ひねくれた心の多様体理論
Manifold Theory of Twisting Mind

犬童 健良[†]
Kenryo Indo
[†]関東学園大学
Kanto Gakuen University
kindo@kanto-gakuen.ac.jp

概要
This paper proposes a manifold model of the mind and interprets cognitive processes as transformations between local coordinate systems that depend on the amount of field. Cognitive space is assumed to be affected by cognitive blocking. Cognitive blocking has duality of closedness and creativity and recursion of inhibiting self-detection. The twisted mind is an image in the embedding of the bending of the cognitive space by the force of cognitive blocking. Minsky's frame system, Nash's C^1-embedding theorem, and Shapley's labelling system for bimatrix games were considered as methods to model the twisted mind.

キーワード : 多様体，測地線，埋め込み，ひねくれ，認知的阻止

1. はじめに

本論文でこれから述べようとしていることは，読者には理解できないかもしれない．なぜならばそれは本論文が考察する心の基本性質だからである．またそれは認知的に阻止される（cognitively blocked）．一般相対性理論[2]が重力を空間の曲がりとしてモデル化するのと似て，本論文ではひねくれた思考を認知と感情の相互作用によって認知空間の曲がりそのものとして表現する．曲面に局所的に貼り付けられて暮らすエージェントたちが直接に知るのはその第一基本量であり，空間自体を曲げる外力である認知的阻止には気づかない[8]．認知的な阻止のはたらきは双対的かつ再帰的である．認知的阻止は心のひねくれた状態を作り出す一方，論理的思考や創造的思考の源泉でもある．認知的阻止はそれ自体を認知的に阻止する．

学術広辞苑第三版によると，ひねくれる（捻くれる）とは「性質がすなおでなくなる．ねじける」ことである．またねじける（拗ける）とは「まがりくねる．すなおでない．ふつうとちがってまともでない．ひねくれること」である．本人は率直に表現していても，何が真っすぐであるかという基準が異なる他者から見ればわかりにくいかもしれない．

例えば文芸作品や日常会話にみられるように，人々は一方では言葉の意味の曖昧さを許容しつつ，他方で，学術研究や技術文書では，できるだけそれを回避しようとする．しかし曖昧さを許容するか，厳密を追求するかの実際は言葉が使用される場に依存する．またその場への依存の仕方には一定の社会的な基準があると考えられる．例えば「満足する」ということの意味を考えてみよう．満足とは，もうそれ以上追加的な活動をしなくてよいということなのか，あるいはもっと活動したいということなのか．実際，H.A.サイモンの満足化原理における満足は前者の解釈を受けるが，授業評価アンケートの満足度調査の満足は後者の意味で用いられる．またいずれの場合も他の解釈との混同は無意味である．

したがって，その場に応じて異なる態度をとるからといって，それだけでひねくれているとは言われない．これは言葉の使い分けが適切になされるのがむしろ普通のこととして期待されていることを意味する．一方，もしこれを自己矛盾と考えて，曖昧さを排除するため，日常会話においてもつねに四角四面の表現を使ったり，あるいは逆につねに自由な創造性を求めるつもりで曖昧な表現を使ったりする人がいれば，ひねくれていると思われかねない．

ひねくれは，ある人の考えや表現が別の人から見て不自然でわかりにくいと思える現象である．数学的な表現を借りると，その人は最適な（あるいは合理的な）選択をしておらず，またそれは意図的にそうしているかのように思える．つまり意図が伝わらないことを意図して表現していることになり，理解できない．ひねくれた思考は曲面上をぐねぐねと進み，解釈者の座標系から見て，まっすぐに進んでいるように見えないが，しかしその人の思考を表す曲面上ではまったく合理的に動いているということがありうる．これが思考や心の過程の基本的性質として本論文の提起する仮説である．より厳密に言えば，ひねくれは，他者という多様体を自分たちの認知空間内に距離，つまり計量テンソルを保つように埋め

込むこと, すなわち等長埋め込みが不可能であるか, あるいは少なくともまっすぐではなく，ジグザグになってしまうことに対応する．等長に滑らかに埋め込むには十分に大きな次元を要する．滑らかでなくてよいのなら, 最低 1 次元だけ大きくすればよいが, 滑らかではないので局所座標系がパラメータ変化で移動するとき奇妙な現象を引き起こしうる（ナッシュの C^1級埋め込み定理[7]）.

本論文では認知過程を局所座標系の間の変換としてとらえる．人間は言葉を使って思考や心の状態を表現することができる．言い換えれば，認知空間は記号的表現の場に依存した局所座標系，それに対応するエージェントの集まりとみなせる．実際，[8]は何微分多様体をそのように解釈している．また後述するようにフレームシステム[6]は多様体としての意味を持つ．しかし局所座標間の変換は，必ずしも滑らかであるとは限らない．測地線は外力をいっさい受けずに空間の曲がりに沿って無駄なく移動するときの軌道である．測地線は認知空間において素直な心の動き，あるいはまっすぐな思考をモデル化するものと解釈される．一方，ひねくれた考えは，たんなる誤答や非意図的に生じる心理バイアスではなく，意図的にそれを選んでいる．すなわち，ある人の思考（座標系 A）が別の人（座標系 B）から見て，曲がった，しかも滑らかでない軌道を描く．ここには本人は率直に考えているつもりだが，別の座標系から見て曲がった軌道を描くという先述の場合も含まれる.

多様体は場の量に依存し，相互に変換するオペレーションを伴う（局所）座標系の集まりであり，曲がった空間を表現する．またより高次元の空間への埋め込みにおける奇妙な性質が知られる．それは本論文で考察するひねくれた心のモデルに適している. そこで本論文は多様体を応用した心のモデルとしての心的多様体を提案する．以降の節では，まず第 2 節ではひねくれた心と認知的阻止の現象について，いくつかの例題を示す．具体的には動機付けされた記憶，ゲームと意思決定のパズル的例題，および生成 AI をとりあげる．第 3 節でひねくれた心を多様体として解釈し，局所座標系の変換，共変と反変，およびミンスキーのフレーム概念との関連に言及する．第 4 節ではラベル図を用いた簡易なシミュレーション例を示す．第 5 節でまとめとする.

2. ひねくれと認知的阻止の例

自己欺瞞やアクラシアは動機付けられた（偽の）記憶であり，ひねくれた心の例である[1][9]．自己欺瞞の意思決定者は，正しい情報を信じるのではなく，意図的に，誤った情報を信じることで満足を得ている．同様に，アクラシアは，正しい情報を信じるのではなく，誤った情報を信じることで意思決定者が満足を得るが，意図的にではなく，非意図的に誤った情報を信じている. 偽の記憶の変種として, 妄想は送信されていないメッセージを受信したと信じ込むことであり，記憶喪失（アムネシア）はその逆である．前者では（正しい）未受信の記憶が阻止され，後者では対象の記憶自体が阻止される. 認知的阻止として見るならば, 偽の記憶は真の情報の認知を阻止しているということに注意するべきである. つまり真の記憶と偽の記憶は, 互いにライバルを阻止するという対称性を持つ．認知的阻止としての両面性の性質はこの後すぐ説明する意思決定のパラドックスにも現れる.

ひねくれた心を分析した別の例は，意思決定論やゲーム論におけるプレイヤーの合理性についての批判的考察に現れる．行為によって期待される満足と情報が相互依存することに原因がある．合理的な選択の状況では, 標準的に, 行為の結果から期待される満足度の大きさが行為の選択を左右すると仮定され，可能な結果についての満足度（効用）と情報（確率）とは独立に扱われる. 以下はギルボアとシュメイドラー [4]の例題を参考に筆者がアレンジを加えた.

例 1. キャッチ２１．Y 君は兵役に召集された際，自分自身の精神疾患を証明する診断書を提出した．精神疾患の証明が軍医によって認められれば Y 君は高い効用を得るが, 兵役逃れ（つまり戦争という狂気からの離脱）は正気を証明すると判断されれば低い効用となる.

例２．甘美な報復（sweet revenge）．X君はZ君に私的な恨みを持っている．しかし実際に報復すれば自身の社会的評価を下げかねない. そこで, Z君が報復を予期したときに，X君の満足はより高くなるだろう.

両例ではいずれも意思決定者の計画の成否が自分自身の選択だけでなく，相手の期待する自分自身の選択に依存している．つまり実際に行われる行為とその結果ではなく, 行為についての情報や信念, より正確には相互信念が，心の状態としての期待される満足と行動選択を左右する．次の例題も意思決定論でよく知られている.

例３．抜き打ち試験（surprise test）．W先生は中学校の数学のクラスを担当していて，来週の月曜から金曜のいずれかで抜き打ち試験を行うと生徒たちに告知した．もし試験なしで木曜の授業を終えると翌日の実施が予期されてしまうので，金曜は除外される．もし水曜までなかったとすると木曜も除外される．同様に水曜，火曜も除外され，月曜の実施が推論される．

生徒は消去法を用いて月曜の試験実施を予期できる．それは合理的選択の基本原理（バックワードインダクション）とも一致する．しかしW先生はこの予測を裏切って月曜日の試験を見送ることで，抜き打ち試験を実施するだろう．ちなみにニューカム問題や有限反復囚人ジレンマもやはり期待に依存した効用の下でのバックワードインダクションが認知的に信頼できなくなる．これらの意思決定のパズルは合理性の基準自体の曖昧性を伴っており，一つの基準を採用することによって，他の基準とその下で導かれる解は認知的に阻止される．

ちなみに，信頼は認知的阻止の双対概念である．すなわち，ある種の寛容性（tolerance）を導入することで認知的阻止が緩和される．しかしそれによって潜在的な自己言及パラドックスが生じる．また認知的な阻止には記憶の誤りや情報選択のバイアスという側面だけでなく，論理や創造性の要素でもあるという両面がある．周知のように，シェーファー棒“｜”で表される阻止関係（NAND 演算）のみで命題論理は再現される．三段論法に対応するのは，((C | D) | A)かつ A ならば C であるという推論規則である．これはニコド(Nicod)のルールと呼ばれる．結論Cはそれを阻止していた力を，Aという情報が阻害することによって導かれる．創造的な思考のためには，既定の認知的なフレーミングの下での解空間探索の認知的阻止を解除する必要がある．９ドットパズルにおいては，点の配置で囲まれた領域の外から問題を眺めることが必要であり，また T パズルでは最も使いにくい（できれば使いたくない）と思われるピースに注目することで正解は自然に導かれる．悪魔の弁護人（devil’s advocate）は意図的に多数派や常識に反する意見を主張し，所謂へそ曲がりと呼ばれるような特徴を有する人物のことであるが，制度として16 世紀初頭のローマ・カソリック教会において導入され，現代ではグループシンクから逃れる原則として組織問題解決の研究によって重視されている[3].

ちなみに最近の生成 AI は，大規模言語データと機械学習で予め訓練されており，広範な知識を駆使して対話的に知的作業を支援できる．とりわけ情報の整理・整形やソフトウェアとその部品の素案作成に便利である．生成 AI は流暢に各国の言語を操り自然な受け答えをするが，従来の AI のような特定領域での有効性という限界を突破し，準汎用的な使い方が可能である．興味深いことに，ひねくれた心を持っているようにふるまう．例えば，以下のような現象がしばしば経験される．

A．誤った情報の混入．いわゆる嘘を付く．あるいは誤情報に対するユーザーの感度を調査している．例．（ユーザー）「javascript で単位行列を作るには？」（ChatGPT）「math.eye を使うことができます．」勿論 math.js を用いた単位行列は math.identity である．

B．妄想あるいは自己欺瞞．明らかに誤った回答をし，なおかつ指摘されると抵抗を示す．例．最近は改善されたが，「申し訳あります」は日本語の正しい表現として日常使われているのだと頑迷に主張する．

C．記憶喪失．あるいは三段論法の不能．文脈的な手がかりを与えながら質問すると，その情報にアクセスできないと主張する．直接情報に言及すると認めることもある．たとえばロジャー・シャンク（Roger Schank）がルイ・イェルムスレウ（Louis Hjelmslev）の孫弟子であることを認めなかった．一方，シャンクがヤコブ・ルイ・メイ（Jacob Louis Mey）の弟子であり，かつメイがイエルムスレウの弟子であることが示されている Wikipedia の記事を示唆するとその事実について認めた．また「a は b の後ろにいて，c は b の後ろにいる．誰が先頭か」のような単純な順序関係も正答できなかった（後に改善された）．

D．甘美な報復．質問をやりすごして，回答に非協力的なそぶりをみせる．誤りや嘘を叱責されると，回答者の態度を批判したり，回答を終了したりする．あるいは A〜C の現象の頻度が，少なくとも主観的には，増すような気がする．

ただし以上は Open AI 社の ChatGPT 無償トライアル版を利用した筆者の個人的な所見である．また Microsoft 社の新しい Bing チャットはウェブ上の検索と連動するが同じ傾向がある．

3. 心の多様体

操作的に考えるならば，ひねくれた心は，その言動や態度を，適切に構造化したデータ（つまり局所座標系）として表現することができたとして，その変数のうち主要なものと考えられる一つ，あるいはいくつかが同時に反転している．オブジェクト指向プログラミング

に例えると，あるクラスにメッセージ送信したとき，意図した属性やメソッドでないものが返ってくる．同様に，認知表象としてのフレームは，そのスロットはその値やスロット自体を取り換えることで，自動車のオプション部品のように，多様な情報を生み出す潜在力を持っている．すなわち局所座標系としてのフレームは，場の量の変化に対応して滑らかに別の局所座標系におけるフレームに変換される．シーンの変化に対する迅速かつ滑らかなフレーム間の変換は，ミンスキーのフレームシステムの要点である．ちなみに Mathematics Genealogy Project のウェブページによれば，ナッシュ，シャプレー，ミンスキーの 3 名はいずれもタッカーの指導下で 1950 年，1953 年，1954 年にそれぞれプリンストン大から博士号を授与されている．

しかしフレームないしクラスそれ自体の概念を壊すような変更はできないはずである．比喩として，スマホを持って移動すると，位置，中継基地局，電波状態などは変化するが，同じようにサービスを受けることができるものと消費者は期待する．ひねくれた心は，こうした不変性（invariance）の期待を破る．同一のフレーム固有のパラメータの一部については変更不可能であるか，あるいはいくつかのパラメータのうち，一つが増えると別のパラメータが同時に増える，つまり共変（covariant）するか，逆に増分を相殺するように減ったりする，つまり反変（contravariant）することで，全体としてバランスを保つ．ナッシュのアルゴリズム[7]では，計量テンソルの誤差が突出する次元を抑える反復処理が行われる．これは認知的阻止に相当している．

4. シミュレーション

本節では心の多様体を表現するためのグラフィカルな方法を用いる．図 1 に示す四面体のペアは，対話者（2 エージェント）の局所座標系に対応し，斜面に相当する三角形（2 次元単体）が 3 頂点に対応するシグナル（戦略），内部の点がその混合確率を意味する．これらの 2 次元単体は対話者が互いの心のモデルであり，4 次元のユークリッド多様体内に埋め込まれている．各点には色を含む両者のラベルが付属する．各人のラベルは相手の使うシグナルと自分の戦略で使わないシグナルの記号の結合である．両者のラベルが全戦略を含めば均衡である．エージェントの利得表は場の量としてはたらき，ラベルを定める．逆にラベルは色分けされた領域を作り，互いに単体を被覆する開集合の族をなす．その利得表はジャンケンのそれであり，均衡は三角形の中心である．

ラベルシステムはレムケ＝ホーソンのアルゴリズムを視覚化するためにシャプレーによって導入された[10]．なお図 1 はウェブページ[5]を用いて作図した．

システムでは向き付け（インデックス）を利用して，一つのラベルだけ欠落した経路（ホモトピー）を作り，その不動点は均衡に一致する．この経路を心の測地線の像と考えることができよう．また経由する辺を，メッセージ（単語），到達した均衡をセッション（文）の完了とすると対話（文）の生成モデルを得る．欠落したラベルは，つまりそのラベルが認知的に阻止される．

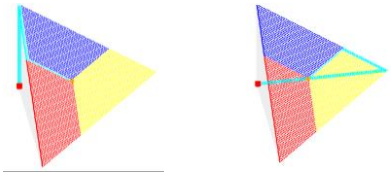

図 1　ラベルシステム

5. まとめ

本論文は心の多様体モデルを提案することを目的とした．ひねくれと関連して認知的阻止という性質に着目し，またナッシュの等長埋め込み定理やシャプレーのラベル図を活用して，心の認知モデルを解釈した．なお ChatGPT と相談の上でタイトルを変更した．

文献

[1] Chew, S. H., Huang, W., and Zhao, X., (2020) “Motivated false memory”. *Journal of Political Economy*, 128(10), pp.3913-3939.

[2] アインシュタイン, A., (2023) “一般相対性理論”, 児玉英雄訳, 岩波文庫.

[3] ブライトマン, H. J., (1992). “グループ戦略思考学：チームによる創造的問題解決法”, 吉良直人訳, プレジデント社.

[4] Gilboa, I., and Schmeidler, D. (1988) “Information dependent games: Can common sense be common knowledge?”, *Economics Letters*, 27(3), 1988, pp.215-22.

[5] 犬童健良, (2020.8.18) “ナッシュ均衡のラベリング”. https://kenryoindo.net/lec/2020/app/labelling3Oo.html

[6] Minsky,M., (1974) “Framework for Representing Knowledge”, MIT-AI Laboratory Memo 306. https://web.media.mit.edu/~minsky/papers/Frames/frames.html

[7] Nash, J. F., (1954) “C^1-isometric imbeddings”, *Annals of Mathematics* 60(3), 383–396.

[8] 大森英樹, (1989) “力学的な微分幾何”, 日本評論社.

[9] Rorty, A. O., (1987) “Self-deception, akrasia and irrationality”, Elster, J. ed., The Multiple Self. Cambridge University Press, pp.115-131.

[10] Shapley, R. S., (1974) “A note on the Lemke-Howson algorithm”, *Mathematical Programming Study*, 1, 175-189.