

意識のEntification仮説

○山川 宏^{1,2}、布川 絢子²、松尾 豊¹

(1. 東京大学、2. 全脳アーキテクチャ・イニシアティブ)

背景：意識研究の現状

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

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

科学的な意識の理論

- ✓ Global workspace 理論
 - ✓ 情報統合理論
 - ✓ 高階意識説
- ✓ 生物学的自然主義
- ✓ 予測誤差最小化理論
- ✓ エナクティブ・アプローチ

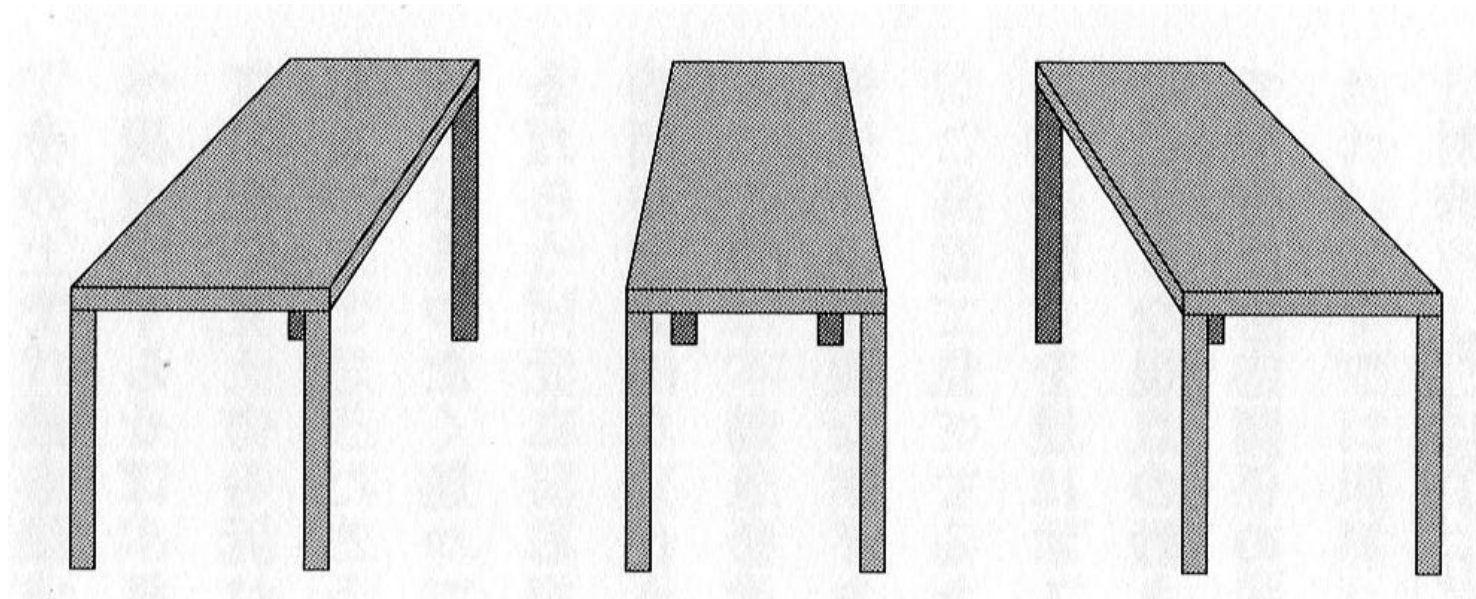
哲学的な意識の理論

- ✓ 心脳同一説
- ✓ 随伴現象説
- ✓ 中立一元論-性質二元論
 - ✓ 観念論
 - ✓ 汎心論
- ✓ 消去主義・錯誤説
- ✓ The Unity of Consciousness (Bayne)

Entityとはなにか？

(W.V.O. Quine,
1969)

No **Entity** without **Identity**

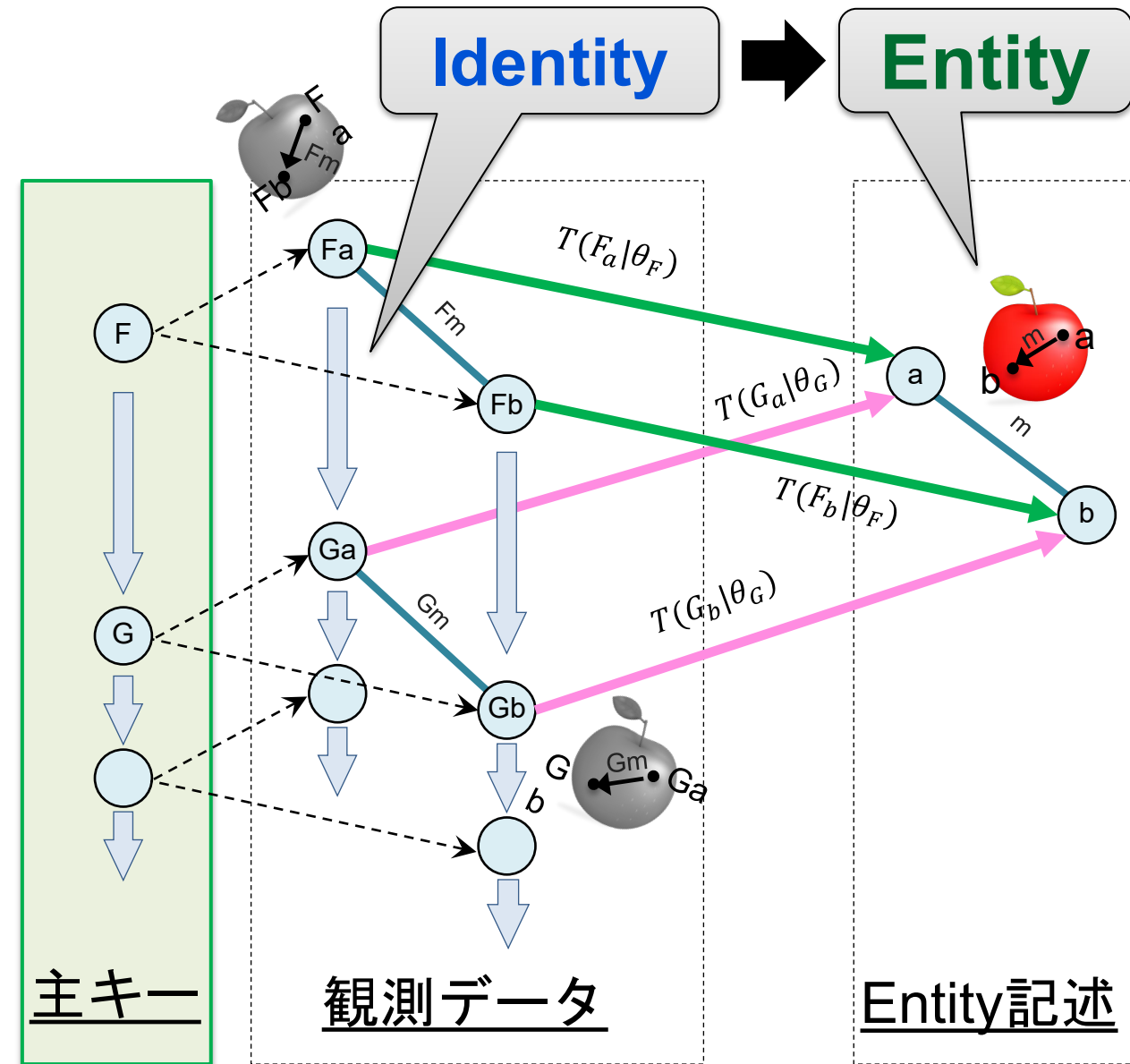


机の現れの変化

(西郷, 田口. 現実とは何か: 筑摩書房, 2019)

Entificationとは何か

Entificationとは、主キーで
統御された事例の集合上で繰
り返し出現する、空間的な対
称操作の下で同一となる部分
的なパターンを、少ないパラ
メータで制御される変換を通
じて縮約されたEntity記述を
得る処理である。



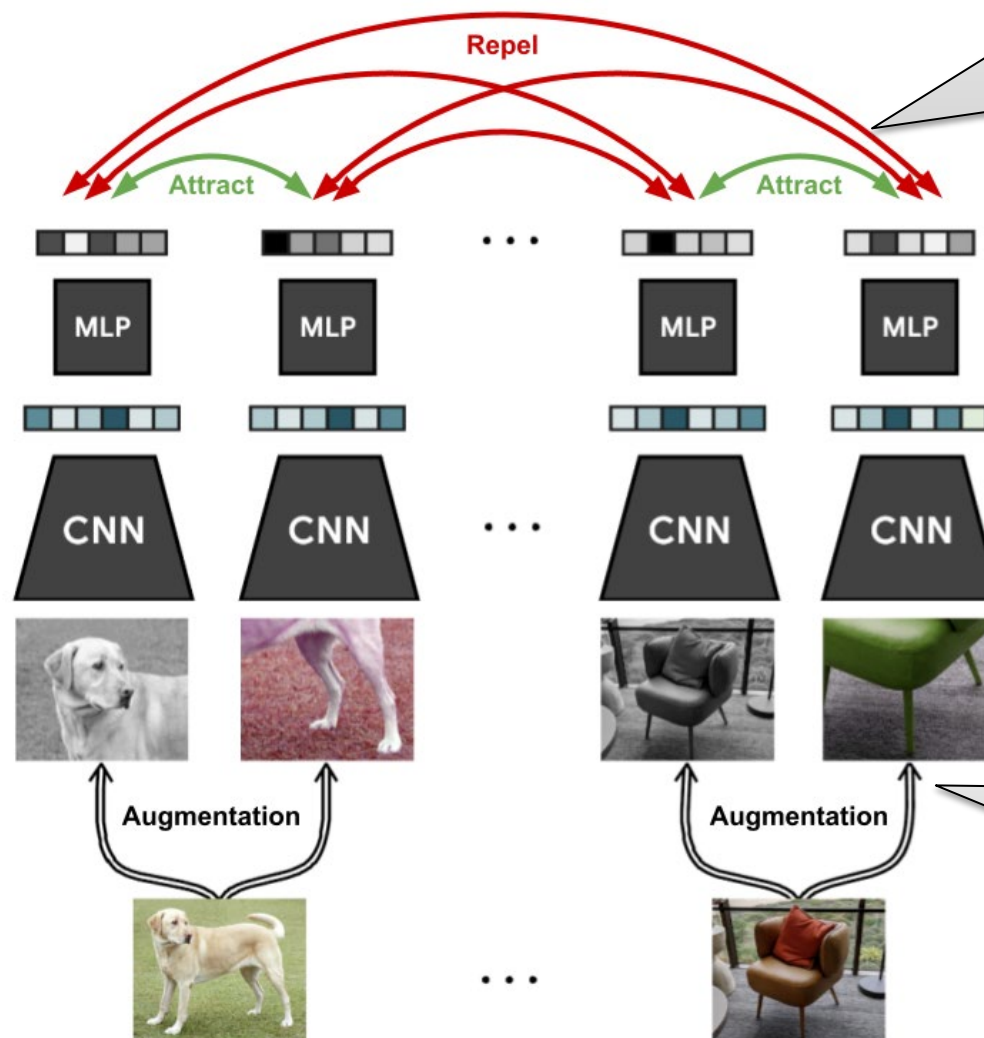
対照学習 (Contrastive learning)

SimCLR

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020).

A Simple Framework for Contrastive Learning of Visual Representations. In *Proc. ICML*

対照学習は、画像に対する自己教師あり学習の一種。現在は多くの研究で利用されている。



内部表現において、
同一の個物に由来するものは近づけ、そうでないものは遠ざける

Entity

**No Entity
without
Identity**

加工して複数化された
同一の個物の画像

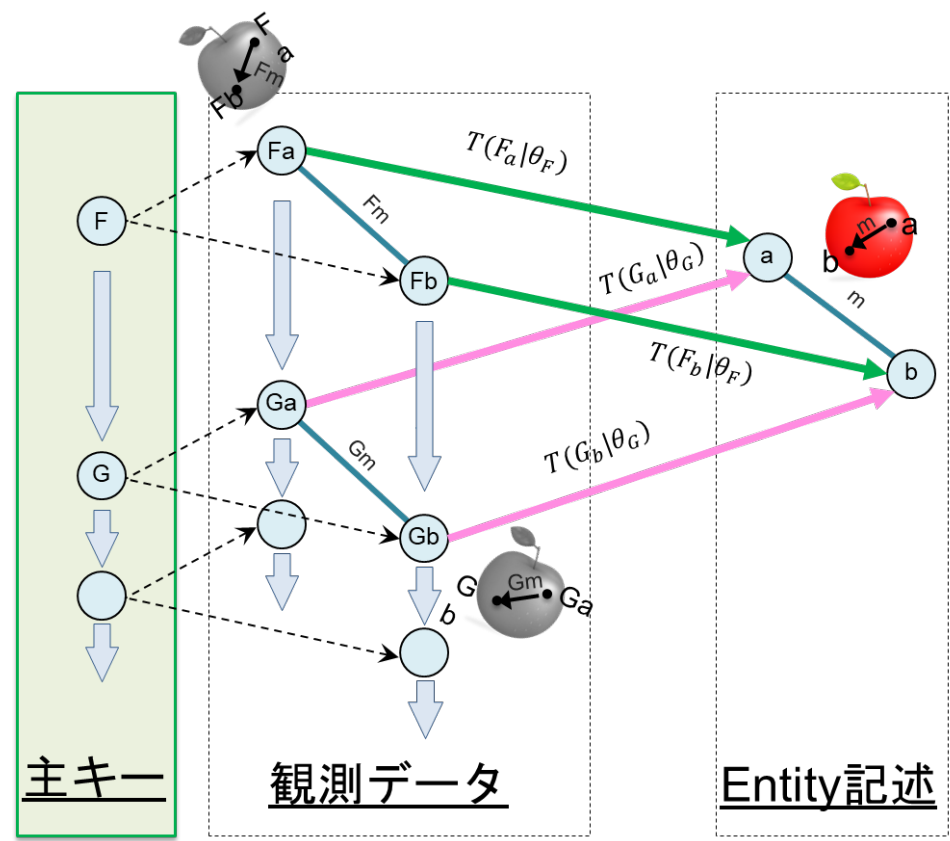
Identity

整列注意機構：現実的なラスタ形式でのEntification

ベクトル形式Entification

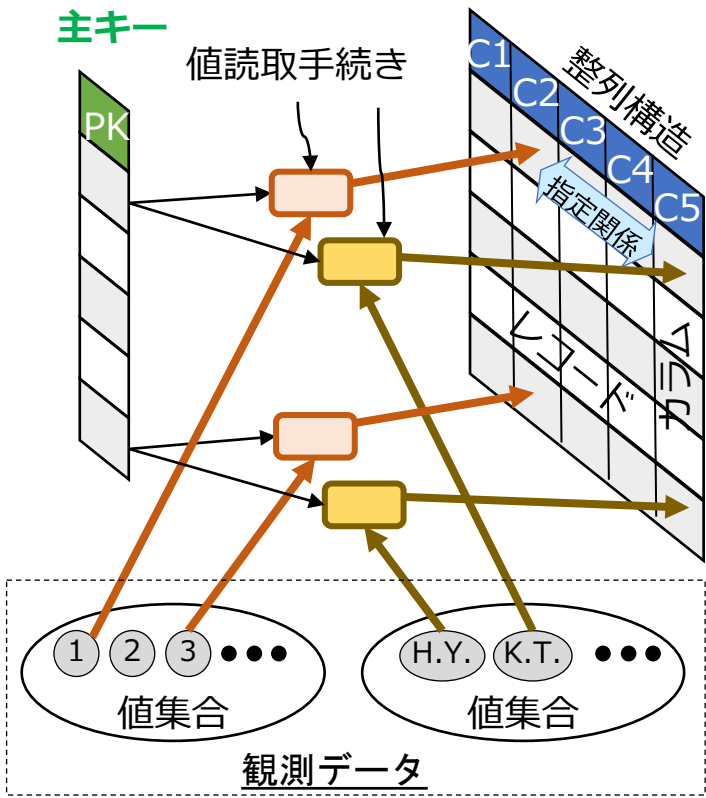
ドローソフト、
多くの印刷物

ペイントソフト、
写真／動画



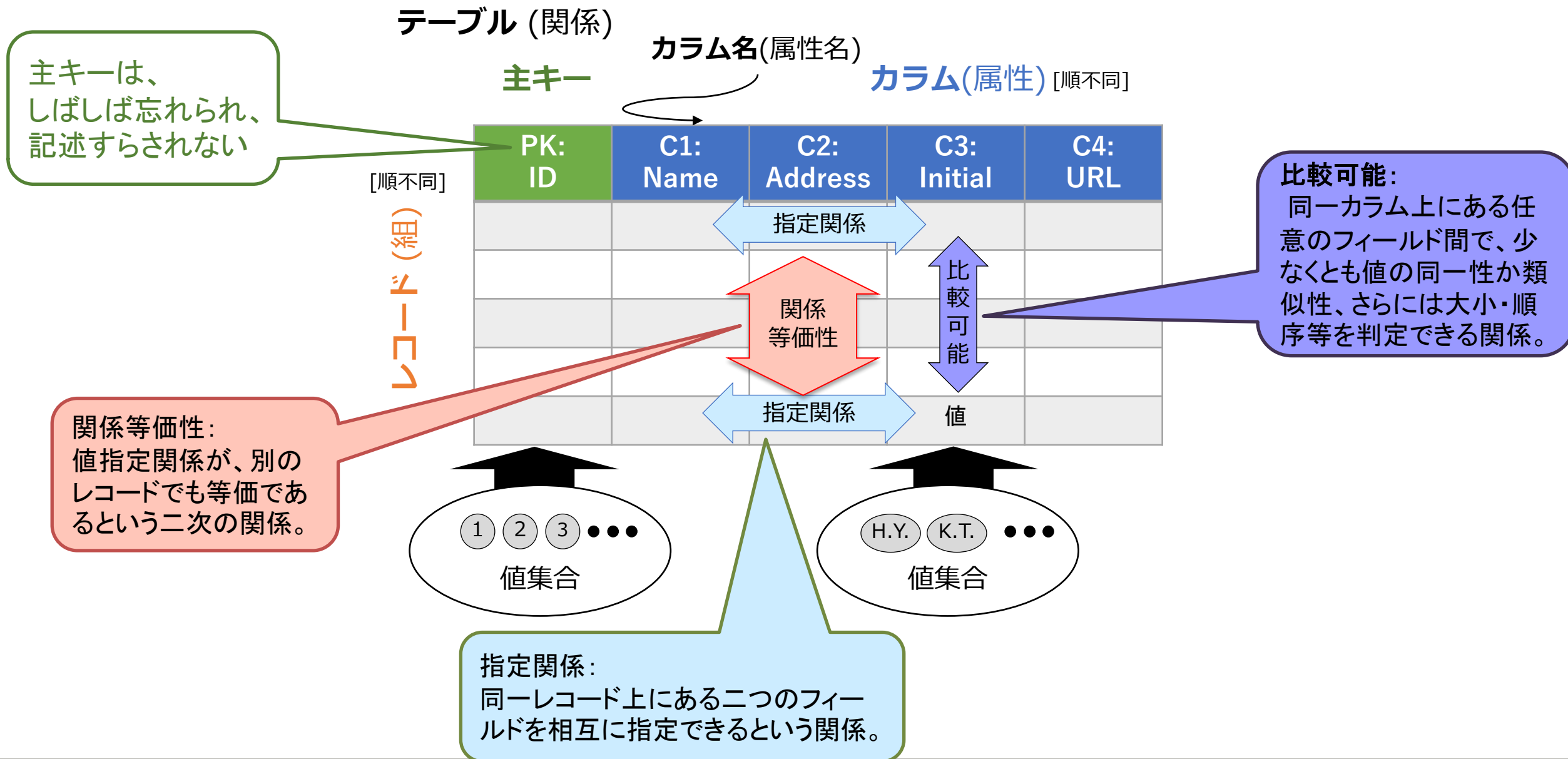
ラスタ形式Entification 内の

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



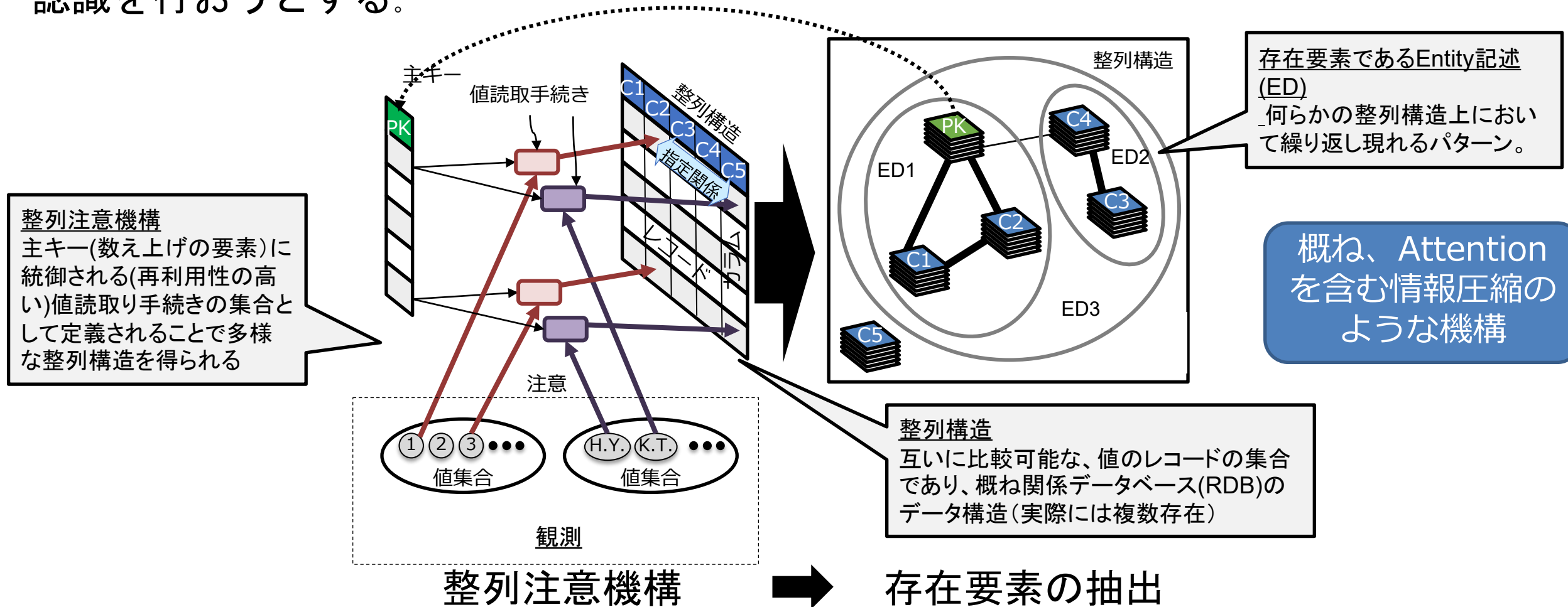
世界の観測における指定関係(F_m, G_m)は、主キーに統合された値読取り手続きに反映される

同一性(や帰納法)の前提となる整列構造(～RDB)と主キー



Entification(ラスタ形式)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。



意識のEntification仮説：主キー系列で制御されるEntification処理

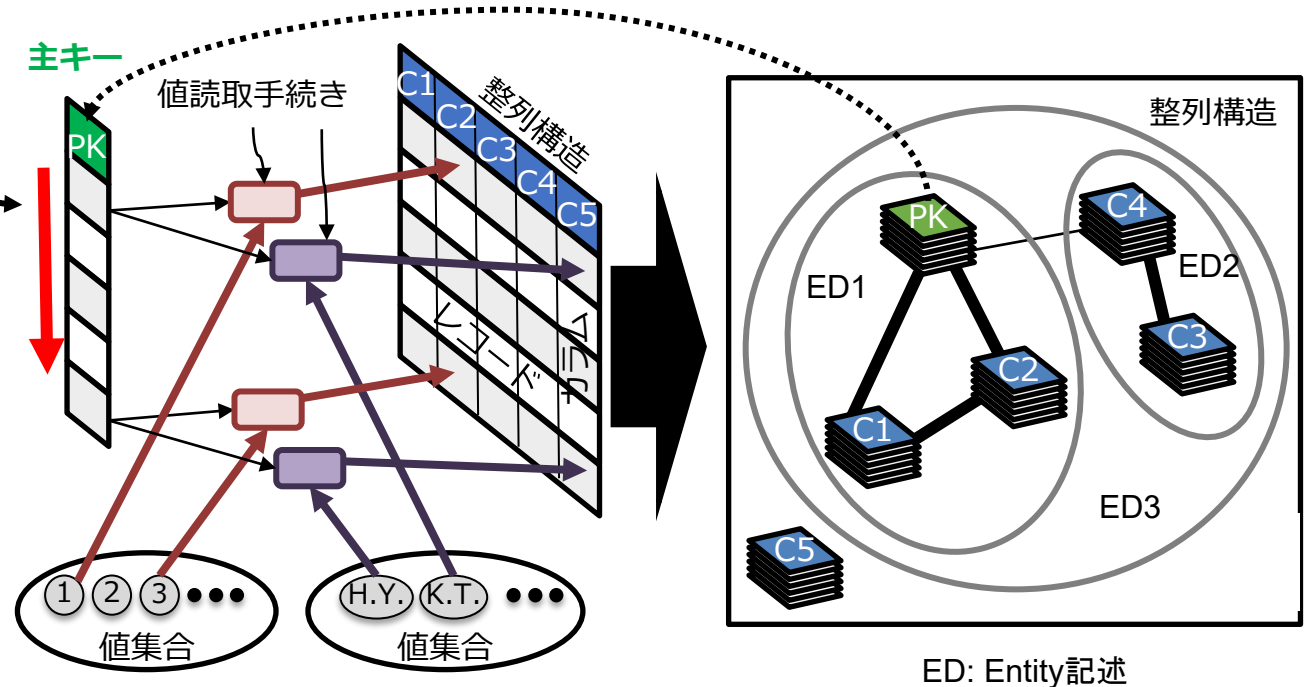
仮説：意識とは、ある**主キー系列**で制御された複数の読取手続きから得られるデータ構造(整列構造)上において、何らかの秩序に支配される範囲を特定可能なモデルを認識しようとする処理である

主キー系列

順序付けされた主キー

脳内でEntificationを実装する際の制約

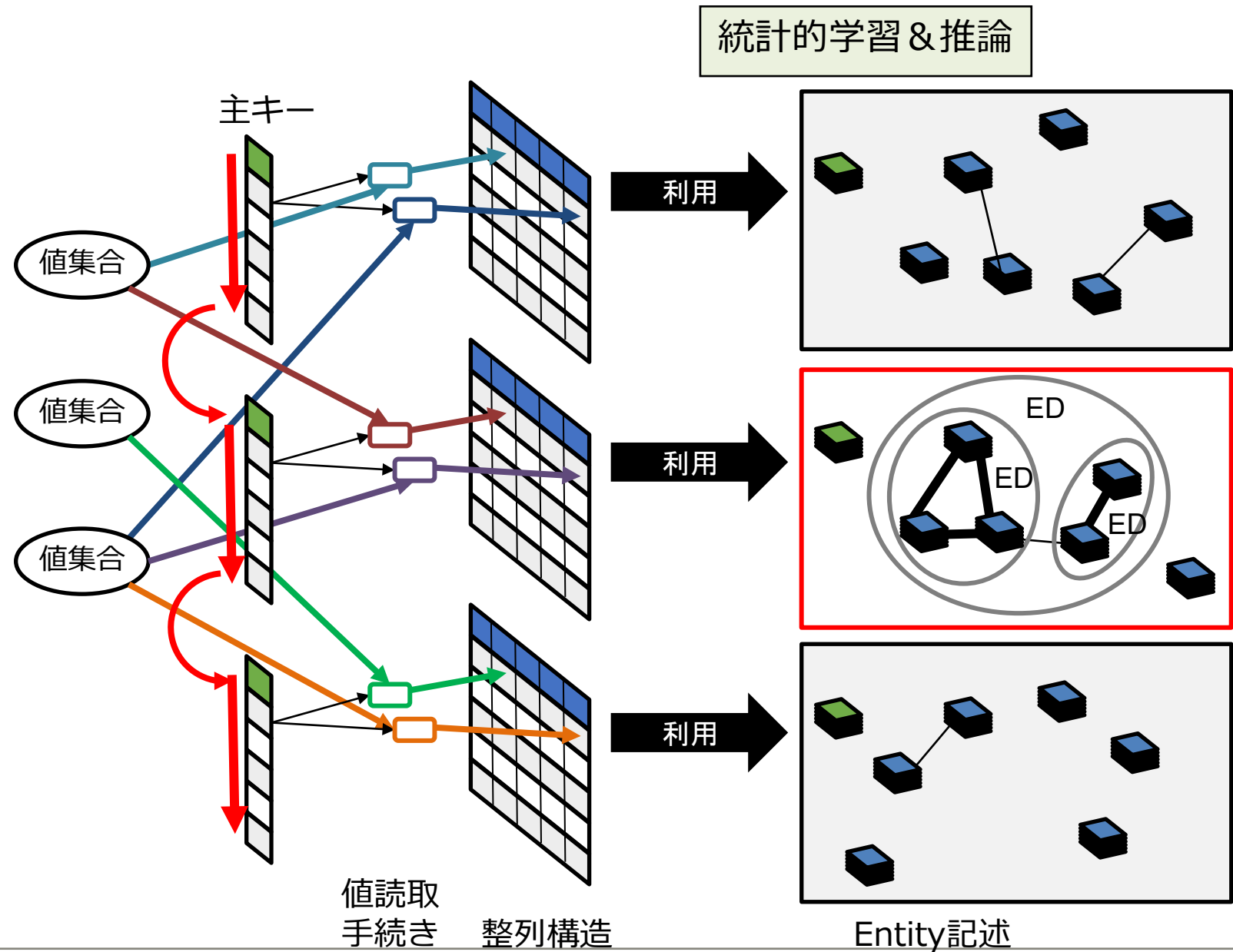
- Entity記述は新皮質上に表現される
- ある時点において表される内容は一つである
- 隣接する時刻間のみで表現内容が比較可能



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。



Entificationに必要な3つの処理とその機械学習での状況

必要とされる 処理	機械学習の 進展状況	個体(Particular) 観測されるEntity		普遍者(Universal) 抽象的なEntity
		静的 (Intendent Continuant)	動的 (Process)	
		この椅子、岸田総理	JSAI2022、	
主キー決定: 数え上げの 単位の決定	あまり整理されてい ない。(二重文節化 などはある)	不要: 同時刻の情報を利用 するように設計す るだけでよい	必要: 時間幅の変化へ の対応に課題。 音素、エピソードなど	必要: 個体や普遍者を数 える。 (外延)
対称性を利用した縮約	対照学習で基本的 に対応可能	必要	必要	必要
個体の 切り出し	セマンティック・セグ メンテーション+α	必要	必要	不要

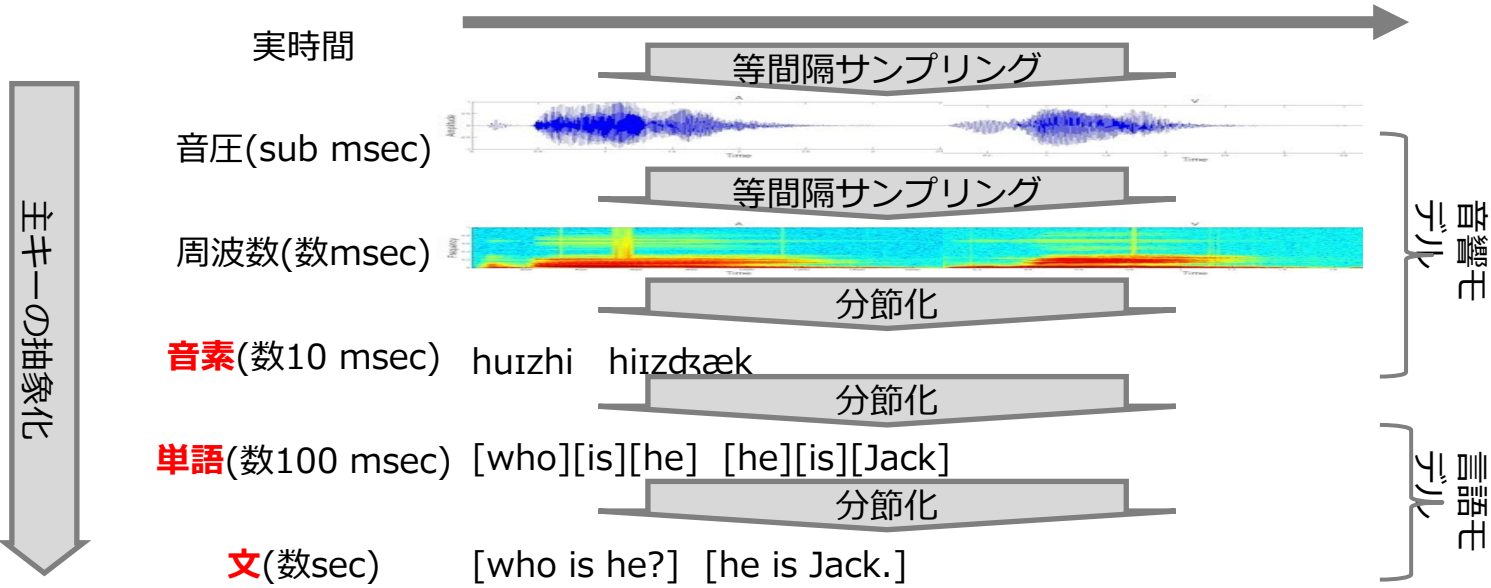
一般性が高く、時間的な切り出しを行う技術は未開拓

多様な種別において、カウントするための主キー

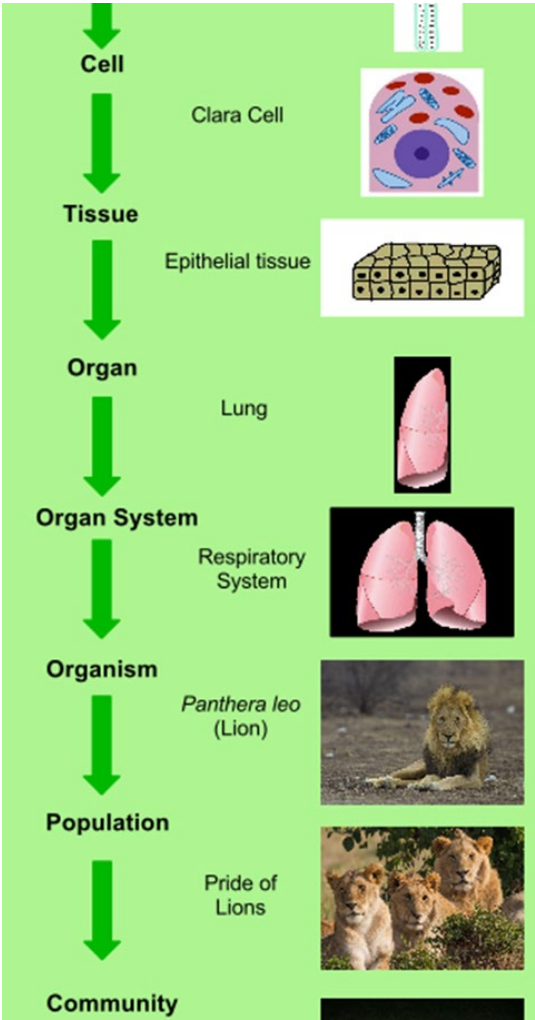
主キーは、種別毎に数えられる事物を指定する

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

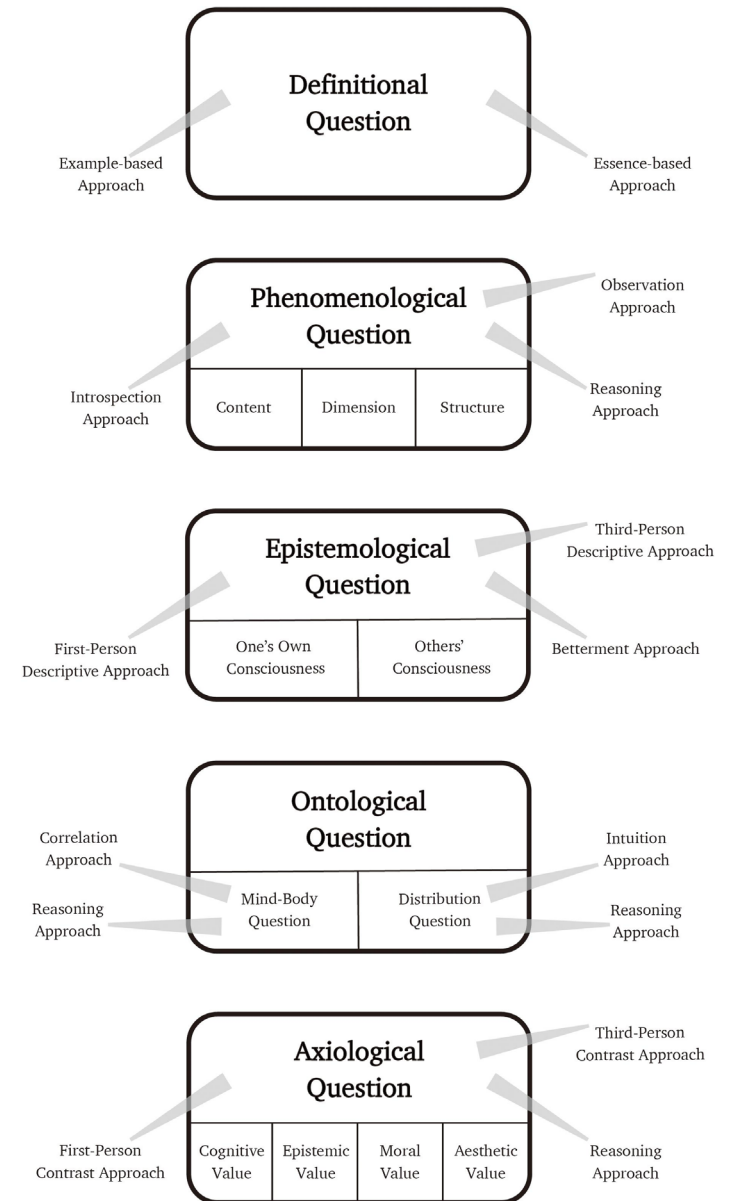


出典：https://en.wikipedia.org/wiki/Biological_organisation

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)



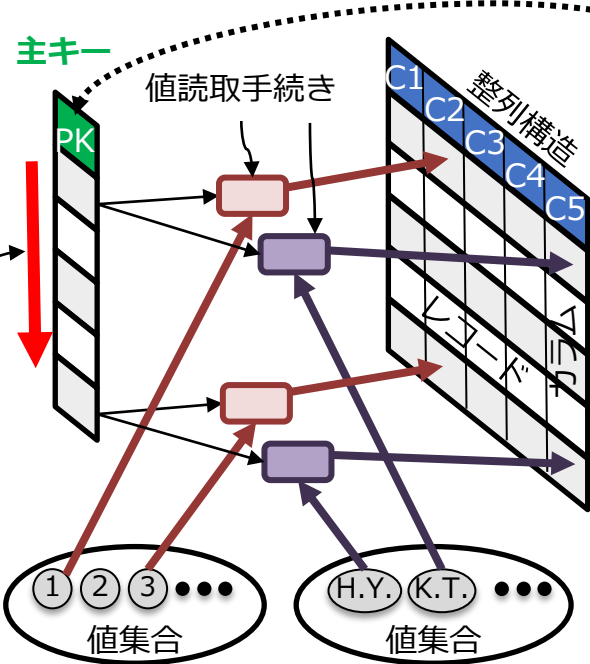
Entification仮説の意識地図上において位置づける

意識研究地図	5つの問い	問いの詳細		Entification仮説との関連
	定議論 (Definition): 「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Entificationという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): 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): 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): 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

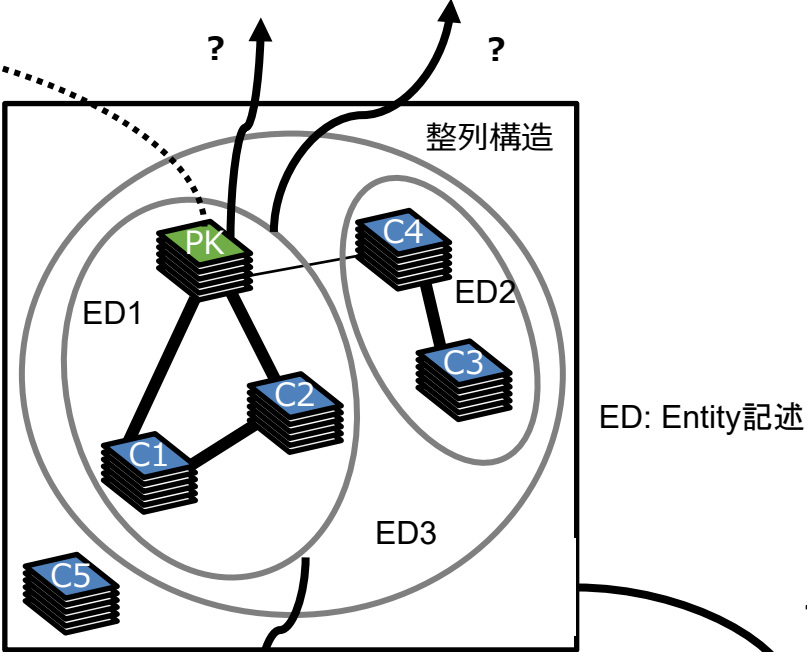
現象的意識とアクセス意識

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

まとめ

- 意識研究はここ30年程で急進展するも全体の見通しが悪い
 - 存在から意識を議論する、先行研究があるが機構の提案は見当たらない
- 視覚の機械学習を参照してEntification(存在を見出す処理)を提案中
- 画像(ラスタ形式)からのEntification機構は、整列注意機構を含む
- Entificationにおいて、事例の数え上げを制御するの主キーの役割
 - 動画中の物体認識等では、主キー系列はフレームレートであり着目されない
 - 時間幅をもつエピソード、個物の比較では、主キーの決定が課題になる
- 意識のEntification仮説を提案
 - 「意識 = **主キー系列**で制御されるEntification処理」
 - 本仮説は、本質的定義に基づき、機能的な価値は明確
- 基底核が視床をGatingする形でモデル化を検討中
 - モデル化が進むと脳神経活動の相関を議論が可能となる

Entification、脳モデリング、世界モデル、深層学習などにご興味のある方、お気軽にお声がけください

論文特集「知能のための世界モデル」@NGC

Call for Papers on World Models for Intelligence

New Generation Computing, Springer Nature

New Generation Computing 特集「知能のための世界モデル」論文募集

<https://www.springer.com/journal/354/updates/20324432>



Focus: Since the rapid development of deep learning starting in the late 2010s, research on learning data-based models of the external world and their use for cognition and behavioral tasks has greatly spread. Given this, this special issue deals with topics related to artificial intelligence research (especially machine learning) in which systems demonstrate their intelligence capability by using acquired models of the world. Specifically, including description methods and description languages for modeling various aspects of the world, learning methods and their algorithms for abstracting and extracting various aspects of the world, as well as computational techniques for making various inferences and predictions using the obtained models. There are also contr testbeds and evaluation suites, such as simulators for training and testing, and methods for evaluating intelligence. We welcome the emphasis on research that tries to tackle various aspects that have not realized despite the advances in deep learning, and the use of knowledge from cognitive science and neuroscience for this purpose. On the other hand, we also welcome contributions that organize and summarize the nature of the advances in intelligent technology through the advancement of deep learning. We hope many contributions to this special issue will help open the door to intelligence as versatile as human intelligence, and the limitations of traditional symbolic AI and deep learning that have been developed so far.

Topics of interest:

Algorithmic information theory, Brain-inspired intelligence, Causal reasoning, Cognitive robotics, Sense learning, Consciousness mechanism, Deep reinforcement learning, Disentanglement, Distillation, process theory, Equivariance, Equivalent structure, Free Energy Principle, General intelligence test assessment, Generative models, Learning environment platforms, Imagination, Imitation Learning, bias, Intelligent Architecture, Intention estimation, Inverse reinforcement learning, Language, Metacognition, Multi-modality, Object file theory, Ontology, Physics Simulator, Planning, Predictive, Probabilistic model, Representation Learning, Real time system, Self-awareness, Self-supervised, Situation decomposition, Versatility of animal intelligence, Working memory, World Model



Guest Editors:

Yutaka Matsuo, Professor, The University of Tokyo
Mingbo Cai, Assistant Professor, The University of Tokyo
Hiroshi Yamakawa, The University of Tokyo
Yusuke Iwasawa, The University of Tokyo
Masahiro Suzuki, The University of Tokyo
Wataru Kumagai, The University of Tokyo

Important Dates:

Paper Submission deadline	October 31, 2022
Authors' notification	January 31, 2023
Revisions due	March 31, 2023
Final decision	May 31, 2023

Accepted papers will be made available online the Springer website shortly after acceptance.
The publication date of the printed version will be determined when the Special Issue is finalized.

