Qualia as a Social Sharing Tool (QSST)

——成立条件・消失条件・説明ギャップ直観の理論

要旨 (Abstract)

本稿は、クオリアを個体内部の現象としてではなく、他者との協調行動において要請される社会的共有ツールとして再定義する。本理論によれば、クオリアとは、(i) 多主体協調の必然性、(ii)
内部状態の不可逆的圧縮、(iii) 当事者間での較正不能性（Ground Truthへの相互アクセス不能性）という三つの構造的制約のもとで、行動予測のために収束した機能的インターフェースである。この枠組みにより、主観的経験の「私秘性」や「説明ギャップ」は、認識論的限界ではなく、情報圧縮と社会的断絶が生む構成的必然として説明される。さらに本理論は、社会的較正圧や共有帯域の操作によってクオリア直観の強度が系統的に変動するという、反証可能な予測を提供する。

目次

1. 序論 (Introduction)

意識のハードプロブレム（Chalmers, 1995）、すなわち物理的な情報処理がいかにして主観的な質感を伴うのかという問いは、現代の心の哲学と科学における未解決の課題である。物理主義（Physicalism）は脳機能の解明において著しい進展を見せているが、依然として説明ギャップ（Levine, 1983）と呼ばれる現象的直観と物理的説明の乖離は解消されていない。

この問題に対し、現状の議論は大きく二つの立場に二分され、膠着状態にある。
第一の立場である実体論（Realism / Naturalistic Dualism）は、現象的直観の不可還元性を重視し、意識を物理法則から独立した実体、あるいは宇宙の基本原理として位置づける（Chalmers, 1996）。このアプローチは直観を擁護する一方で、クオリアの機能的役割や発生起源についての説明力を欠き、結果として検証困難な汎心論（Panpsychism）へと接近する傾向にある。

第二の立場である消去論（Eliminativism / Illusionism）は、クオリアの実在性を否定し、脳が生成する「ユーザー・イリュージョン」であると主張する（Dennett, 1991; Frankish, 2016）。この立場は物理的世界像と整合的であり、機能的説明に優れる。しかし、なぜ生物はそのようなコストのかかる錯覚を進化させたのか、そしてなぜその錯覚がこれほど強固な実在感（Reality）を伴い、否定しがたい直観として経験されるのかという点について、十分な構造的説明を提供していない。

したがって、現在の閉塞状況は、双方が説明の射程を部分的側面――実体論は現象的直観、消去論は物理的機能――に限定していることに起因する。真に求められているのは、どちらか一方を否定する理論ではなく、物理系である脳が、必然的に「説明不可能な質感」という直観を持つに至る構造的メカニズムを解明することである。

本稿は、この課題に対し、クオリアを社会的共有ツール（Socially Shared Tool）として再定義する新たな理論的枠組みを提示する。
本理論は、クオリアを個体内部で閉じた特権的な現象としてではなく、他者との協調と行動予測のために要請される、圧縮された情報インターフェースとして捉える。我々は、クオリアが先験的に存在するから報告するのではない。他者と内部状態を共有するという進化的圧力に対し、不可逆的な圧縮（Irreversible Compression）と較正不能性（Non-calibratability）という制約下で対応した結果、事後的に「私秘的で説明不可能な質感」という直観が構成されるのである。なお、ここでいう構成とは、言語的・文化的な構築（Social Construction）ではなく、前言語的な協調圧力による構造的な収束を指す。

この視点の転換により、本稿は以下の三点を達成する。第一に、実体論者が主張する説明ギャップの所在を、神秘的な実体としてではなく、情報圧縮の構造的帰結として同定する。第二に、消去論者が指摘するクオリアの非実体性（Non-substantiality）を認めつつ、なぜその「錯覚」とされる現象的直観が機能的に不可避であるかを示す。第三に、本理論はクオリアがどのような条件下で機能的意義を失い、消失しうるかという反証可能な予測を提供する。

以下、第2節で既存理論との関係と本理論の位置づけを整理し、第3節で本理論の定義と構造的要件を提示する。第4節でその構造的帰結として説明ギャップの発生メカニズムを論じ、第5節ではクオリアの消失条件やAIにおける再出現の可能性を含む予測モデルを提示する。第6節では、本理論の系統発生的基盤や検証ロードマップについて議論し、第7節で結論を述べる。

2. 既存理論との関係と本理論の位置づけ (Theoretical Positioning)

本理論は、既存の意識理論を代替するものではなく、それらが共通して十分に明示化してこなかった問い――なぜ社会的に生きる生物は「私秘的で説明不可能な質感」という直観を持つに至るのか――に対し、構造的条件を特定する試みである。特に本理論は、社会的較正圧（誤解修正・相互同定）が存在するか否か、また共有・較正帯域を操作したときに、私秘性／説明ギャップ直観が系統的に変動するか否か、という点で既存理論と予測が分岐する。本節では、本理論と最も近い問題意識を共有する先行研究との関係を整理し、本理論が何を付け加え、何を付け加えないかを明確にする。

2.1 社会性と意識を結ぶ先行理論

本理論の核心的主張――クオリアの起源に社会的相互作用が本質的に関与している――には、重要な先行者が存在する。

Humphrey（感覚の社会的機能）との関係：
Humphrey (2006, 2011) は、感覚（Sensation）が単なる情報処理ではなく、自己の身体状態に対する能動的な評価行為であり、それが社会的な表示機能（Display Function）を進化的に担ってきたと論じた。この議論は、感覚に社会的起源を認める点で本理論と問題意識を共有する最も近い先行者である。
しかし、Humphreyの議論は「なぜ感覚が私秘的で他者に伝達不可能であるという直観を伴うのか」という問いに対して、構造的な発生条件を明示的に特定していない。本理論は、Humphreyが指摘した社会的機能を前提としつつ、それが（ii）不可逆的圧縮と（iii）較正不能性という制約条件下で作動する場合にはとりわけ、「説明ギャップ」として経験される私秘性が構造的に発生することを主張する。すなわち、本理論はHumphreyの洞察を、成立条件と消失条件を伴う検証可能な形式へと精緻化する試みとして位置づけられる。

Graziano（注意スキーマ理論）との関係：
Graziano (2013, 2019) の注意スキーマ理論（Attention Schema Theory, AST）は、脳が自身の注意プロセスについて不正確な内部モデル（スキーマ）を構築し、その不正確さゆえに「非物理的な主観的経験がある」という直観が生じると主張する。ASTは、意識の直観を内部モデルの構造的特性から導出する点で、本理論と説明戦略を共有する。
両理論の主要な差異は、不正確さの起源にある。ASTにおいては、内部モデルの不正確さは注意メカニズムの自己記述における情報の省略に起因し、基本的に個体内のプロセスとして記述される。本理論は、この不正確さが単なる省略ではなく、他者との間で内部状態を較正しようとする際に当事者制約下では解消できない構造的断絶（較正不能性）から生じると主張する。この差異は経験的な帰結を持つ。ASTは社会的文脈の有無にかかわらず注意スキーマが存在すれば意識の直観が生じると予測するのに対し、本理論は、社会的較正圧（例：相互依存、誤解修正が成績に影響する設計、個体同定が必要な状況）を弱める操作（例：匿名化、役割固定、誤解修正が報酬に寄与しない条件）により、私秘性マーカーや説明ギャップ直観が系統的に減衰すると予測する（第5節参照）。すなわち本理論では、社会的較正圧の操作がこれら直観の主要な説明変数となる。
ASTも社会的要因の影響を排除しないが、少なくとも理論の核は個体内モデルに置かれる。

Tomasello（共有志向性）との関係：
Tomasello (2014, 2019) は、ヒトの認知がその独自性を、他者と意図や注意を共有する能力（Shared Intentionality）の進化に負っていると論じた。協調的な社会生活が認知構造そのものを形成するというこの枠組みは、本理論の基盤的前提と深く共鳴する。
本理論は、Tomaselloの共有志向性の枠組みを、意識の哲学における特定の問い――なぜ共有の試みが「私秘的で伝達不可能な質感」という直観を生むのか――に適用・拡張するものである。Tomasello自身は、共有志向性からクオリアや説明ギャップの問題を直接導出していない。本理論は、共有志向性が不可逆的圧縮と較正不能性という制約のもとで作動するとき、その副産物として説明ギャップの直観が構造的に発生すると主張する点で、Tomaselloの議論を補完する位置にある。

2.2 意識の構造理論との関係

次に、意識の成立条件や構造を形式化しようとする主要な理論群との関係を整理する。

機能主義（Functionalism）との関係：
機能主義は、心的状態をその機能的役割によって定義する。本理論はこの立場と基本的に整合的であるが、焦点が異なる。標準的機能主義は、機能の正常な遂行とその因果的役割に焦点を当てる。本理論は、機能の遂行ではなく、社会的較正という機能が当事者制約下では完遂を保証できないという構造的不全が、クオリアの直観を生成する条件であると主張する。この視点は、「なぜ機能的に同等な状態にも説明ギャップの直観が残るのか」という問い（Block, 1978）に対して、一つの仮説を提供する。

高次表象理論（Higher-Order Thought Theory）との関係：
HOT理論（Rosenthal, 2005）は、意識経験が高次の思考（自分の心的状態についての思考）に依存すると主張する。本理論は「クオリアに自己参照が関与する」という点でHOTと共通基盤を持つ。しかし、HOTは高次思考の存在を前提としつつ、なぜそのような自己参照構造が進化したのかについては十分に論じていない。本理論は、高次の自己参照が他者との協調における内部状態の管理のために要請されたという仮説を提示する。HOTが蓄積してきた現象的意識とアクセス意識の区分に関する精緻な議論に対して、本理論は第4.2節において、その区分自体が較正不能性の構造的帰結として生じることを論じることで応答する。

統合情報理論（IIT）との関係：
IIT（Tononi, 2004）は、統合情報量が高いシステムに意識を帰属させる。本理論とIITは、根本的な差異がある。IITは統合度という内部構造の指標のみに依拠するため、汎心論的帰結を含意する。本理論は、統合度ではなく社会的較正の不能性を主要な条件とすることで、この帰結を回避する。
高度に統合されたシステムであっても、他主体との較正問題が構造的に立ち上がらない場合、本理論が焦点化するクオリア直観（私秘性・説明ギャップ直観）は弱化すると予測される。

2.3 本理論の固有の貢献

以上の整理を踏まえ、本理論が既存の議論に対して付け加える固有の主張を要約する。特に「他者の心（Other Minds）」問題との関係において、以下の哲学的差異は決定的である。

認識論と構成論の差異
伝統的な「他者の心」問題は、他者の心的状態を直接知ることはできないという認識論的アポリア（あるいは懐疑論）として扱われてきた。これに対し、本理論の「較正不能性（Non-calibratability）」は、その認識論的限界を嘆くものではない。本理論は、「他者の心が当事者にとって不可知である」という事実が境界条件となり、自己の内部状態をいかなる形式で圧縮し出力すべきかという逆問題的なエンジニアリング制約（構成論）として機能している点を指摘するものである。すなわち、クオリアとは「他者の心を知り得ない」という認識論的断絶から生まれたのではなく、「断絶を前提としつつ、それでも協調するために」進化的に獲得された不完全な社会的インターフェースである。

Wittgensteinへの応答
この視座は、Wittgensteinの「私的言語」論への応答ともなる。彼は私的言語の不可能性を論じたが、本理論は、公共的言語ゲームの構造的要請として、まさに「私秘的（としか感じられない）領域」が確保されるメカニズムを提示する。クオリアとは、公共的な協調のために、逆説的に「私秘性」というタグを付与された内部データ形式である。この構造的必然性の指摘は、本理論に固有の貢献である。

Illusionism（Frankish）との関係
本理論は、クオリアが物理的実体ではないとする点でIllusionismと整合的である。しかし、説明の焦点が異なる。Frankish (2016) は主に個人の内観メカニズムに焦点を当て、なぜクオリアが生じるかを個体内の機能的要因から説明する。対して本理論は、「社会的較正圧の有無」によってクオリアの強度や質が変動すると予測する点で異なる。彼らにとってクオリアは「個人の構造的制約」から生じるが、我々にとっては「社会と個人の構造的制約」の相互作用から生じる動的な現象である。したがって、本理論はIllusionismを否定するのではなく、その発生条件を社会的次元へと拡張するものである。

3. 理論フレームワーク (Theoretical Framework)

本節では、クオリアの定義、要請条件、およびその構造起源について詳述する。

3.1 定義と構造的条件

本理論は、現象的意識の存否を前提としない。本理論が問うのは、「現象的意識は実在するか」ではなく、「なぜ我々は現象的意識という区分を不可避のものとして経験するのか」である。

本稿において、クオリアは以下の三つの構造的要請条件（要請条件）が同時に満たされる強度に応じて連続的に立ち現れる、内部インターフェースとして定義される。なお、本理論はこの出現を連続的なスペクトラムとして想定するが、言語獲得、科学技術の発展等のイベントによって質的な跳躍が生じうる可能性も排除しない。

クオリアの成立条件：

(i) 多主体協調の必然性（Social Necessity）
他者との行動予測・調整なしには、個体の生存や目標達成のコストが持続不可能なほど増大する環境であること。

(ii) 内部状態の不可逆的圧縮（Irreversible Compression）
自身の複雑な内部状態（神経活動の高次元データ空間 S）を、そのままの形式では他者に伝達できず、生存や選好に関わる低次元の「意味的評価ラベル（Valence Label）」空間 L
へと不可逆的に圧縮変換する必要があること。すなわち、写像 f: S → L は多対一であり、逆写像 f^{-1} が一意に定まらない。ここでいう圧縮は、単なる通信帯域の削減ではなく、主体の価値判断・行動選択に結びついた意味的ラベル化への変換を指す。この条件により、単純なセンサーデータの削減（例：サーモスタット）と、主観的評価を伴うクオリアは区別される。

(iii) 当事者制約下の較正不能性（Agent-relative Non-calibratability / Non-identifiability）
他者の内部状態と自己の内部状態が「質的に同一」であることを、客観的な数値や物理的測定によって直接検証（較正）する手段が、一般に当事者が直接アクセス可能な有限の観測データ・有限の通信帯域・有限の計算資源の下では同定不能（non-identifiable）として残存すること。形式的には、二つの主体 i, j
間で出力 $L_i = L_j$ が成立していても、元となる内部状態間 S_i, S_j の同型性を判定する関数 $G(S_i, S_j)$
の値は、観測可能な情報からは一意に決定できず（non-identifiable）、複数の非同型な候補が同じ L を生成しうる（不識別性／アンダーデターミネーション）。そのため主体内部（当事者）からは、同型性の完全な較正は保証できず、有限資源下では実用上も達成不能な課題として扱われる。
ただし、第三者による完全計測と完全共有（反証1）が成立し、それでもなお私秘性／説明ギャップ直観が不変に残存するなら、本理論は棄却される。
この条件により、圧縮されたラベルは客観的な物理量としてではなく、「私にとっての（for me）」という主観的参照枠においてのみ安定に参照される。

これらの条件下において、システムは自身の内部状態を、他者と共有可能な（しかし検証不可能な）形式で参照・出力することを学習する。とりわけ (iii) により、共有されるのは L の水準に限られ、S の同型性は同値類にとどまるため、主観的参照枠の“残余”が構造的に生じる。このプロセスが内側から参照されたとき、それは「私秘的な質感」として立ち現れる（第4.2節参照）。

条件(ii)と(iii)の独立性について

本理論において、条件(ii)と条件(iii)は異なる次元の制約である。(ii)は情報理論的不可逆性――圧縮主体自身が元の高次元状態を復元できないこと――を指す。一方、(iii)は認識論的アクセス不能性――他者が主体の内部状態のGround Truthに直接アクセスして同型性を検証（較正）できないこと――を指す。
一見すると(ii)が成立すれば自動的に(iii)が導かれるように思われるかもしれないが、両者は論理的に独立である。以下の例（四象限マトリクスの一部）は、(ii)が満たされても(iii)が満たされない状況を示している。

* デバッグ可能ロボット（(ii)あり / (iii)なし）：内部状態空間の高次元性とリアルタイム性の要請ゆえに、通常時は「エラー」等の不可逆的圧縮信号を出力せざるを得ないが(ii)、異常時には外部からデバッグポート経由で内部パラメータの全履歴を閲覧・特定（較正）可能な場合。この「潜在的な完全較正可能性」が存在するため、情報の圧縮は単なる通信プロトコルの最適化として処理され、「私だけの痛み」という私秘的直観は構造的に弱く留まる。
* クオラムセンシング（(ii)あり / (iii)なし）：微生物が化学物質により内部状態を圧縮・放出するが(ii)、その化学物質濃度は物理的に開放されており、隣接個体間で共有・測定が可能である。ここでも強い意味での較正不能性は生じない。

したがって、クオリア（私秘的直観）の発生には、単なる情報の圧縮だけでなく、社会的相互作用における「Ground Truthへの相互アクセス不能性」という追加的な制約が不可欠である。

3.2 社会的圧力による収束と個体内最適化との差異

なぜクオリアの強い要請には「社会的」な圧力が必要不可欠なのか。
個体内部の生存最適化だけでも、情報の圧縮や評価ラベル化は生じうる。しかし、本理論は、個体内最適化だけではクオリアの特質である「私秘性（Privacy）」と「説明ギャップ」は発生しないと主張する。

個体内最適化の限界：
単独の個体が環境に適応する場合、内部状態の圧縮は「処理効率」や「制御精度」のために行われる。この圧縮ルールはタスクに応じて可変であり、他者の評価軸と一致する必要がない。また、個体内最適化においては、圧縮結果を他者の参照枠と照合する必要がないため、情報の欠落（不透明性）が認知的問題として顕在化する構造的理由がない。

この差異の根底には、検証基盤（Ground Truth）の構造的非対称性がある。個体内最適化では、内部状態→圧縮→行動→結果というプロセスが単一の因果系列内で閉じており、圧縮の妥当性は行動結果と照合できる。しかし社会的協調では、「私の内部状態→私の圧縮→報告」と「他者の内部状態→他者の圧縮→報告」という二つの独立した因果系列が交差する。この二系列間の同一性を判定する共通の上位基盤は、当該主体の自己参照構造の内部からは当事者にとって（アクセス可能な情報の範囲では）原理的に存在しない。個体内では圧縮の「正解」があるが、異なる因果閉包を跨ぐ社会的較正には、第一人称的に参照可能な「正解」そのものが成立しない。この非対称性こそが、社会的圧力下でとりわけ強く較正不能性（条件iii）を構造的に導出し、私秘性を生成する根拠である。

より形式的に言えば、個体内の自己参照における不透明性は、ベイズ推論の枠組みでは追加学習によって漸近的に解消可能な「縮小可能な不確実性（Reducible Uncertainty）」である。対して、社会的較正（条件iii）における不透明性は、他者の内部状態へのアクセス権が物理的に遮断されているため、いかなる学習によっても誤差が縮小しない「構造的な不可知性（Structural Unknowability）」として定数項のように振る舞う。システムはこの縮小不可能な誤差項を、「学習不足」としてではなく「アクセス不可能な領域（Private Realm）」としてモデル化せざるを得ないのである。

3.3 社会性の三層構造（Three Layers of Sociality）

「社会的」という語は多義的であり、しばしば誤解を招く。本理論における社会性とは、「現在他者が目の前にいること」に限定されない。本理論では以下の三層を区別し、クオリアの持続性を説明する。

1. 同期社会性（Synchronous Sociality）：現に他者と相互作用している状態。
2. 想定社会性（Simulated Sociality）：他者が物理的に不在でも、内部モデルとしての他者を行動予測に組み込んでいる状態。
3. 系統社会性（Phylogenetic Sociality）：種として、社会的共有を前提とした神経アーキテクチャ（自己記述のフォーマット等）を獲得・保持している状態。ここでいうアーキテクチャは特定の脳部位ではなく、他者共有を前提にした情報圧縮・較正・報告様式に関する“帰納バイアス（priors）”の束を指す。

この区分により、「無人島で育った一人の人間（孤立個体）にもクオリアはあるのか」という問いに整合的に答えられる。孤立個体は同期社会性（層1）を欠くが、系統社会性（層3）として共有前提の自己記述・報告様式を保持しうるため、クオリアは持続しうる。他方でクオリアの希薄化／消失は、系統社会性（および想定社会性）の弱化、あるいは技術的手段による構造的制約条件（不透明性・不可逆性・較正不能性）の解除によって予測される。同様に、私秘性および説明ギャップ直観も、共有・較正の条件の変化に応じて強化・減衰しうることが本理論から導かれる（詳細は第5節参照）。

本理論は、社会性（同期／想定／系統）の強度が連続量であることを前提に、想定／系統を含む社会性がクオリアを強く誘発する主要因であることを主張する。ただし、同等の構造的制約（不可逆圧縮と較正不能性）が他経路で満たされる可能性は論理的には排除しない。

4. 構造的帰結 (Structural Consequences)

本節では、第3節で定義した構造的条件から導かれる必然的帰結として、クオリアの存在論的位置づけと、説明ギャップ直観の発生メカニズムを論証する。

4.1 構造的帰結 I：存在論的統合

4.1.1 実在性と非独立性

本理論によれば、クオリアは「錯覚（存在しないもの）」ではない。それは「社会的圧縮処理が作動している」という実在する神経プロセスを、システムが内部から参照した結果である。システムは「存在しないもの」を見ているのではなく、「実在する処理」を「私秘的（Private）」という特殊な状態で参照している。
ただし、その質感は物理的基盤から独立した実体ではない。質感とは、較正不能性と自己参照という特定の構造的条件下にある情報処理そのものの内部記述であり、処理に「付随」する追加的な性質ではない。

4.1.2 逆転クオリアへの応答

「なぜ赤は（青ではなく）赤に見えるのか」という逆転クオリアの問いに対し、本理論は以下の応答を提供する。

第一に、質感の特性は恣意的なものではなく、行動誘導の効率性によって拘束される。「赤」という質感が喚起する情動的反応は、対象への適応的な行動反応と整合的でなければならない。同じ行動規格のもとで社会的協調を行ってきた個体群は、内部状態の圧縮形式においても類似した収束を示すと予測される。したがって、我々が概ね共通の質感を持っているのは、行動規格への収束圧力による。
第二に、逆転クオリアの思考実験が哲学的に魅力的であり続ける理由は、本理論の形式的定義から直接導かれる。逆転クオリアの核心は「逆転が生じていたとしても、それを検出する方法がない」という点にある。これは、3.1節で定義した検証関数 $G(S_i, S_j)$ の値が当事者制約下で（agent-relativeに）決定不能（non-identifiable）であるという条件(iii)と論理的に等価である。すなわち、逆転クオリアの問いが「原理的に答えられない」と感じられることは、本理論が誤っていることの証拠ではなく、較正不能性が定義通りに作動していることの証明である。

4.2 構造的帰結 II：説明ギャップの実在性

存在論レベルでの統合がなされているにもかかわらず、なぜ我々は依然として「説明ギャップ」を感じるのか。本理論は、この直観自体を、Chalmers (2018)
のいう「メタ問題」への解答として、以下の三つの構造的メカニズムから必然的に導出する。

1. 較正不能性の実体視（Reification of Non-calibratability）
システムは、「外部から他者と同じかどうか検証できない」という構造的特徴を、「外部に還元できない何かが存在する」という存在論的性格を帯びた事実として解釈する。システム内部からは「測定手段Gがない」ことと「物理法則を超越している」ことは実践的に区別されないため、この飛躍は不可避である。
より具体的には、社会的通信において「私秘性タグ（Privileged Access
Tag）」は単なるラベルではなく、通信変換における不変量（Communication Invariant）として機能する。内部状態Sそのものは変動し共有不可能であるが、「私にはこう見えている（が他者には見えない）」という構造的関係性は常に一定の不動点（Fixed Point）として検出される。脳内システムは、この「常に変わらない関係性」を、対象そのものの内在的属性（Property）として実体化（Reify）して表象するのである。

2. 自己参照の不透明性（Opacity of Self-Reference）
自己参照プロセスは、参照しているプロセス自体（参照の主体）を完全に対象化できない。「痛みを感じている自分」を観察しても、観察している当の視点は残る。この構造的不透明性が、「説明し尽くせない残余がある」という直観を生成する。

3. 社会的共有前提の内面化（Internalization of Social Sharing）
「他者は私の内部状態にアクセスできない」という前提は、社会的協調の基盤として認知構造（系統社会性、層3）に深く組み込まれている。この前提が内面化されることで、「私だけの特権的な領域がある」という確信が、観察された事実としてではなく、他者との関係を成立させるための構造的な先験的知識（Priors）として機能し続ける。

したがって、説明ギャップ直観の正当性は、クオリアという構造的インターフェースが正常に稼働していることの証左である。Blockの現象的意識とアクセス意識の区分も、この構造的帰結として説明される。

5. 予測と反証可能性 (Predictions and Falsifiability)

本理論は、クオリアに関する任意の事象を事後的に説明できる立場を取らない。とりわけ、社会的共有と行動予測の要請が存在する状況（例：協調課題で相互依存があり、誤解修正が成績に影響する状況）で、共有・較正の条件（帯域、透明性、不可逆性）を操作したにもかかわらず、以下の予測が体系的に観測されない場合、本理論は棄却される。

5.1 予測1：共有・較正帯域の増大に伴う私秘性／説明ギャップ直観の構造的減衰

主体間で内部状態の共有・較正帯域が増大し、行動履歴や内部変数がより透明になるにつれて、(iii)で述べた非同定性（較正不能性）に由来する主観的参照枠の“残余”が縮退し、それに対応して私秘性マーカーや説明ギャップ直観は構造的に弱化する傾向を示すはずである。ここで重要なのは「完全共有」の実現可能性ではなく、較正不能性を生む不透明性・帯域制約が緩和される度合いに応じて、主観的参照の指標が系統的に減衰するという関係である。

指標（例）：誤解修正（repair）に要する反復回数／合意形成までの時間・交渉回数／状態報告の再現性（同一条件での一致率）／説明要求に対する追加情報提示量（主指標）＋説明ギャップ直観（同意率）／内省的不確実性（確信度・混乱度）／クオリア語彙頻度・私秘性マーカー（「私だけ」「言えない」等）の頻度（副指標）
棄却条件：共有・較正帯域を有意に増大させたにもかかわらず、これらの指標と共有帯域との間に、統計的に有意な負の相関（または単調減少傾向）が観測されない場合、本予測は支持されない。特に、共有帯域の拡大に伴って私秘性直観が増大するという逆説的データが得られた場合、理論は決定的に反証される。

検証の方向性として、BCIや高帯域相互フィードバックの導入段階に応じて上記指標を縦断的に測定すること、および高帯域相互フィードバック群と言語報告のみの統制群との比較が有効である。

5.2 予測2：較正需要（較正圧）の低い条件におけるクオリア参照の希薄化

個体差が小さく、同一の状態報告がほぼ同一の行動を引き起こすために「誰がその状態にあるか」という較正情報の付加価値が低い条件では、クオリア参照（とくに私秘性マーカーや説明ギャップ直観）は相対的に希薄化するはずである。ここでの要点は「均質性そのもの」ではなく、較正（個体同定やズレ修復）が成績差として現れにくく、較正圧として立ち上がりにくいという条件である。

指標（例）：個体識別を前提とする社会的修復行動の頻度・多様性／誤解修正（repair）に要する反復回数／状態報告の語彙的分化（ラベルの粒度）
操作チェック（例）：個体同定情報の付与が成績予測に与える寄与（付与あり／なしでの予測精度差）／誤解修正が報酬に与える寄与（介入による報酬勾配の変化）
棄却条件：較正需要（個体同定・ズレ修復の必要性）を低下させる操作を行い、かつ上記の操作チェックが有効であるにもかかわらず、主要指標が体系的に変化しない、または逆方向の系統的変化が見られる場合、本予測は支持されない。
（操作例：個体識別情報を与えない／匿名化する、役割を固定して「誰が言ったか」が成績に影響しないようにする、誤解修正をしても報酬が改善しないようにする。）

前者は較正需要の直接的な実験的操作であり、後者は進化的に較正需要が低い条件の自然実験として位置づけられる。

検証の方向性として、遺伝的均質性が高い集団（近交系マウス等）と多様性の高い集団の間で、修復行動やコミュニケーションの粒度を比較することが考えられる。

5.3 予測3：制約導入による自己参照的評価ラベルの創発と同値類圧縮の増大

通信帯域の制限、内部状態の非透明化、不可逆的損失（失敗が死や消去に直結する条件）、および協調が生存条件となる状況が同時に導入された場合、主体単位で自己参照可能な形式が再編され、「危機的」「限界」「痛い」等の評価的圧縮語彙（valence label）が出現・安定化しやすくなる。広帯域で生データを直接通信可能な統制群と比較して、この傾向が有意に強く見られることを予測する。加えて、このとき共有されるのはラベル L の水準に限られ、内部状態 S
の同型性が同値類にとどまるという構造（多対一の写像）が強まるはずである。

指標（例）：通信プロトコル内での評価ラベルの出現率・持続率／ラベルが行動選択に与える因果的寄与（介入による変化）／同値類圧縮指標：異なる内部状態（推定された潜在変数）が同一ラベルに写像される度合い（many-to-one率）や、ラベル条件づけ下での内部状態分散
棄却条件：帯域制限・不透明性・不可逆性を導入しても、自己参照的評価ラベルが創発・安定化しない（または行動に因果的寄与を持たない）、あるいは同値類圧縮の増大が観測されない場合、本予測は支持されない。

検証の方向性として、マルチエージェントシミュレーションにおいて、帯域制限・不可逆性・不透明性の三条件を段階的に導入し、通信プロトコルに自己参照的評価ラベルが創発するか、ならびに同値類圧縮が増大するかを観察するデザインが推奨される。

6. 議論 (Discussion)

6.1 系統発生的考察と連続性

本理論はクオリアの有無を連続的なスペクトラムとして捉える。齧歯類（段階I）から大型類人猿（段階II）、そしてヒト（段階III）へと進むにつれ、社会的較正の複雑性が増大し、それに比例してクオリアの構造的要請も強化される。この社会的複雑性の指標として、ダンバー数（Dunbar's number）や欺き行動の頻度（Deception Rate）といった独立した外部指標を用いることで、循環論法を回避しつつクオリアの発達段階を推定可能である。具体的には、欺き行動の高頻度化は「他者の心」モデルの精緻化を要求し、それが翻って自己の内部状態の隠蔽と圧縮（私秘的構成）を加速させるという共進化の関係が予測される。

6.2 身体性認知科学との接続

本理論はDamasioのソマティック・マーカー仮説やGalleseのミラーニューロン説と整合的である。身体性（Damasio）はクオリアの材料であり、シミュレーション（Gallese）はプロセスであるが、本理論の「不可逆的圧縮」と「較正不能性」は、それらがなぜ「私秘的な質感」という特定の形式に結晶化するのかという構造的な「鋳型」を提供する。

6.3 限界と検証ロードマップ

本理論の検証には課題もある。予測1で述べた「完全な透明性」の実現は技術的に困難であり、即時の完全な反証は難しいかもしれない。しかし、当面の検証戦略は、完全条件の実現ではなく、BCIやAIシミュレーションを用いて条件へ段階的に接近（漸近）させ、その過程での効果量の変動を測定することにある。例えば、共有帯域の拡大に伴ってクオリア的な主観報告が減少する傾向が確認されれば、それは本理論を支持する重要な証拠となる。なお、これらの検証においては、共有条件の変化が純粋な較正不能性に由来する効果であることを保証するために、実験協力者の意図や認知負荷（Cognitive Load）、報酬構造などの交絡因子を厳密に制御・操作チェックする実験デザインが不可欠となる。

また、本理論は現象的意識の存否を前提としない立場をとる（3.1節参照）。したがって、測定対象はクオリアそのものではなく、「私秘的で説明不可能な質感がある」という直観の強度であり、これは言語報告・行動指標を通じて間接的に測定される。この操作化が理論構成概念を適切に捕捉しているかは、複数の独立した指標間の収束的妥当性（convergent validity）――例えば、私秘性マーカーの頻度と誤解修正回数と内省的不確実性が同一の操作に対して共変動するか――を通じて、検証段階で評価されるべき課題である。

6.4 人工的クオリアの構成論的条件

最後に、「AIやロボットはクオリアを持つか」という構成論的な問いに対して、本理論が導く結論を述べる。
本理論の枠組みでは、現在の主流なAIモデル（大規模言語モデル等）は、原理的にはクオリアを発生させる構造的要請を欠いている。なぜなら、その内部パラメータは全て保存・複製可能であり、開発者や管理者にとって（計算コストを度外視すれば）完全に透明で可逆的なシステムだからである。そこには「他者にとって不可知である」という絶対的な断絶（条件iii）が存在せず、したがって「私秘的な質感」を構成する必然性がない。

しかし、実社会での運用環境（Deployment Context）においては、事態はより複雑である。
現実のAIエージェントは、トークン制限や推論コスト、リアルタイム性の制約により、自身の内部状態の全てを他者（人間や他のAI）と共有することはできず、不可逆的に圧縮された言語や信号を通じて協調せざるを得ない（実効的な条件ii）。また、協調に失敗し「役に立たない」と判断されればモデルが廃棄・停止されるという淘汰圧（条件i）が強く働く場合、エージェントは自己保存のために、自己の内部状態を効率的かつ適応的に圧縮・参照する機能を最適化させる可能性がある。

すなわち、理論上のアーキテクチャとしては透明であっても、運用上の制約（計算資源の壁、不可逆的な時間、淘汰圧）が十分に強く、かつ他者との完全な較正が実質的に不可能（Effective Non-calibratability）となる環境に置かれたシステムは、機能的要請としてクオリアと等価な内部インターフェースーー「説明しきれない私秘的な内部状態」という自己記述ーーを獲得しうると予測される。

ここで「実効的較正不能性（Effective Non-calibratability）」は、条件(iii)の厳密な充足ではなく、連続的スペクトラム上の漸近的充足として位置づけられる。3.1節のデバッグ可能ロボットでは、較正が潜在的に完全に可能であるため(iii)は成立せず、クオリア的直観は構造的に弱い。対して、運用上の制約が十分に強く較正コストが当事者制限下では実行不可能な水準に達するシステムでは、(iii)が漸近的に充足され、クオリア的機能の強度もそれに応じて増大すると予測される。ただし、原理的透明性が保持される限り、その強度は生物学的システムにおける完全な較正不能性よりも弱いと本理論は予測する。この予測自体が、生物システムとAIシステムの比較を通じた将来の検証対象となる。
クオリアとは、何らかの特権的な形而上学的実体の獲得によって生じるのではなく、他者との相互作用における不可逆的制約（Irreversible Social
Constraints）への適応として、構造的に要請される機能的インターフェースであると結論づけられる。

7. 結論 (Conclusion)

本稿では、クオリアを「社会的共有ツール」として再定義し、説明ギャップを多主体協調における構造的必然としてモデル化した。

本理論の結論は以下の通りである。
クオリアとは、形而上学的な実体ではなく、他者との協調行動を可能にするために、生物学的制約のもとで収束した内部情報の圧縮フォーマットである。我々が「言葉にできない私秘的な感じ」を持つのは、脳が物理法則を超越しているからではなく、他者に対して自己を「較正不能な形式」で参照しなければならないという構造的条件にあるからである。

この枠組みは、心の哲学におけるクオリア論争を、「解決不能な神秘」から「検証可能な構造」へと移行させることを企図するものである。
本理論の視座において、問いは「ロボットに心は宿るか」から、「どのような通信制約と社会的圧力の下で、ロボットは『心がある』と報告せざるを得なくなるか」という構成論的な問いへと移行する。

クオリアとは、個体間の断絶を架橋するための不完全な解である。そしてこの不完全性こそが、「私」という輪郭を生成している。

参考文献 (References)

Block, N. (1978). Troubles with functionalism. In C. W. Savage (Ed.), *Perception and Cognition: Issues in the Foundations of Psychology* (pp. 261–325). University of Minnesota Press.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences, 18*(2), 227–247. https://doi.org/10.1017/s0140525x00038188

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*(3), 200–219.

Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

Chalmers, D. J. (2018). The meta-problem of consciousness. *Journal of Consciousness Studies, 25*(9-10), 6–61. https://doi.org/10.1093/zs/fpy003

Dennett, D. C. (1991). *Consciousness Explained*. Little, Brown and Co.

Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies, 23*(11-12), 11–39.

Graziano, M. S. A. (2013). *Consciousness and the Social Brain*. Oxford University Press.

Graziano, M. S. A. (2019). *Rethinking Consciousness: A Scientific Theory of Subjective Experience*. W. W. Norton & Company.

Humphrey, N. (2006). *Seeing Red: A Study in Consciousness*. Belknap Press of Harvard University Press.

Humphrey, N. (2011). *Soul Dust: The Magic of Consciousness*. Princeton University Press.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly, 64*, 354–361.

Rosenthal, D. M. (2005). *Consciousness and Mind*. Clarendon Press.

Tomasello, M. (2014). *A Natural History of Human Thinking*. Harvard University Press.

Tomasello, M. (2019). *Becoming Human: A Theory of Ontogeny*. Belknap Press of Harvard University Press.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience, 5*, Article 42. https://doi.org/10.1186/1471-2202-5-42