするため、世界にあると想定される「存在の体系」を探索しつづけるEntificationについてGOFAIや存在論哲学的な観点を踏まえた検討を行っている。

# 世界モデルにおいて「存在」を記述する意義

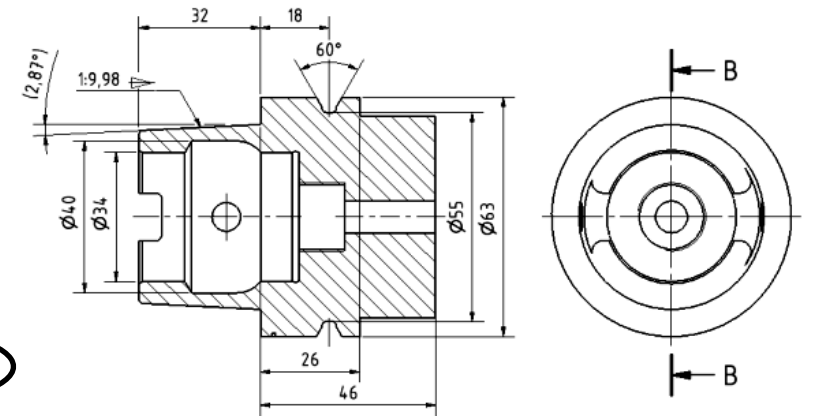
良い予測・行動を行うために世界をモデル化する知能において、事物を存在(Entity)として扱うの能力は、記述の縮約、記号化、思考の効率化の基盤と重要である。

## No **X** without Entity

**X = Tool**

**X = Language**

**X = Design**



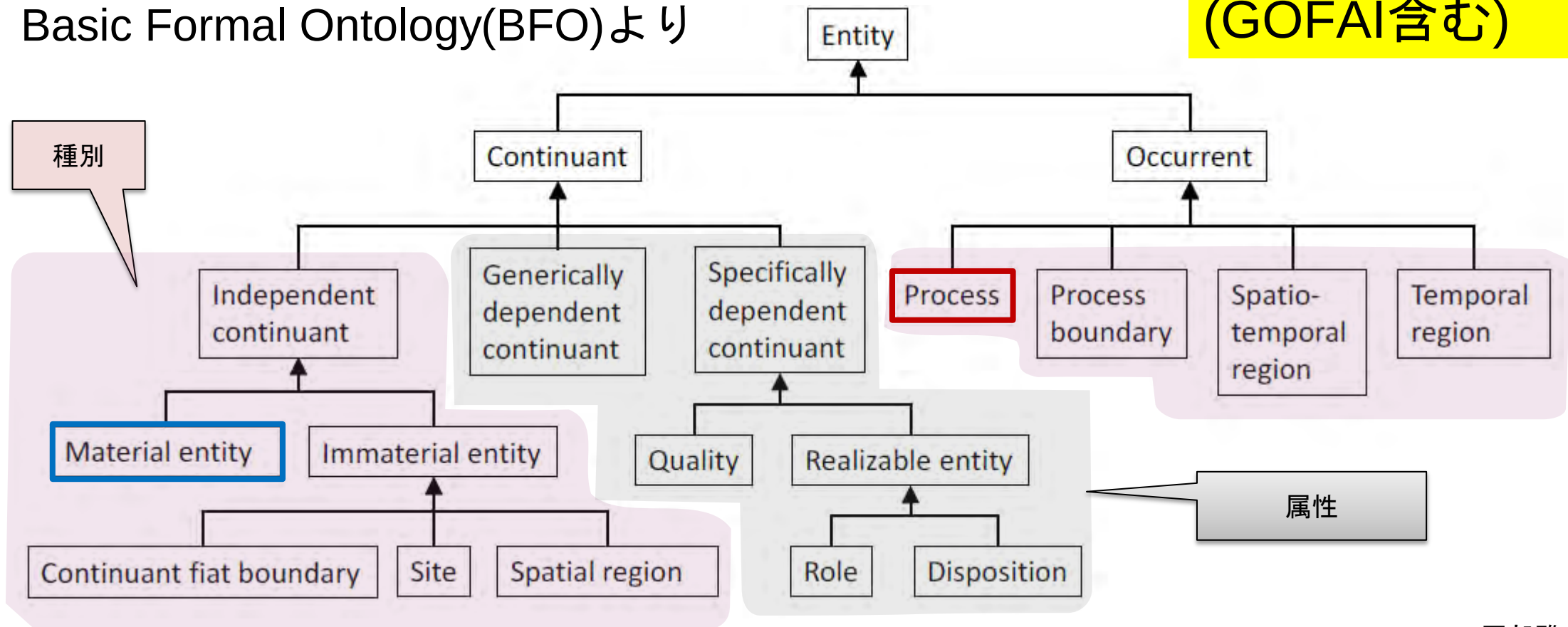
# 記号で記述されてきた存在 (Entity)の広がり

存在は2つに分けられる

- 種別(Kinds) : 数えられる (Independent) ← 着目
- 属性(Attribute) : 種を特徴づける (Dependent)

ヒトは様々な設計を  
記号(図・数式等)を  
用いて行ってきた  
(GOFAI含む)

Basic Formal Ontology(BFO)より



# Process Entity化 が AGIを導きうる

汎用人工知能(AGI)：未知を含む様々なタスクを解決できるAIという目標

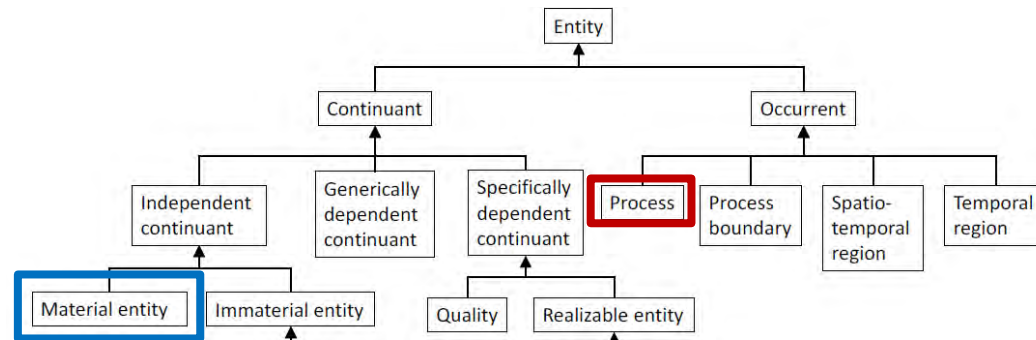
深層学習は、言語指示に基づいて物理的な画像の結合を行うことができる。



「バイクに乗ったピカチュウ」「アボカドの椅子」の静止画を生成するなど

OpenAI (2021). DALL-E: Creating Images from Text, <https://openai.com/blog/dall-e/>

タスクは、存在の一種である Processに含まれるだろう。



ProcessのEntity化ができれば、

AGIの課題である、タスクの柔軟な結合へ発展しうるかも：

例えば、キッチンで料理をする作業と片付ける作業を区別しながら、同時に組み合わせて効率的に実行

# アジェンダ

---

1. 世界モデルにおいて存在を記述する意義
2. Entification処理： 存在を見出す処理
3. 意識のEntification仮説
4. Entification仮説の位置づけ
5. 脳に学ぶEntification機構の探求に向けて
6. まとめ

# 認知的な個別化の定義例 (横路佳幸, 同一性と個体, 2021)

認知主体 $S$ の視覚に基づく「認知的な個別化」の成功とは、「問題の対象 $o$ はどれか」に対して答えを与えるような**同一性**に関する特定の命題を $S$ が知っていること、すなわち  $\exists x(K_S(o = x))$  が成立することであると約定すると、次が成立する。すなわち、ある一定の視覚能力を備えた知覚主体 $S$ が通常の状態において $o$ を認知的に個別化することに成功するとき、

- $o$ を特定の**種別概念 $F$** を例化するものとして正しく表象する $S$ の知識体験 $e_1$ 、
- さらに同じ **$F$** であるという特定の**種別的同一性**が $o$ と特定の個体 $a$ の間に成立するという $S$ の真なる信念 $e_2$ にもとづいて、 $S$ は $o = a$ と信じる。

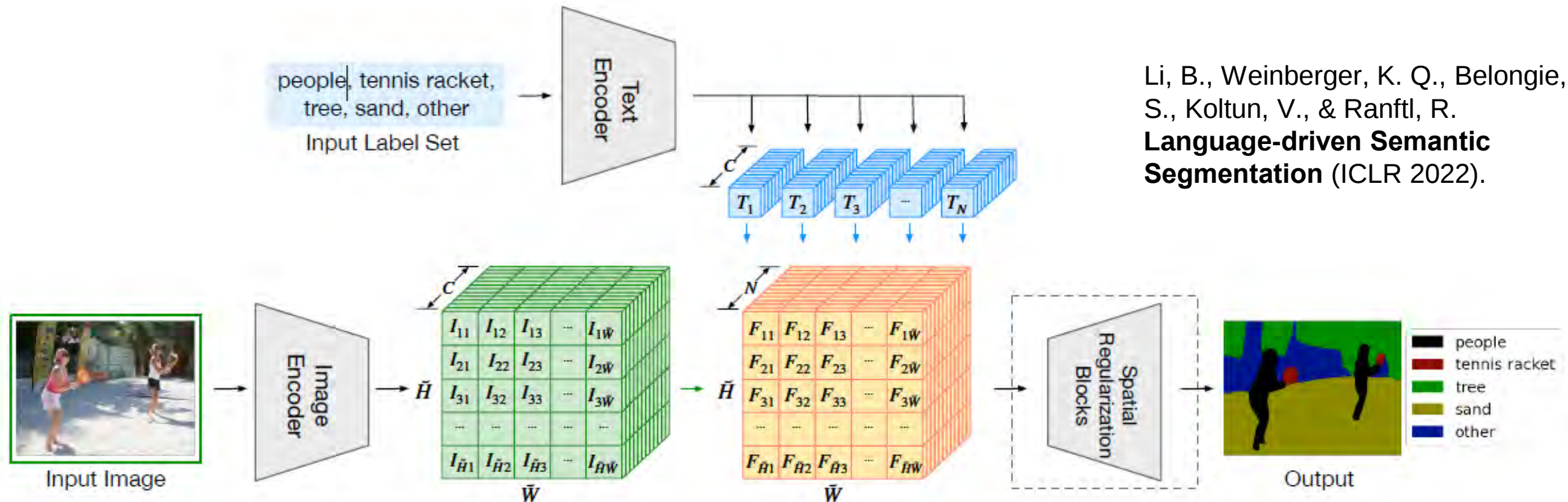
このとき $e_1$ と $e_2$ は、 $o = a$ という**同一性**に関する命題が真であることを指し示すものであると同時に、 $o = a$ という単称的な信念を認識的に正当化する叙述的な証拠である。

- $S$ : 認知主体
- $a$ : 個体
- $o$ : 対象
- $e_1$ : 知識体験
- $e_2$ : 真なる信念
- $F$ : **種別概念**
- $K$ : 命題知

# 同一性を利用したSemantic Segmentationの発展状況

データ拡張を通じて**同一性**の情報を与える対象(contrastive)学習(自己教師あり学習の一種)により、pixel単位で切り出された画像領域をテキストと対応付ける。

(主な関連研究としてCLIP: *Contrastive Language-Image Pre-training* (OpenAI, 2021)がある)



# 同一カテゴリの画像範囲に似たようなピクセルを統合する仕組み

Step 1: Look for regions that likely belong together

→ Shared pixel ownership assumption

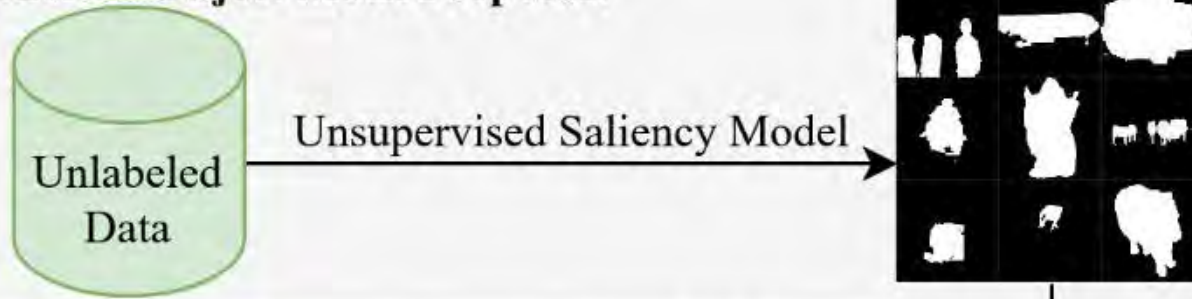
→ Use a mid-level visual prior

Step 2: Generate semantic pixel embeddings

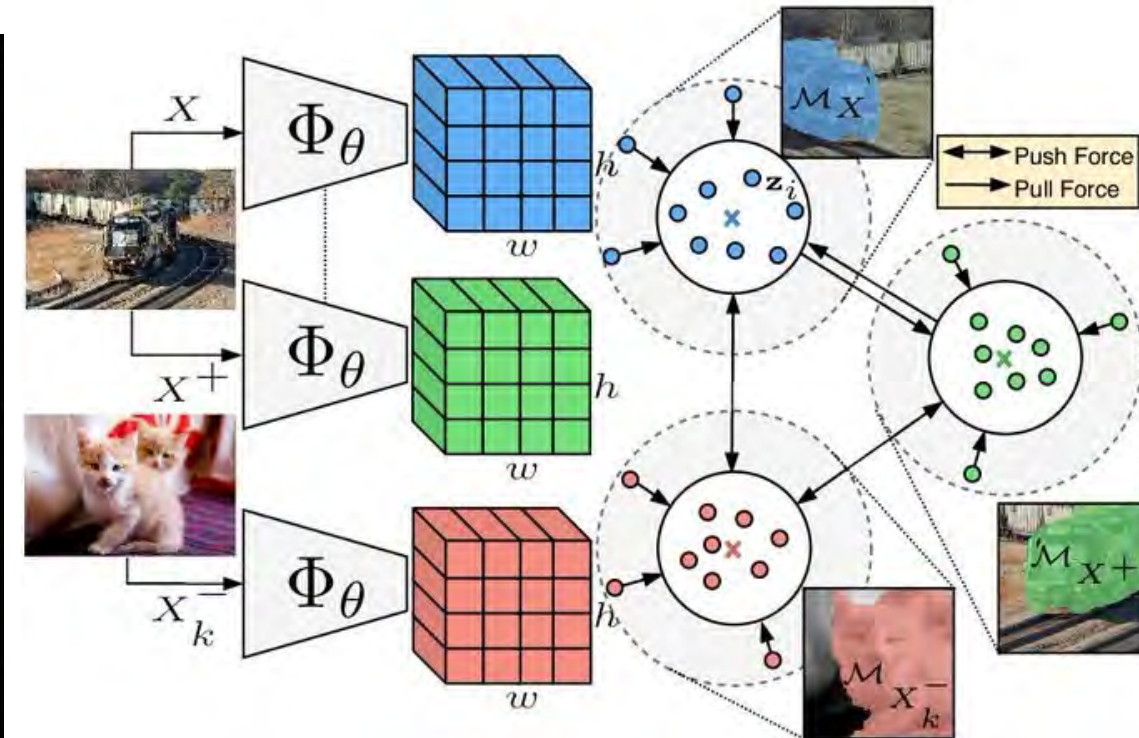
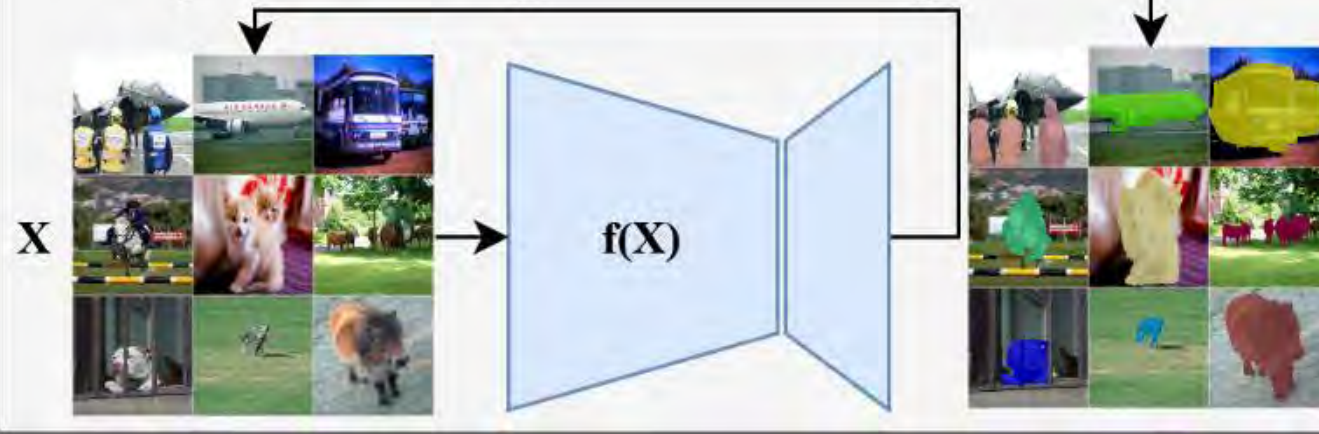
→ Leverage object mask proposals

→ Maximize or minimize the agreement

## I. Generate Object Mask Proposals



## II. Unsupervised Learning of Pixel Embeddings



Van Gansbeke, W., Vandenhende, S., Georgoulis, S., & Van Gool, L. (2021). Unsupervised Semantic Segmentation by Contrasting Object Mask Proposals. In 2021 IEEE/CVF International Conference on Computer Vision (ICCV).

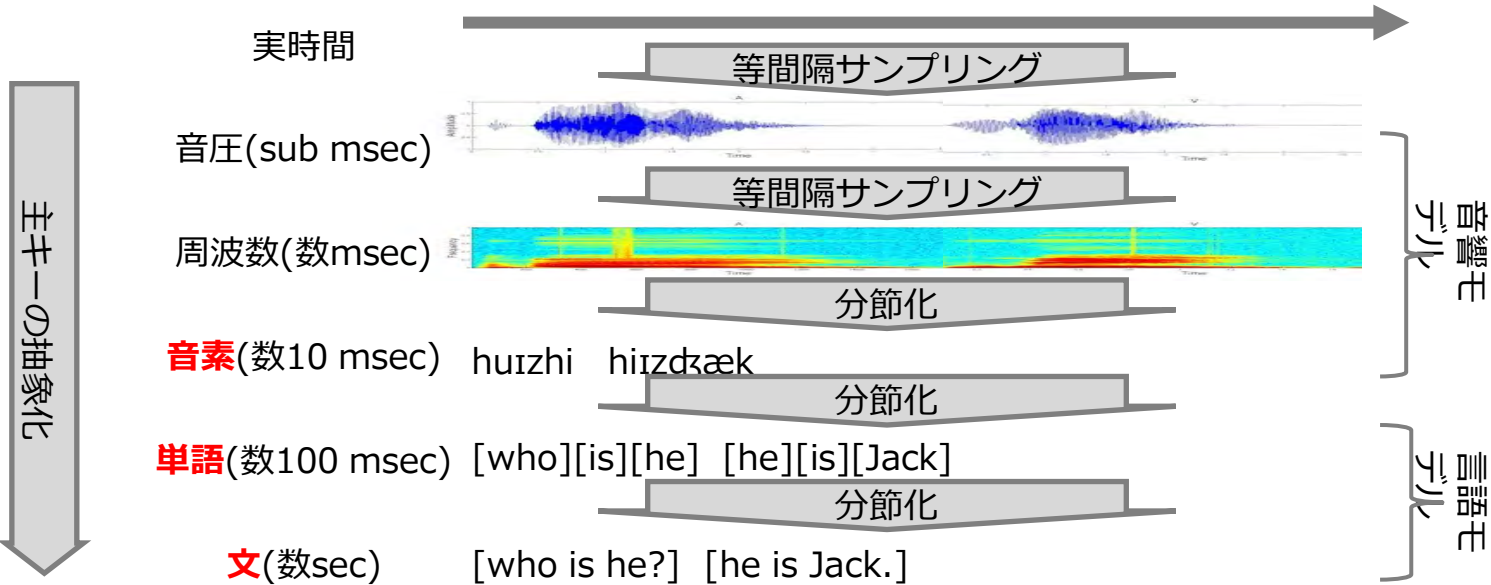
<https://doi.org/10.1109/iccv48922.2021.00990>

# 多様な種別において、カウントするための主キー

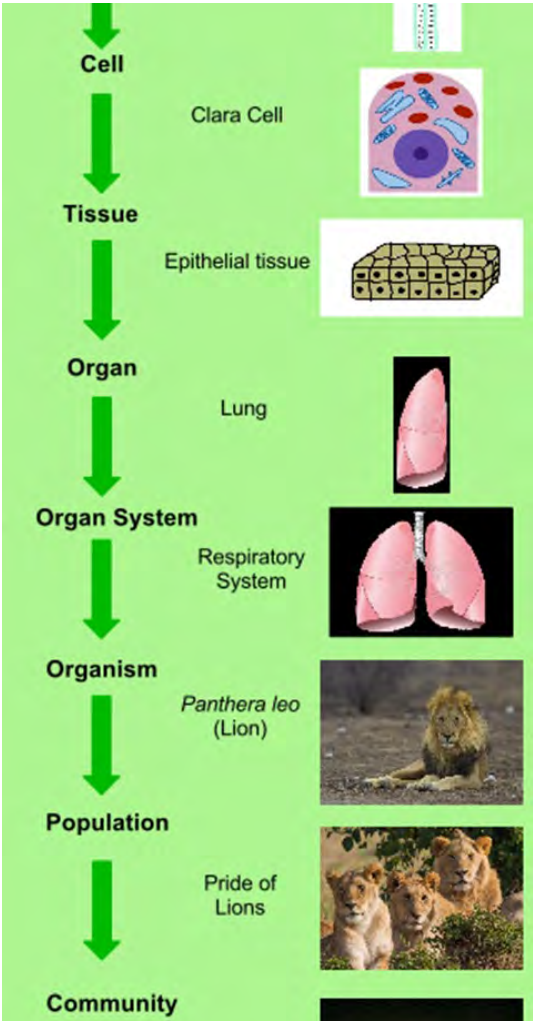
主キーは、種別毎に数えられる事物を指定する

例：時刻の集合、空間位置の集合、物体の集合、音素の集合など

言語認識における種別：主キーが段階的に抽象化される

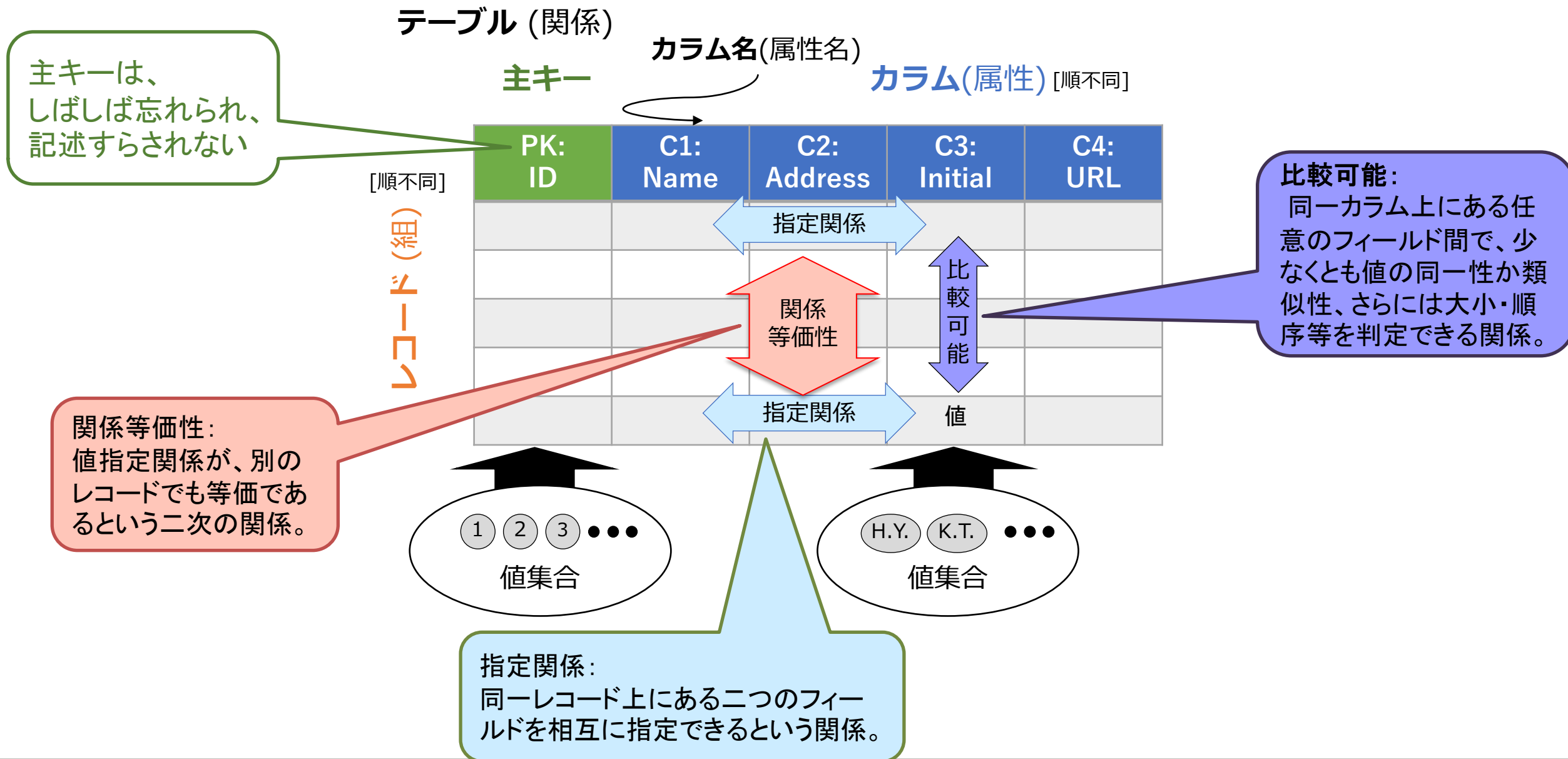


## 生命における種別概念の階層性の例



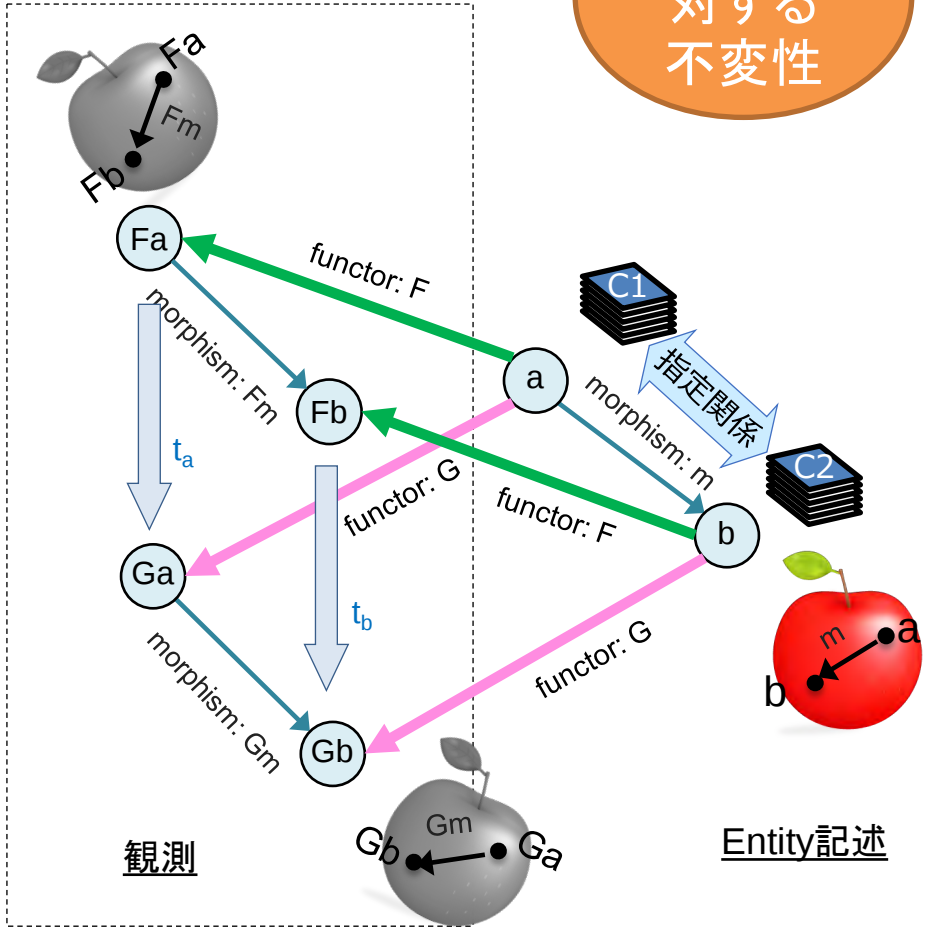
出典: [https://en.wikipedia.org/wiki/Biological\\_organisation](https://en.wikipedia.org/wiki/Biological_organisation)

# 同一性(や帰納法)の前提となる整列構造(～RDB)と主キー

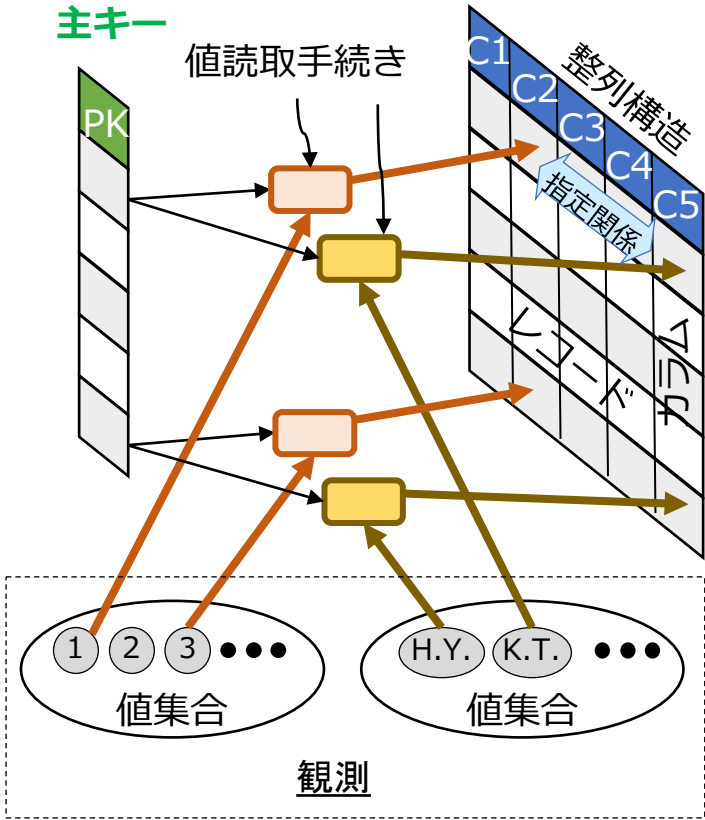


# 整列注意機構：整列構造を得るための一般的なメカニズム(山川)

## 存在の認識：



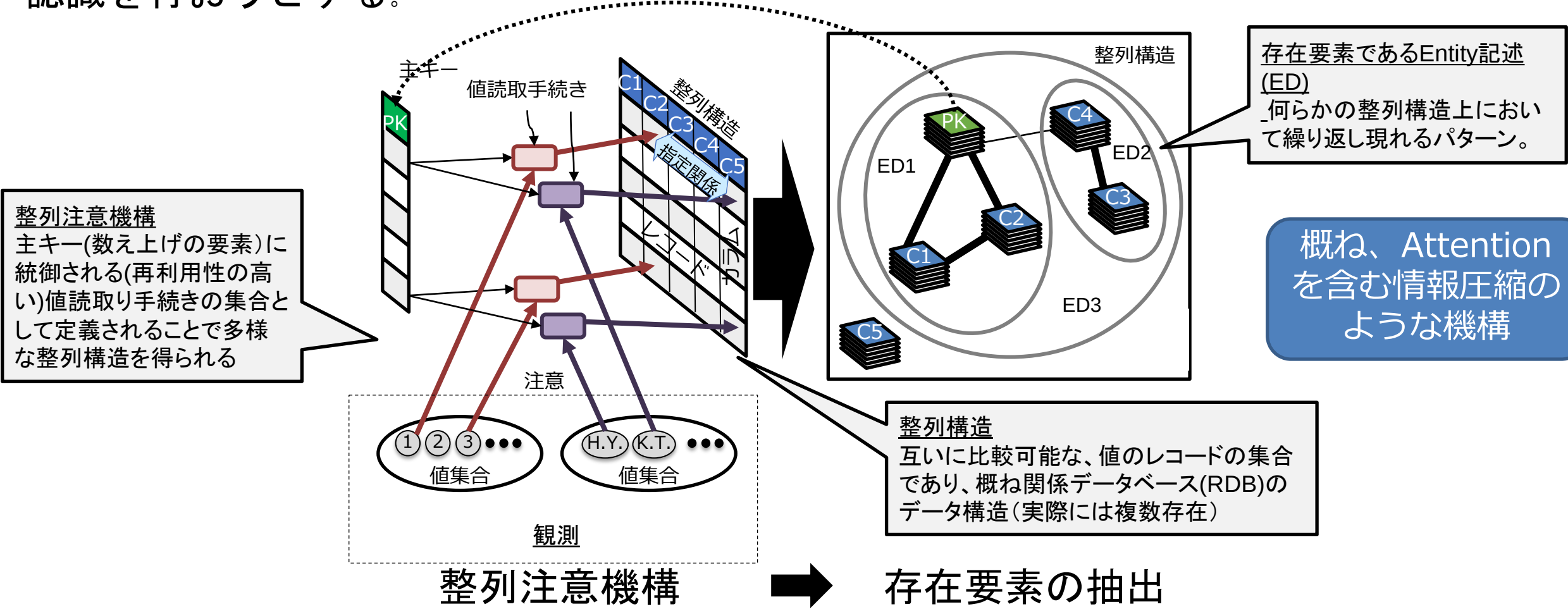
## 整列注意機構：主キーの各値に応じて、特定の値集合から値を読み取る読取り手続きの集合からなる注意機構



世界の観測における指定関係( $Fm$ ,  $Gm$ )は、主キーに統括された値読取手続きに反映される。

# 世界から存在を取り出すEntification処理の全体像

観測される情報から、複数の視点から見て不変となりうる部分を、より一貫性が高くコンパクトな、Entity記述(存在要素)からなる体系として記述し、それらを用いて認識を行おうとする。



# アジェンダ

---

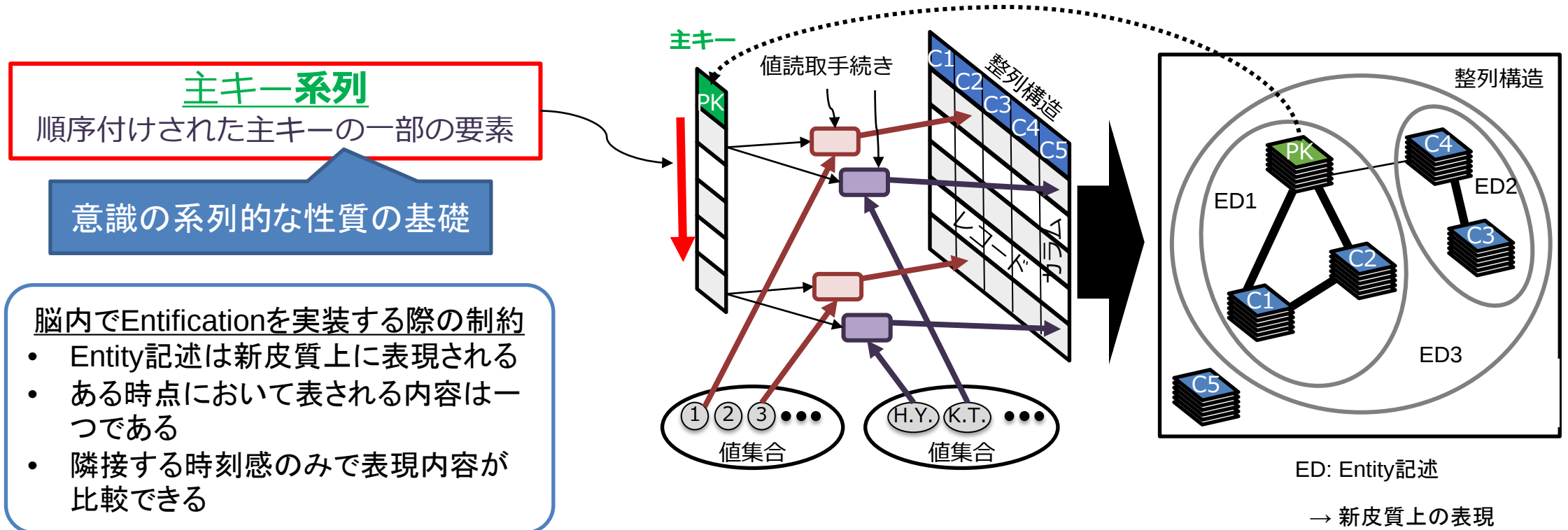
1. 世界モデルにおいて存在を記述する意義
2. Entification処理： 存在を見出す処理
3. 意識のEntification仮説
4. Entification仮説の位置づけ
5. 脳に学ぶEntification機構の探求に向けて
6. まとめ

## 意識を知るために、忘れられ勝ちな主キーを呼び起こす

- データ(RDB)とそれを用いた統計処理において、数え方を決定する主キーは本質的に重要である
- だが、特定のデータだけに着目すれば、主キーを記述せずとも、データ分析が可能（主キーは暗黙化されている）
- 複数のデータを扱う場合には、データ毎の主キーとそれらの間の関連性を把握する必要がある
- 機械学習では、見過ごされ勝ちな主キーに光をあて、それと意識との関連を検討することから進展が期待できる

# 意識のEntification仮説：

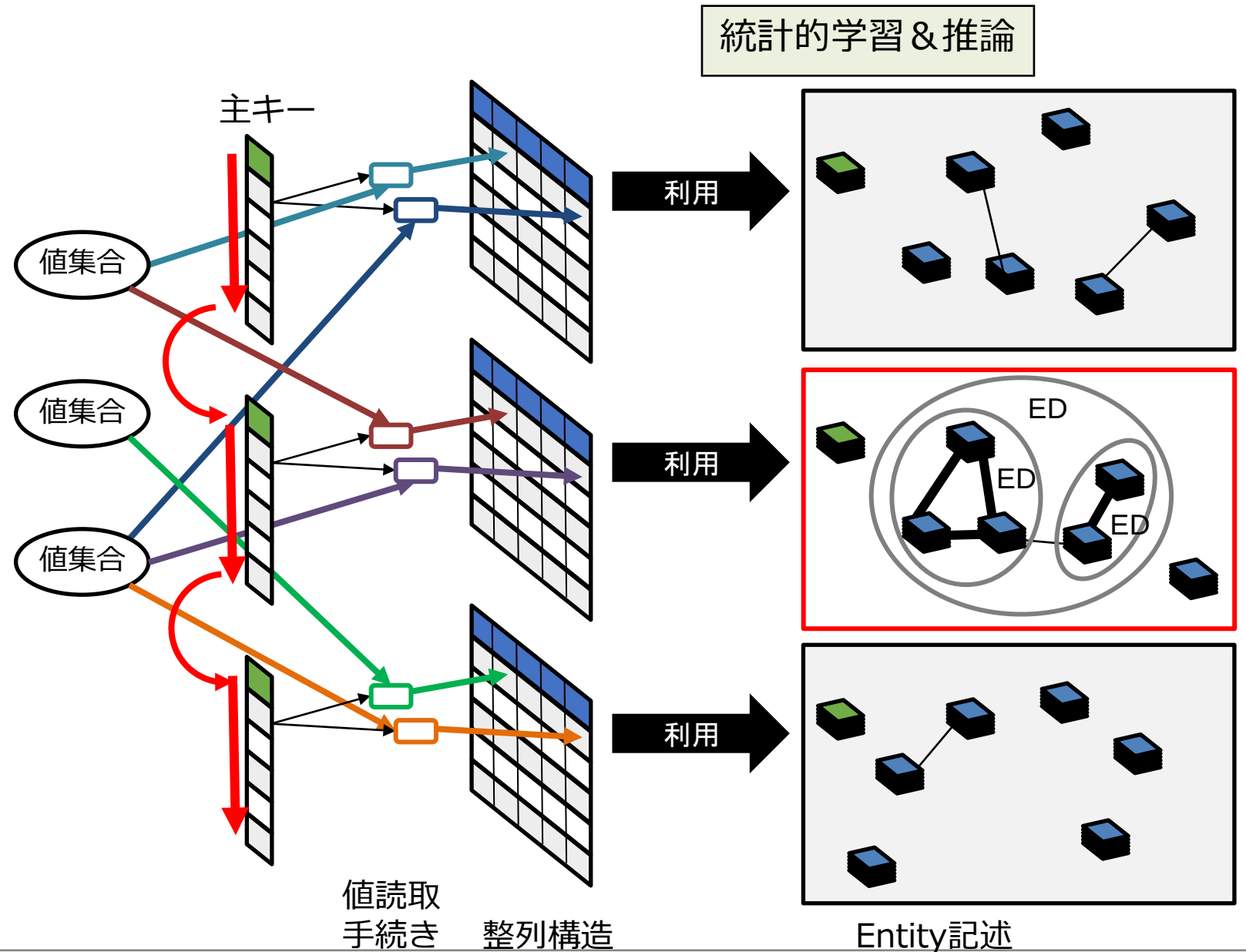
意識とは、ある**主キー系列**で制御された複数の読取手続きから得られるデータ構造（整列構造）上において、何らかの秩序に支配される範囲を特定可能なモデルを認識しようとする処理である



# 意識は、様々な主キーを渡り歩くであろう

脳内においては、  
固定的および動的  
に作られる整列構  
造に対応する主キ  
ーが存在する。

意識は、そうした  
主キーの間を渡り  
歩くのではないか  
と想像している。



# アジェンダ

---

1. 世界モデルにおいて存在を記述する意義
2. Entification処理： 存在を見出す処理
3. 意識のEntification仮説
4. Entification仮説の位置づけ
5. 脳に学ぶEntification機構の探求に向けて
6. まとめ

# 意識研究の地図

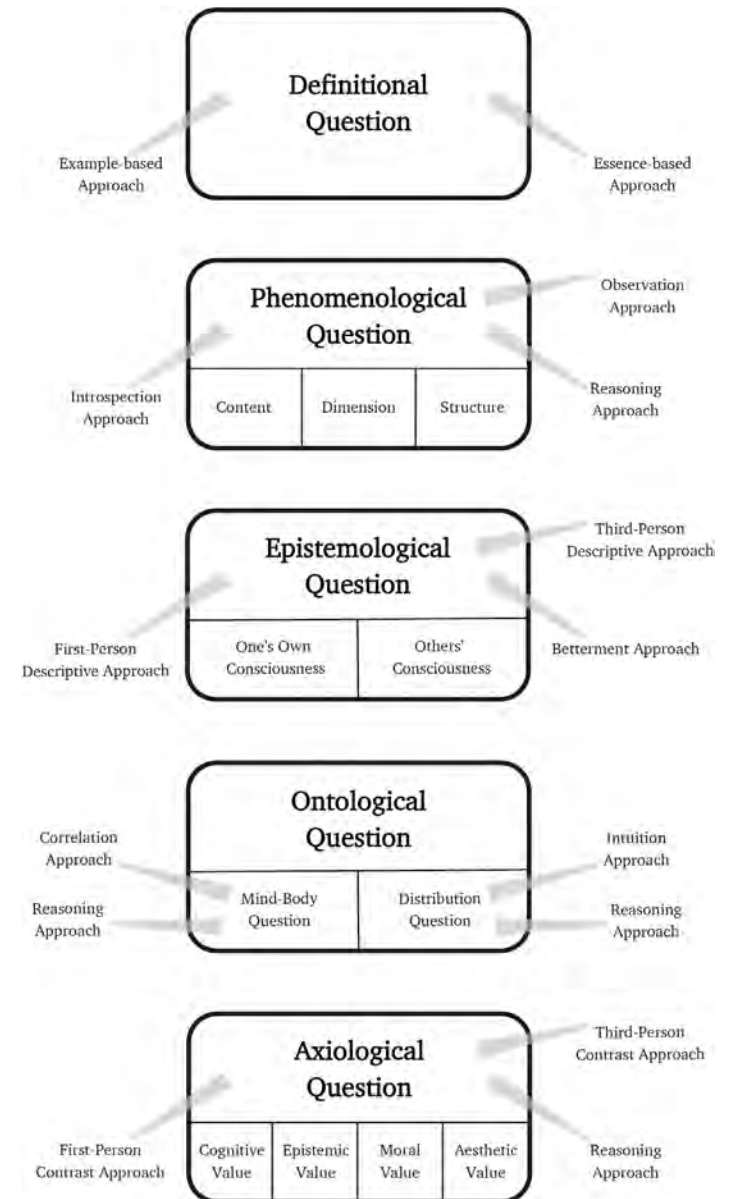
- ここ30年程で意識研究は急進展するも全体の見通しが悪い
  - 意識研究の学際化が進み、異なる学問領域で様々な理論が提案
  - それら理論間の関係が不明瞭であり、比較検討が難しい
- 「問いとアプローチ」を中心とした意識研究の整理
  - 意識理論の体系的な分類(地図)が可能になると期待
  - 定義論、現象論、認識論、存在論、価値論の五つの問い
  - 個別の意識の理論がどの種の問いにどの手法を用いて答えているかを明らかにした

[Niikawa, 2020] Niikawa, T. A Map of Consciousness Studies: Questions and Approaches. Front. Psychol. 11, 530152 (2020)

# 意識地図：意識をめぐる五つの問い

- 定義論 (Definition) :  
「意識」という語をどう定義すべきなのか？
- 現象論 (Phenomenology) :  
意識主体からみたとき、意識はどのようなあり方をしているのか？ (一人称的／三人称的)
- 認識論 (Epistemology) : 【方法論】  
私達はどのように意識について知るのか？
- 存在論 (Ontology) :  
意識は世界のうちにどのように位置づけられているのか？
- 価値論 (Axiology) :  
意識にどのような価値があるのか？ (一人称的／三人称的)

[Niikawa, 2020] Niikawa, T. A Map of Consciousness Studies: Questions and Approaches. Front. Psychol. 11, 530152 (2020)



## (1) 定義論 (Definition) :

【問い】 意識という語をどう定義すべきなのか？ (研究の出発点において足並みをそろえるために)。

- 意識を情報処理として定義するEntification仮説は、意識の本質を選び出すことで定義する、本質的アプローチ
  - 典型的な意識経験から共通点を抜き出す実例アプローチではない
- (Entification仮説に近い)既存の定義論からの意識研究
  - 自然種である [Bayne, The unity of consciousness, 2011]
  - 実在を自然変換から把握する[西郷,<現実>とは何か、 2019]

※ 本質的アプローチでは、存在との関連付けが共通である

※ 「**主キー**」の導入は、Entification仮説に固有と思われる

## (2) 現象論 (Phenomenology) :

【問い】 意識主体からみたとき、意識はどのようなあり方をしているのか？

- 意識の内容と種類

- Entification仮説では、外界の事物のみならず思考のメタ認知や自己などのEntity記述を含むため、内容や種類は多岐にわたる

- 意識の構造

- Entification仮説は、主キーを起点としてEntity記述を見つけ出そうとする処理を意識とみなすプロセス的な意識理論
  - 知覚・情動・思考などのまとまりを持つ存在者に注意を向ける志向性
  - 着目すべきEntity記述がまだ世界の中で特定されていない処理段階でも、探索する処理自体は存在するためそこには意識があるとみなす。

### (3) 認識論 (Epistemology) : 【方法論】

【問い】 私達はどのように意識について知るのか？

(一人称) 自分自身の意識のあり方の分析、 (三人称) 他者の意識のあり方についての分析

- 覚醒しているが存在を探索する活動(つまりEntification)を行っていない状態についての内観研究
  - 「無の境地」 (雑念を排除し、完全にものに執着しなくなった心持ち) に達した人にインタビューなどを行う方法である。
  - ただし、自己を認識することもEntificationと思われる点が課題

## (4) 存在論 (Ontology) : 説明問題

【問い】 意識は世界の内にどのように位置づけられてるか？  
(説明問題)なぜその相関が成立するのかを問う。

これに対しては、同一性、随伴現象説、中立一元論、観念論などがある

- 現段階でEntification仮説は、意識をEntification処理とみなすという、**同一説**の立場をとっている。
- その評価方法としては、後述するように意識にかかわる既存の研究やアイディア（クオリア、アクセス意識、自己認識）と整合性を議論することがあげられるだろう。

## (4) 存在論 (Ontology) : 相関問題

【問い】 意識は世界の内にどのように位置づけられてるか？

(相関問題) 意識のあり方と相関している神経活動を特定する。

- 特定の意識内容に十分な最小限の神経活動を特定する
  - それらは脳内のどこにあるかを特定する
  - そうした意識内容に対して情動的・機能的意味を与える
- Entificationとしての機能を備えたモデルの挙動が、脳の解剖学的構造とタスク実行時の神経活動と対応づける研究
    - Entification処理を実行する計算モデルの研究開発の進展後に行われる

## (5) 価値論 (Axiology) :

【問い】 意識にどのような価値があるのか？

- 機能的/認識論的価値(システム自身有能力)
  - Entity記述を獲得することで世界を効率的に記述
  - 言語的コミュニケーションの基盤
  - システムの行動決定に有益
- 道徳的/美的価値(外部からそのシステムをとらえた場合)
  - Entificationの定義それ自体からこれら価値を導くことは難しい
  - ヒトとは異なる実装と振る舞いを示す意識に対して道徳的・美的価値をどのように認めるかという議論の土台を提供しうる

# Entification仮説の意識地図上の位置まとめ

|                |                                                                   |         |     |                                                                  |
|----------------|-------------------------------------------------------------------|---------|-----|------------------------------------------------------------------|
| 意識<br>研究<br>地図 | 5つの問い                                                             | 問いの詳細   |     | Entification仮説                                                   |
|                | 定議論 (Definition):<br>「意識」という語をどう定義すべきなのか？                         | 本質アプローチ |     | 意識を情報処理として定義する(特に主キー系列)                                          |
|                |                                                                   | 実例アプローチ |     |                                                                  |
|                | 現象論 (Phenomenology):<br>意識主体からみたとき、意識はどのようなあり方をしているのか？(一人称的／三人称的) | 内容・種類   |     | 外界の事物のみならず思考のメタ認知や自己などのEntity記述を含むため、内容や種類は多岐にわたる(種・事物を数えるなど)    |
|                |                                                                   | 構造      |     | 主キーを起点としてEntity記述を見つけ出そうとする処理を意識とみなすプロセス的な意識理論                   |
|                | 認識論 (Epistemology):【方法論】私達はどのように意識について知るのか？                       | 自己      | 記述的 | 覚醒しているが存在を探索する活動(つまりEntification)を行っていない状態についての内観研究              |
|                |                                                                   |         | 改良的 |                                                                  |
|                | 存在論 (Ontology):<br>意識は世界のうちにどのように位置づけられているのか？                     | 心身問題    | 相関  | Entificationとしての機能を備えたモデルの挙動が、脳の解剖学的構造とタスク実行時の神経活動と対応づける研究       |
|                |                                                                   |         | 説明  | 同一説: 意識にかかわる既存の研究やアイディア(クオリア、アクセス意識、自己認識)と整合性の議論は可能              |
|                |                                                                   | 分布問題    |     | 動物やAI内部の世界モデル内                                                   |
|                | 価値論 (Axiology):<br>意識にどのような価値があるのか？(一人称的／三人称的)                    | 機能      |     | 帰納推論の基盤、世界を効率的に記述、言語的コミュニケーションの基盤、システムの行動決定に有益                   |
|                |                                                                   | 道徳/美    |     | 間接的価値として、ヒトとは異なる実装と振る舞いを示す意識に対して道徳的・美的価値をどのように認めるかという議論の土台を提供しうる |

# 現象的意識とアクセス意識[Block, 2005]

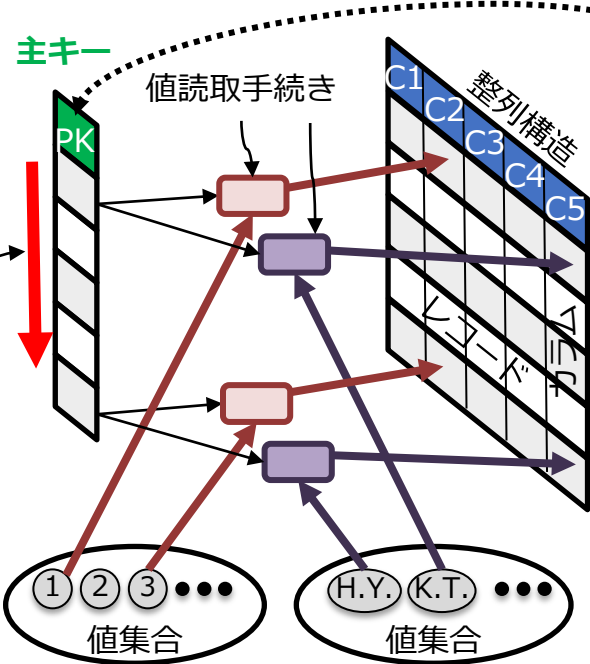
Entification処理に現れる概念との明確な対応付けは未だ不明

- 現象的意識：読書中に注視点周辺で報告できないが意識にのぼる部分
  - ある時点において認識しているEntity記述もしくは、それを含む整列構造に対応づくかもしれない。
  - クオリアはEntity記述のカラム(属性)と関連づくかもしれない
    - クオリアは現象的意識に含まれる個々の対象についての質的な性質
- アクセス意識：読書中において注視点の付近で単語として報告できる部分、各タイムステップで、行動／計画／想像／想起の条件となれる部分
  - 主キー系列と関連づけるのが自然

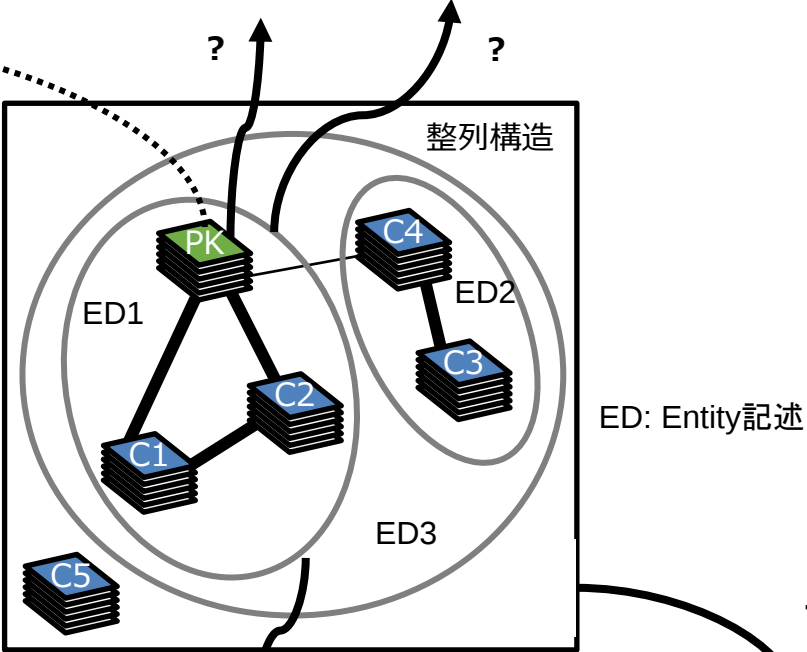
# 現象的意識とアクセス意識

意識の系列的な性質の基礎

**主キー系列**  
順序付けされた主キーの一部の要素



**アクセス意識：**  
読書中において注視点の付近で単語として報告できる部分、各タイムステップで、行動／計画／想像／想起の条件となれる部分



**現象的意識：**  
読書中に注視点周辺で報告できないが意識にのぼる部分

# アジェンダ

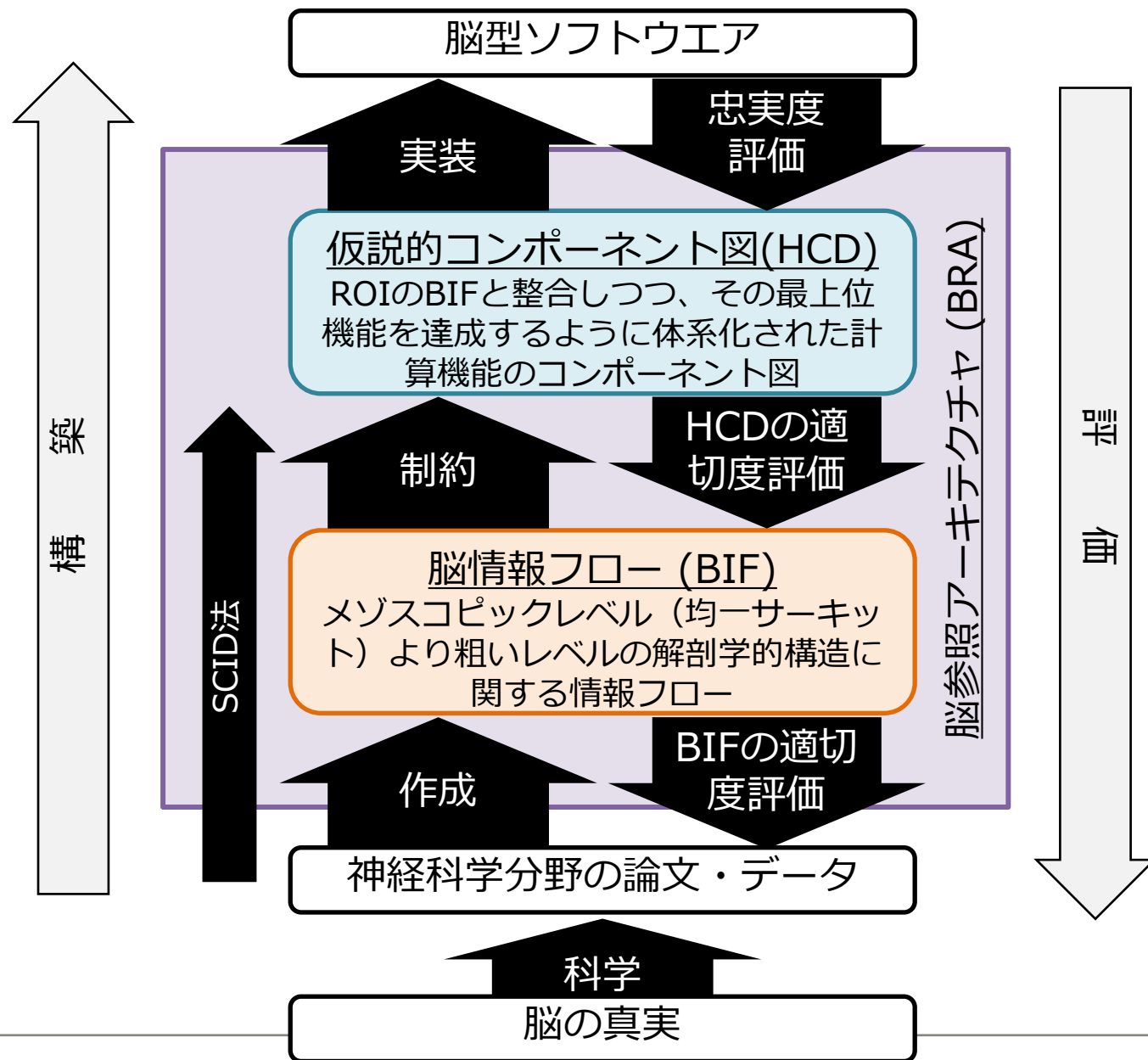
---

1. 世界モデルにおいて存在を記述する意義
2. Entification処理： 存在を見出す処理
3. 意識のEntification仮説
4. Entification仮説の位置づけ
5. 脳に学ぶEntification機構の探求に向けて
6. まとめ

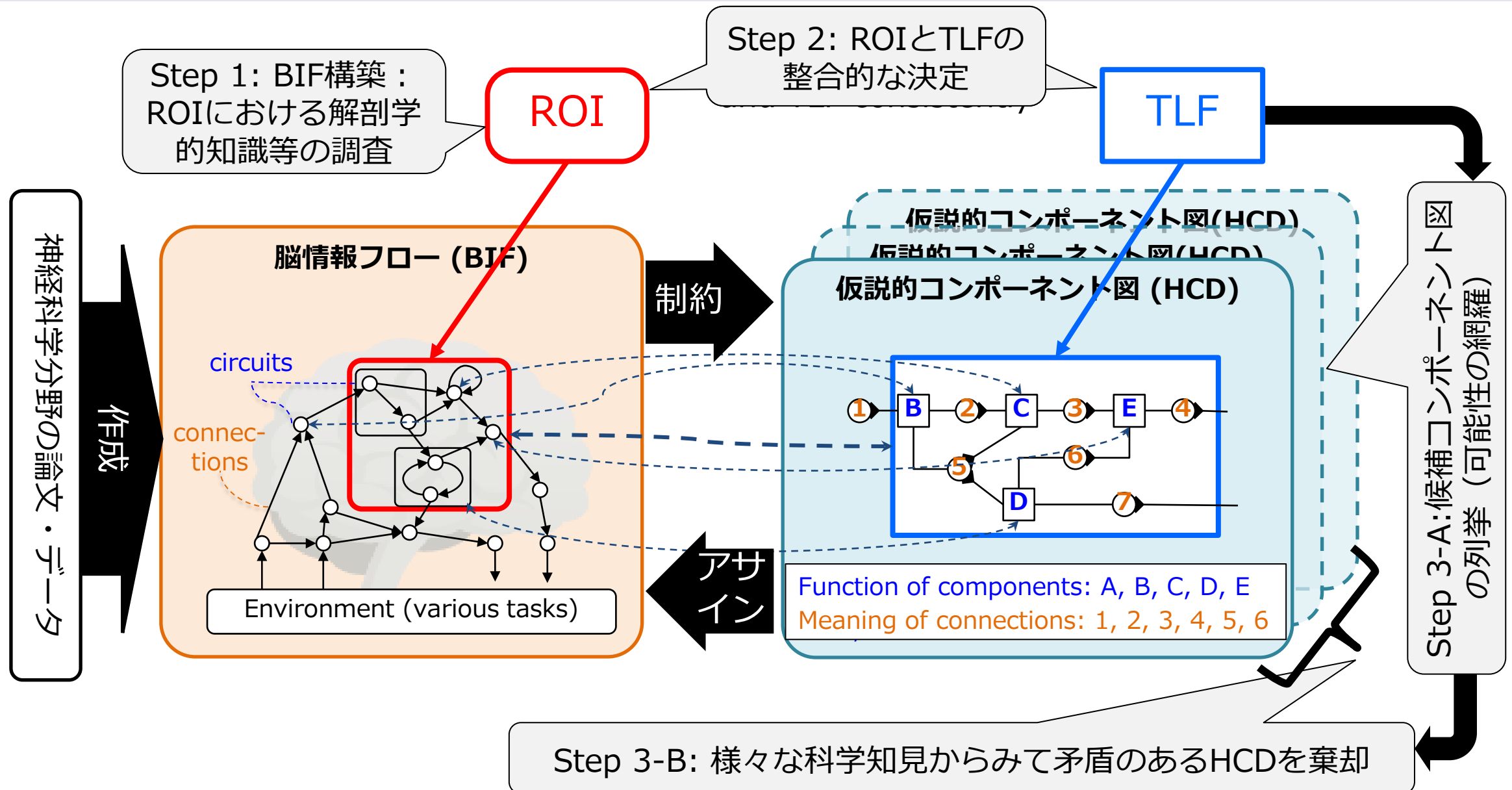
# BRA駆動開発

脳全体の神経回路を参照することで、人間の認知機能を再現するソフトウェアを構築する手法

Entification仮説の下では、進展著しい意識とその脳に関わる知見を参照したSCID法によりEntification機構の機能仮説を設計できるだろう。



# Structure-constrained Interface Decomposition (SCID)法

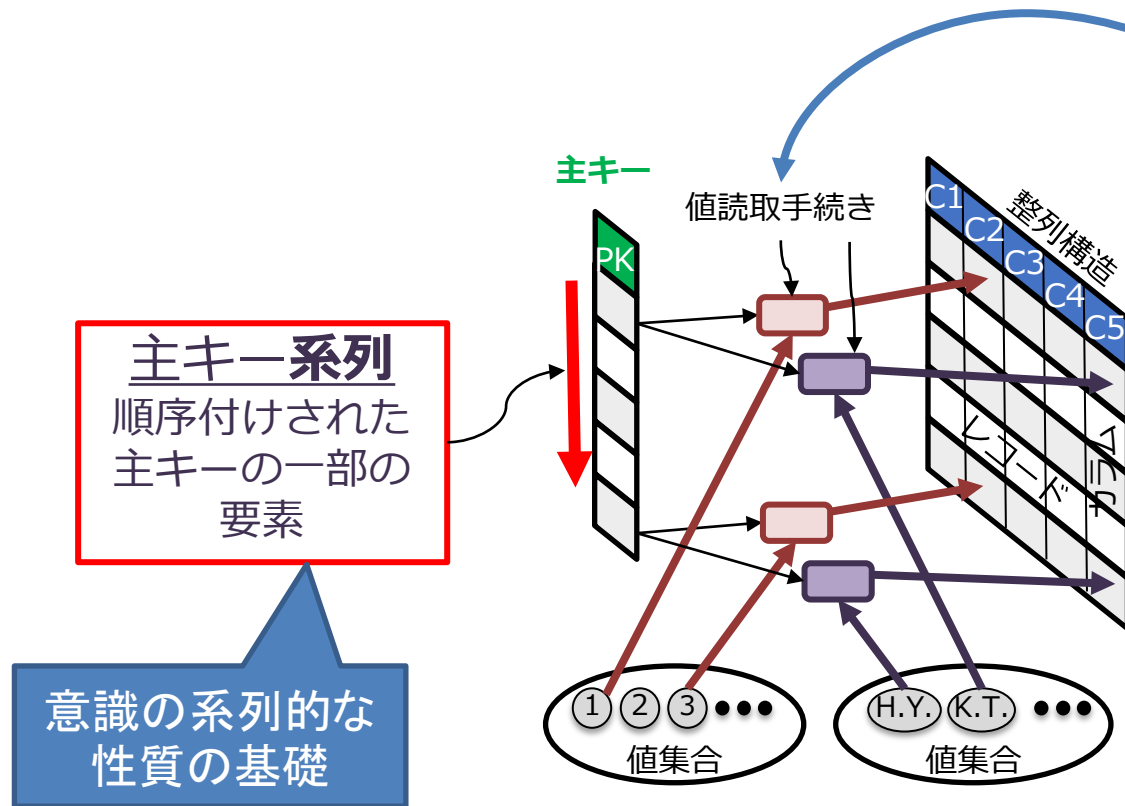


# SCID法の準備状況(Step.2)

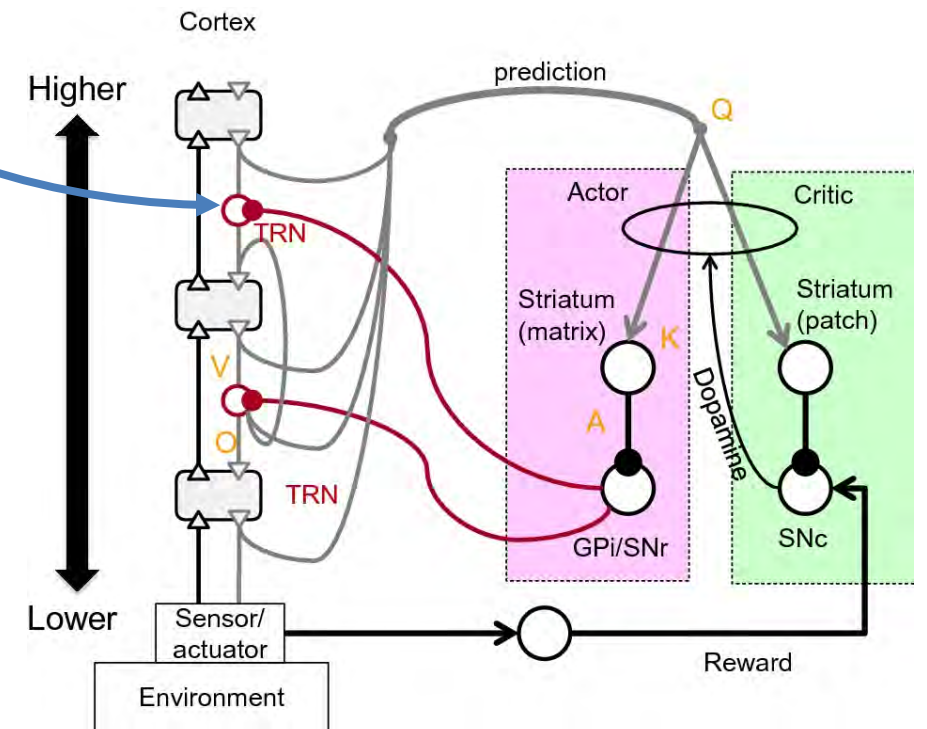
- 対象脳領域 (ROI)
  - 新皮質と視床:
    - 意識にかかわる脳器官として伝統的に着目されている
  - 大脳基底核:
    - 近年、視床-皮質結合のみに注目した意識障害モデルでは、実験データに適合しないことが指摘された[Crone, 2017]
- 最上位機能(TLF)
  - 意識のEntification仮説
  - 「様々な整列構造上における、何らかの秩序に支配される範囲を特定可能なモデルを獲得し認識する機能」

# Entificationのために大脳基底核が担う機能の検討

今後、値読取り手続きが、視床(TRN)のGating機能に対応づく可能性を探る予定



Yamakawa, H. (2020). **Attentional Reinforcement Learning in the Brain.** *New Generation Computing.*



# アジェンダ

---

1. 世界モデルにおいて存在を記述する意義
2. Entification処理： 存在を見出す処理
3. 意識のEntification仮説
4. Entification仮説の位置づけ
5. 脳に学ぶEntification機構の探求に向けて
6. まとめ

# まとめ

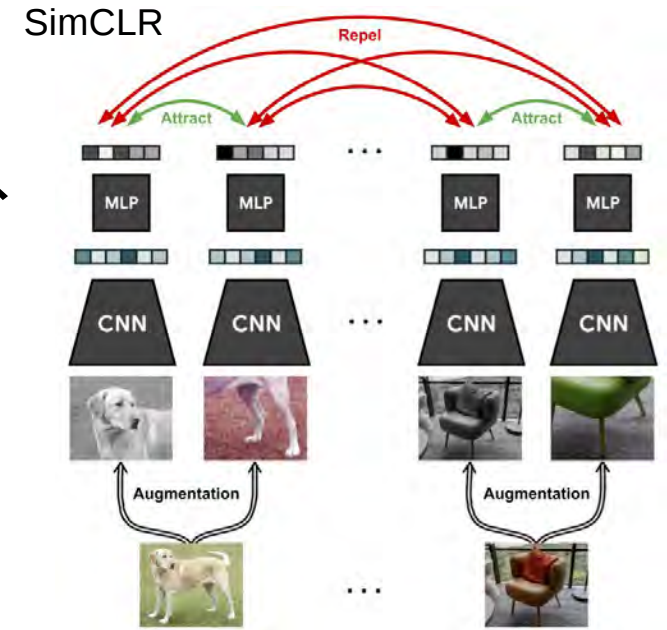
---

- 最近、外界からの情報を様々なEntity記述に変換するEntification処理の提案を行った。
- 主キーが系列化したものが意識と関連するとしたEntification仮説を作業仮説として設定した。
- 作業仮説を意識研究地図上で位置づけ現象的意識とアクセス意識との関連を議論した。
- Brain-inspiredなEntification研究の可能性について検討した

# おわりに

- AIに宿る意識とは？

- Entification仮説を前提とした意識は、AIシステムにおいては、Multi-threadのように複数  
が並行して動作するであろう。
- ヒトの価値において「意識をもつこと」を重  
視する立場をとると、未来に残すべきもの  
についての議論への影響がある。



Contrastive learningのような形で同一性の獲得をおこなえば並列処理になる。

- 無の境地

- 意識の本質が、存在を探し求め続けるEntificationであるならば、意識がある限り、あらゆる煩悩を消し去った「無の境地」に到達することが難しいのは自然な帰結であるかも知れない。