新進研究者 Research Notes
感情の構成主義理論は我々の素朴心理学的営みに どのような問題点をもたらすか
How the Theory of Constructed Emotion Challenges Our Folk-Psychological Practices

冨田夏美

Abstract

Recent neuroscientific evidence challenges Basic Emotion Theory, increasingly supporting the Theory of Constructed Emotion. This paradigm shift implies that folk emotion concepts are not natural kinds, but socially constructed categories derived from continuous affective experiences so called “core affect.” This paper examines the implications of this constructivist turn, identifying three levels of structural uncertainty in mapping affective experiences to discrete labels: cross-cultural variance in concept mapping, individual classification tendencies and intra-personal instability. We conclude that it will be crucial to build emotion theories that can tolerate and appropriately reflect the diversity in human emotion classification. (95 word)

(1) 研究テーマ

近年、感情心理学においてパラダイムシフトの兆候が顕在化している。特に fMRI を用いた研究の大規模なメタ分析は、長らく感情科学において支配的であった基本感情理論（Basic Emotion Theory）が想定する感情固有の神経生物学的基盤の存在に対し、深刻な経験的疑義を呈している。この結果、感情は生得的なプログラムとしてトリガーされるのではなく、より基本的な感覚入力や概念知識に基づき構成されるとする、構成主義理論（Theory of Constructed Emotion）が有力な代替仮説として急速に台頭しつつある。

この構成主義的転回が妥当であるとすれば、感情という心的状態をめぐっても、かつて Churchland が信念・欲求モデルに関して提起した素朴心理学と科学的心理学の関係性の問題が、新たな論点として浮上することになる。Churchland は、信念や欲求といった素朴心理学の基本語彙は、神経科学の発展によって将来的に消去されるべき不正確な理論的枠組みであると論じた。我々が日常的に用いる「怒り」や「悲しみ」といった感情の素朴概念もまた、Churchland にとっての信念・欲求モデルと同様に、科学的探究の進展によって消去されるべき概念なのだろうか。

本稿では、基本感情理論と構成主義理論の双方おける感情経験と感情概念の関係性を整理し、構成主義理論が正しいと判明した場合、我々の感情にまつわる素朴心理学的営みにどのような問題点が生じることになるかを個人内のケースと個人間のケースに分けて考察する。

(2) 研究の背景・先行研究

具体的な先行研究の確認に入る前に、語彙を整理する。本稿では、頭に血がのぼるような感覚などに代表される現象的な意識経験としての感情を感情経験とよぶ。そして、この感情経験を指して我々が使っている概念、例えば「怒り」を感情概念とよぶ。この感情概念は、複数の粒度を持つ。信念や欲求など、他の心的状態と比較される対象としての感情の概念を類としての感情概念とする。また、類としての感情概念のなかには「怒り」や「悲しみ」など個別具体的な感情経験を指すものが含まれる。これらを類の下位レベルであるカテゴリーとしての感情概念とする（戸田山 2025: 97）。本稿で検討する基本感情理論と構成主義理論の対立は、このカテゴリーとしての感情概念（以下、感情カテゴリーとよぶ）群と、それらが指示する感情経験群の関係性に関するものである。

歴史的な経緯として、感情心理学では基本感情理論が先に成立し、長い間支配的な立場を維持してきた。そしてそれを反証するような形で現在影響力を増しつつあるのが構成主義理論である。そこで、まず基本感情理論の主張を確認する。

基本感情理論の主旨は、素朴心理学における感情カテゴリー群（以下、素朴感情カテゴリー群）のあり方は感情経験群のあり方と対応している、つまり感情経験群と素朴感情カテゴリー群の関係は実体 X と X の概念という関係であるというものである。基本感情理論によれば、これは各感情経験が進化の過程で自然選択された結果残った生存に資する生得的な機能であるからにほかならず、この機能を実現するための神経生物学的基盤が存在する、つまり感情経験群が自然種だからである。

実際、素朴心理学における「怒り」や「悲しみ」といった素朴感情カテゴリー群が、類としての感情概念という集合の要素として離散的に存在しているのは事実である。我々の素朴な認識において、「怒り」は「怒り」以上の何かに還元できず、「怒り」と「悲しみ」は質的に異なる存在であるとされている。基本感情理論は、この素朴感情カテゴリー群の離散的なあり方が、それぞれの感情カテゴリーの指示する感情経験群の実在的なあり方と対応している、すなわち感情経験群は離散的な仕方で実在していると主張する。

この主張を反証するのが、構成主義理論である。もし、基本感情理論の主張が正しく感情経験群が離散的に存在する自然種ならば、特定の感情カテゴリーに固有かつ一貫性のある神経生物学的基盤が見出される必要がある。具体的には、脳内に「怒り」基盤が存在し、その「怒り」基盤が「怒り」ケースでしか動かない(「悲しみ」「喜び」ケース等では動かない)固有性を持ち、かつ、「怒り」ケースでは必ず動く一貫性を持っていることを示す必要がある。しかし、構成主義理論の主要論者である Barrett らによるメタ分析は、そのような神経生物学的な基盤が見出されなかったことを示している(Lindquist, Wager, et al. 2012: 139)。

このように、構成主義理論は基本感情理論の主張する感情経験群の離散的な実在を否定し、感情経験群は連続的な仕方で実在していると主張する。具体的には、構成主義理論は感情経験群がコア・アフェクトという快－不快および覚醒－非覚醒の二次元で表現される連続的な経験の集合であると主張する。基本感情理論が素朴感情カテゴリーこそが感情経験の最小単位であるとしたのに対し、構成主義理論はそれを更に還元しこの二つの次元の組み合わせこそが感情経験の最小単位であると主張しているのである(Barrett 2006: 47)。

では、感情経験が連続的な仕方で実在しているのならば、素朴感情カテゴリーの離散的なあり方との不整合はどのように考えればよいのだろうか。構成主義理論によれば、我々は二次元に還元される連続的な感情経験を離散的な感情カテゴリーに分類(カテゴライズ)することを後天的に学ぶのである。そしてこの分類体系を確立する過程を Barrett は「構成」とよぶ(Barrett 2017: 12)。

この点が本稿の分析において非常に重要である。基本感情理論は、素朴感情カテゴリー群が感情経験群という離散的に存在する実体を指示する概念にほかならない主張する。つまり、感情カテゴリー群と感情経験群は、ある実体とその概念として明確な対応関係が存在していることになる。しかし、構成主義理論においてはそうではない。構成主義理論は、感情経験群という連続的に存在する実体を目的に基づいて人工的に分類したものが素朴感情カテゴリーであると主張する。この場合、感情経験群という実体を複数の仕方で分類して概念化することが理論上可能になり、実体と概念の間の対応関係は明確にひとつに定まらないことになる。

この構成主義理論の主張は、感情経験を色のようにとらえることを提案するものである。光の波長は連続的に変化するスペクトルであり、私たちが「青」と呼ぶ範囲と、「緑」と呼ぶ範囲の間に、物理的に明確な境界線は存在しない。

つまり、我々の素朴概念である「青」や「緑」は自然種として離散的に実在するわけではない。この連続的な光の波長において、どこからどこまでを「青」と呼ぶかという色のカテゴリー分けは、言語や文化によって定義されている。実際、ロシア語には日本語の「青」に相当する単語がなく、「濃い青（sinii)」と「水色（goluboi)」を別の基本色として区別する。この「青」や「緑」に相当するのが、いわゆる「怒り」や「悲しみ」にあたるのである（Barrett 2019: 85)。

構成主義理論において、情動発達はこのような感情概念の発達と同義であると見なされる。構成主義理論において感情概念は特定の文脈における目標への機能をもつ抽象的な実体とされ、情動発達は子どもが自身に有益な行動を導くための感情概念を自らの感覚入力に基づいて構成する能力を獲得するプロセスとして捉えられる。この感情概念を構成する過程において、「怒り」「悲しみ」などの素朴感情カテゴリーが決定的な役割を果たすことになる。感情概念に対応するのは物理的な感覚入力ではなく目標への機能であるため、例えば「怒り」に分類される事例は、障害の克服、脅威からの自己防御、社会的優位性の合図など、様々な文脈を含むことになる。子どもは親が高度に多様な文脈に対して同じ素朴感情カテゴリーを用いてラベル付けするのを観察することで素朴感情カテゴリー群のあり方を学習し、異なる感覚入力をもつ事例に対して目標に基づく機能的類似性を見出すための足場として利用するのである（Hoemann, Xu & Barrett 2019: 8)。

(3) 筆者の主張

前節で確認した通り、基本感情理論と構成主義理論は感情経験と感情概念の関係性について対立する主張を展開している。基本感情理論が主張するように素朴感情カテゴリーは神経生物学的基盤に基づく自然種なのか、もしそうでないならば感情経験は構成主義理論の主張するような二次元に還元できるのか、これらの問いへの答えは経験的研究の進展を待たねばならない。

しかし、ここからは構成主義理論の主張が正しいと判明した場合を想定し、素朴感情カテゴリーが社会的に構成された概念、具体的には感情経験と一貫した対応関係を持たない人工的な分類だとすると、我々の感情にまつわる素朴心理学的営みにどのような問題点がもたらされるのかを検討してみたい。以下、筆者は、この問題を個人内の問題と個人間の問題の二つのケースに分けて検討する。

個人内のケースとしてまず考えられるのが、感情経験の誤分類の問題である。構成主義理論の提出する感情経験の二次元平面（Russell & Barrett

1999: 808）を見ると、例えばSadness（悲しみ）とDisgust（嫌悪）は非常に近い位置にあることがわかる。二次元平面におけるSadness（悲しみ）とDisgust（嫌悪）のこの位置関係を受け入れるのならば、ある刺激に対する反応として強い不快感と中程度の覚醒が生じたとき、それは悲しみである可能性も嫌悪である可能性もあるのである。しかし、一般に素朴心理学において「悲しみ」と「嫌悪」は取り違える可能性があるほど似たものとしては認識されていないため、誤分類の可能性はそれなりに高いといえる。例えば、信頼していた友人に嘘をつかれていたことがわかったとき、そのときの感情を「悲しみ」として認知して行動するか「嫌悪」として認知して行動するかによって、この友人との今後の関係性は大きく変わってくるのではないだろうか。このように、感情経験の分類は非安定的で揺らぎの可能性を大きく孕んでいるように思われる。

これらの個人内の問題よりも一段階複雑さを増すのが、個人間の問題である。素朴感情カテゴリーが言語や文化によって定義される人工的な概念なのだとすれば、違う文化圏の人とはそもそも持ち合わせている素朴感情カテゴリー群の分布が異なることになる。これは、前出の個人内の感情経験の誤分類とはまた違った問題を提起する。なぜなら、前出の問題はある文化の素朴感情カテゴリー群のあり方を受け入れた上でその下で生じる誤分類によるものだったが、この問題は分類の前提とされる素朴感情カテゴリー群自体が異なっていることによるもののだからである。

例えば、前出の色の例において、日本語における「青」に相当する単語はロシア語において存在せず「濃い青（sinii)」と「水色（goluboi)」が基本色と見なされていることを述べた。例えばロシア人の友人に「誕生日に水色（goluboi）の花瓶が欲しい」と言われたとき、それを日本人である私が「青い色の花瓶が欲しい」と解釈し、「濃い青（sinii)」にあたる色の花瓶をプレゼントしたら、その友人は困惑するだろう。なぜこのようなことが起こるかといえば、日本語の「青」はロシア語における「濃い青（sinii)」と「水色（goluboi)」を内包する広い概念だからである。実際、私の手元のスマートフォンで「濃い青（sinii)」と「水色（goluboi)」を和訳すると、どちらも「青」という結果が表示される。この例が示すように、もし素朴感情カテゴリーが社会的に構成される人工的な分類なのだとすると、特定の素朴感情カテゴリーの指示する心的状態を正確に把握するためには、その素朴感情カテゴリーを含む素朴感情カテゴリー群の分布を全体論的に把握する必要が出てくるように思われる。

では、同じ文化圏に所属し、素朴感情カテゴリー群の分布（以下、素朴感

情カテゴリー地図とよぶ）が一致していれば、特定の感情経験の分類結果は必ず一致するのだろうか。そんなこともないように思われる。先ほど個人内における分類の揺らぎの問題について論じたが、これは個人間でも十分に起こりえることだからである。前出の例を利用するならば、私がある感情経験を「悲しみ」と解釈する傾向が高いに対して、私の友人は同じ感情経験を「嫌悪」として解釈する傾向が高い、ということは十分にあり得る話である。このような場合に我々の共通の知人が我々に嘘をついていることがわかったとき、友人は嫌悪感からその知人とは絶交することを提案するのに対して、私はそのような報復行動には賛成せず、まずはその知人の事情を聞くことを提案する、なんてこともあるだろう。現在の心理学において性格は行動傾向として定義されることが多い（Miller 2023）が、上述の友人の性格が「神経質」で私の性格が「寛容」だと評されることが容易に想像できるように、感情経験の分類傾向はまさに性格を大きく決定づける要素のひとつだと考えることができるのではないだろうか。

以上の議論をまとめると、素朴感情カテゴリー群が社会的に構成される人工的な概念だった場合、感情経験の感情概念への分類には以下の三段階の不確実性が構造的に生じることになる。まずは個人間における素朴感情カテゴリー地図の違いによる問題。次に、素朴感情カテゴリー地図を共有する個人間の分類方法の違いによる問題。そして、個人内における分類の通時的な揺らぎの問題である。

このような感情分類における不確実性の問題は、近年の Affective Computing（感情計算論）領域における感情概念の曖昧性や分布を扱う研究潮流と対応するものと考えることができる。従来の感情計算論研究においては、個人間における分類のばらつきは、除去すべきノイズとして扱われるか、平均化や多数決によって単一の正解ラベル、すなわち支配的とされる素朴感情カテゴリーへと集約されるのが一般的だった。しかし、このような還元主義的な仮定は、感情という本質的に主観的であり、表出および知覚の両面において多様な解釈を含みうる現象を扱う上で、必ずしも現実的ではないという見解が研究者間で共有され始めている。したがって、個人間の分類の不一致を単なる誤差として切り捨てるのではなく、人間の感情分類における多様性を許容し、適切に反映できる表現モデルの構築が求められている。しかし、現状の機械学習アプローチは感情のような主観的かつ曖昧な情報を適切に処理する能力が不足しており、個人間の分類の不一致を含む曖昧性のモデル化が重要度の高い未解決課題として注目を浴びている（Sethu, Provost, et al. 2019: 4）。

(4) 今後の展望

本稿では、基本感情理論と構成主義理論における感情経験と感情概念の関係性を整理し、構成主義理論のもたらす感情概念に関する帰結が我々の素朴心理学的営みにどのような問題点をもたらすかを検討した。

冒頭の Churchland が提起した消去主義の問いに立ち返れば、筆者は、素朴感情カテゴリーが自然種としての実在性を欠いていたとしても、それらが対人相互作用において果たしている機能的役割を踏まえれば、必ずしも消去されるべき概念とはいえないと考える。むしろ、これらの感情概念がいかにして社会的に構成され、我々の経験を分節化する道具として機能しているかを解明することこそが、次世代の感情科学のひとつの重要な側面となるのではないだろうか。

同時に、感情経験そのもののメカニズムの解明も引き続き期待される。基本感情理論と構成主義理論の対立では特定の感情経験を引き起こす機能をもつ神経生物学的基盤の有無が焦点となっていたが、構成主義理論の内容にも未だ検討を要する論点は多い。心の哲学の伝統における感情の哲学理論は現在非常に多様なものが提出されており、この問題にも積極的に取り組みたいと筆者は考えている。

このような感情に関する理論的研究は、人間と人工知能の相互作用（Human–computer interaction）という我々の世代にとって非常に重要な領域に新たな視座をもたらすことが期待される。感情経験のメカニズムが理論化され数理的に計算できるモデルが確立されれば、ユーザーが世界をどのように感情的に分節化しているのかという認知の癖を人工知能が学習・模倣できる可能性も出てくるだろう。これは、異文化間のコミュニケーション支援や、個人の主観的現実に寄り添ったメンタルヘルスケア・システムの実現に直結する。我々がよりよい社会を築くために避けて通れない研究領域だといえるだろう。

(5) 参考文献

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences, 35*(3), 121–143.

戸田山和久 (2025). 感情は科学の概念なのだろうか. 植野仙経・佐藤弥・鈴木貴之・村井俊哉（編）*『感情がつくられるものだとしたら世界はどうなるのか』* 金芳社, pp. 80-117.

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science, 1*(1), 28–58.

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12*(1), 1–23.

Barrett, L. F. (2019). *How emotions are made*. Houghton Mifflin Harcourt.

Hoemann, K., Xu, F., & Barrett, L. F. (2019). Emotion words, emotion concepts, and emotional development in children: A constructionist hypothesis. *Developmental Psychology, 55*(9), 1830–1849. https://doi.org/10.1037/dev0000686

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology, 76*(5), 805–819.

Miller, C. B. (2023). Empirical approaches to moral character. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Summer 2023 ed.). https://plato.stanford.edu/archives/sum2023/entries/moral-character-empirical/

Sethu, V., Provost, E. M., Epps, J., Busso, C., Cummins, N., & Narayanan, S. S. (2019). The ambiguous world of emotion representation. *arXiv.* https://doi.org/10.48550/arXiv.1909.00360

（東京大学）