

絵画のスタイル典型度を計測する新しい方法の提案およびドリフト 拡散モデルによる妥当性検証

A New Method for Measuring Style Prototypicality of Paintings and its Validation Using Drift Diffusion Model

緒方 思源¹, 田和辻 可昌², 松居 辰則³
Shigen Fang Ogata, Yoshimasa Tawatsuji, Tatsunori Matsui

¹兵庫教育大学, ²東京大学, ³早稲田大学
Hyogo University of Teacher Education, The University of Tokyo, Waseda University
sogata@hyogo-u.ac.jp

概要

本研究では、絵画のスタイル典型度を計測する実験方法を提案し、ゴッホとゴーギャンの風景画を実験刺激として実験を実施した。また、ドリフト拡散モデル (DDM) を利用して、この方法の認知科学的妥当性を検証した。具体的には、各絵画作品に対して、参加者のスタイル分類タスクにおける回答と応答時間を用いて DDM のドリフト率を推定した。作品のドリフト率の絶対値が分類の容易さを意味し、本実験で計測されたスタイル典型度との間に有意な正の相関が認められた。これにより、提案されたスタイル典型度の計測方法が一定の妥当性を持つ可能性が示唆される。

キーワード：絵画, 典型度, ドリフト拡散モデル

1. 背景と研究目的

絵画の世界には、印象派やポスト印象派といった芸術家集団レベルの流派から、個々の芸術家の画風に至るまで、様々なスタイル (style) が存在する。人が絵画を鑑賞する際、自分が知っているスタイルを用いて絵画作品を分類することが一般的である。なお、同じスタイルに分類される絵画作品でも、そのスタイルにおける典型度 (style prototypicality), すなわち、どの程度そのスタイルの典型例と見なせるかは異なる。

芸術に関する認知科学の先行研究では、絵画作品のスタイル典型度が人々の絵画作品に対する審美的嗜好 (aesthetic preference) に与える影響について、いくつかの実験が行われた。しかし、これらの実験の結果は一致していない。具体的には、Hekkert (1995) の実験では、参加者が「典型的 (typical) - 非典型 (atypical)」と「醜い (ugly) - 美しい (beautiful)」という二つのリッカート尺度を用いて 40 点のキュビズムの絵画を評価した。その結果、絵画のスタイル典型度が美的評価に正の影響を与えていることが報告された。一方、Cutting (2003) の実験では、19 世紀のフランス印象派の絵画を刺激として使用したが、スタイル典型度と美的評価との間には有意な関係が見られなかった。さらに、Farkas (2002)

の実験では、40 点のシュルレアリスムの絵画作品を刺激として使用し、スタイル典型度が最も高い 10 点の作品が他の作品よりも高い美的評価を受けた一方で、スタイル典型度が最も低い 10 点の作品の美的評価と他の作品の評価との間に有意な差は見られなかった。

これらの実験結果の不一致の原因として、使用されたスタイル典型度の計測方法の認知科学的な妥当性が検討されていない可能性が考えられる。この問題を解決するため、本研究では、絵画のスタイル典型度が絵画の感性評価に与える影響を解明することを最終目標とし、絵画のスタイル典型度を計測する新しい実験方法を提案し、その認知科学的な妥当性も検証した。

具体的には、一つのスタイルのみを用いた上記の先行研究とは違って、本研究では、ゴッホ (Vincent van Gogh) の風景画とゴーギャン (Paul Gauguin) の風景画という二つのスタイルの絵画を使用した。これにより、参加者に、絵画作品をどちらのスタイルに分類すべきなのかを積極的に考えさせ、つまり、分類タスクの解決に工夫を凝らすという意味決定の要素を実験に組み込んだ。データ解析の段階では、分類タスクでの意思決定の心理過程を説明する認知モデルであるドリフト拡散モデル (drift diffusion model, DDM) を用いて、実験で計測された絵画作品のスタイル典型度の認知科学的妥当性を検証した。この実験の実施からデータ解析までの手順は次の 3 つのステップに分けられる：

ステップ 1 「スタイル分類の学習」: 参加者は各スタイルの代表作を複数回観察することにより、この二つのスタイルがどの点で異なるのかを吟味した。

ステップ 2 「絵画のスタイル典型度の計測」: 新しいセットの絵画作品を参加者に提示した。参加者は各作品に対して、①その作品がゴッホのスタイルの作品なのか、ゴーギャンのスタイルの作品なのか、②スタイルごとに、その作品がどの程度そのスタイルのものであるのかという二つの設問に回答した。参加者

の設問②での回答を利用して各作品のスタイル典型度を計算した。

ステップ3「スタイル典型度の妥当性検証」:各作品の設問①における参加者の回答を利用して、その作品に対応したDDMのパラメータを推定した。そして、その推定値とステップ2で計測されたスタイル典型度との関係性を分析した。

この妥当性検証の論理について、DDMはドリフト率(drift rate)というパラメータをもっており、その絶対値はある絵画作品の分類の容易さを表す。スタイル典型度が高い作品が、それが低い作品よりも分類が容易と考えられるため、前述の実験方法で計測されたスタイル典型度が認知科学的に適切であれば、スタイル典型度の計測値とドリフト率の絶対値との間に正の相関が見られるはずである。ステップ3ではこの相関が検出されたため、本研究で提案したスタイル典型度の計測方法が一定の妥当性を持つことが示される。

本研究の詳細については、Ogata, Tawatsuji & Matsui (2024)の論文で詳述されており、本稿ではこの研究の主要な部分を取り上げて紹介する。

2. 心理実験

参加者:早稲田大学の学生22名。全員が、①母語が日本語、②絵画の専門教育を受けていない、③色覚が健常であることを条件とした。実験は早稲田大学研究倫理オフィスの承認を得ており(承認番号:2018-HN023)、全参加者からインフォームド・コンセントを取得した。

実験刺激:訓練用セットには、ゴッホの作品8点とゴーギャンの作品8点を、評価用セットには、ゴッホの作品15点とゴーギャンの作品15点を使用した。訓練用セットの作品は各スタイルの代表作である。評価用セットは各スタイルの代表作から非代表作までサンプリングした。実験の準備段階では、13名の西洋絵画の専門家とのインタビューを通じて各作品の代表性を確認した。また、実験中、参加者の回答がゴッホとゴーギャンに関する既存の知識に影響されないよう、ゴッホを「画家A」、ゴーギャンを「画家B」と称した。

実験プログラム:PsychoPy (version 1.90.2)で作成し、MacBook Air PC (15 inch, 2017)で実施した。ディスプレイのキャリブレーションはi1 Display Pro calibratorとDisplayCAL (version 3.7.1.3)を用いて行なった。作品提示時の画像の長辺は650px、背景色は中程度の灰色(CIE $L^* = 50$)に設定した。

手順:実験は2つのセッションから成る。セッション1では、参加者に訓練用セットの作品を注意深く観察し、両スタイルを区別する方法を考案するよう指示した。各作品はランダムな順序で3回、5秒間提示され、作品提示時にはその画家の名前(Painter A または B)が画像の下に表示された。セッション2では、評価用セットの作品がランダムな順序で提示された。各作品に対して、参加者は以下の2つのタスクを完成した:

- 二択強制選択(two-alternative forced choice, 2AFC): 提示されている作品の作者がPainter A か B かを判断した。応答時間はミリ秒単位で記録された。
- VAS評価: まず、その作者がPainter A であると思われる程度を0%~100%のVASで評価した(設問①)。次に、その作者がPainter B であると思われる程度を同様のVASで評価した(設問②)。

刺激間隔(ISI)は2秒であり、その間に一つの白い「+」が画面の中心に提示された。

3. 絵画のスタイル典型度の計算

評価用セットの作品のスタイル典型度を計算する前に、まず、2AFC課題での正解率を用いて、各参加者がこの二つのスタイルを適切に区別できるかを評価した。ゴッホの作品に関して、全参加者が中程度から高い精度を示した($M = 84.1\%$, $SD = 14.4\%$)。ゴーギャンの作品についても、大部分の参加者が中程度から高い精度を示した($M = 77.7\%$, $SD = 20.7\%$)が、2人の参加者は3点の作品のみを正しく判別できた。この2人のデータは以降の分析から除外した。

次に、VAS課題での回答を用いて、評価用セットの各作品を二つの「認知的スタイル」(参加者のスタイル学習によって参加者の脳内に形成された認知的カテゴリ)に分類し、さらに、所属する認知的スタイルに対する典型度を計算した。具体的には、各作品に対して、「ゴッホらしさ」得点は設問①での評価値の参加者間の平均として計算し、「ゴーギャンらしさ」得点は設問②での評価値の参加者間の平均として求めた。ゴッホらしさ得点がゴーギャンらしさ得点よりも大きい場合、その作品を「認知的ゴッホスタイル」というカテゴリに分類し、典型度得点(prototypicality score, PTS)をゴッホらしさ得点の値にした。ゴーギャンらしさ得点の方がより大きい場合、その作品を「認知的ゴーギャンスタイル」というカテゴリに分類し、そのPTSをゴーギャンらしさ得点の値にした。

この分類方法で得られた2つの認知的スタイルは、実世界のスタイルとほぼ同じ作品を含んでいる。「認知的ゴッホスタイル」には12点の作品があり、そのうち11点がゴッホの作品である。「認知的ゴーギャンスタイル」には18点の作品があり、そのうち14枚がゴーギャンの作品である。カイ二乗検定により、認知的スタイルと実世界のスタイルの間に強い連関が認められた ($X^2(1, N=30) = 13.9, p < .001, \text{Cramer's } V = .68$)。

4. スタイル典型度計測の妥当性検証

4.1 妥当性検証におけるDDMの利用方法

第1節で説明された通り、本研究では、DDMを用いて第3節で計算された絵画作品のPTSの認知科学的妥当性を検証した。DDMは、分類タスクにおける意思決定の心理過程を、時間経過に沿って連続的に進行するガウス性のウィーナー過程としてモデリングする。モデリングの際には、刺激ごとに一つのDDMを構築し、すなわち、DDMのパラメータを推定する。Lerche & Voss (2019), Ratcliff & McKoon (2008), Voss et al. (2004)の心理実験の結果によれば、DDMを用いてある刺激に対する人の分類行為をモデリングする際、DDMがカテゴリ（本研究ではスタイルという概念）ごとに、その刺激がそのカテゴリに分類されるという回答の確率、及びその回答の反応時間が従う確率分布とを、高い精度で予測できることが示されている。DDMは、意思決定の閾値 (a)、バイアス (z/a)、ドリフト率 (v) という3つのパラメータを持つ。

意思決定の閾値 (a) は、参加者に速やかな回答を促した場合に生じる回答の正解率と速度のトレードオフを表す。本研究では、回答速度に特別な要求がなかったため、 a の推定はしたが、考察の対象外とした。

本研究では、バイアス (z/a) は、参加者が作品を特定のスタイルに偏って分類する傾向の程度を示す。スタイル学習のセッション（セッション1）では、各スタイルの作品が同数提示されたため、意思決定にバイアスが生じる可能性は低い。そのため、 $z/a=0.5$ （バイアスなしという意味）に設定した。

ドリフト率 (v) の絶対値 (abv) は、ある作品に対する参加者の心理状態が、どのスタイルに分類すべきかに迷う状態から、ある程度の自信を持って判断を下せる状態への移行する（「ドリフト」する）速度を示す。つまり、その作品のスタイルメンバーシップを判断する容易さを示す。一般的にいうと、高い典型度は明確な

スタイルメンバーシップにつながり、作品の分類を容易にする。逆に、低い典型度は曖昧なスタイルメンバーシップに導き、作品の分類を困難にする。したがって、典型度の計測値 (PTS) と abv の間には正の相関が期待される。また、参加者がある作品をゴッホの作品に分類した場合、その回答を「upper response」と称し、 v を正の値にする。ゴーギャンの作品に分類した場合は、「lower response」と称し、 v を負の値にする。

4.2 反応時間の予測分布と実測分布との比較

各作品について、2 AFC 課題での参加者の回答と応答時間という2種類のデータを用いて、一つのDDMを推定した。推定には、Rパッケージ*RWiener* (version 1.3.3) で提供された最尤法の関数を利用した。

各作品に対して、推定されたDDMによって予測された反応時間の分布が実験で記録された反応時間の分布と一致しているかを確認するため、2つのプロットを作成した。図1では、認知的スタイルごとに1点の作品のプロットが示されている。左のプロットでは、反応時間の予測分布からサンプリングされた50個のデータの散布図と、反応時間の実測値の散布図が比較される。右のプロットでは、上記の50個の予測値にフィットした分布曲線と、実測値にフィットした分布曲線が比較される。これらのプロットから、予測分布が実測分布と類似していることが確認できた。

次に、Pearson 相関分析によると、作品のPTSと abv の間に有意な正の相関が認められた（作品全体： $r(28) = .88, p < .001$ ；認知的ゴッホスタイル： $r(10) = .95, p < .001$ ；認知的ゴーギャンスタイル： $r(16) = .79, p < .001$ ；図2で示される）。これらの結果から、本研究で提案されたスタイル典型度の計測方法が認知科学上で一定の程度適切であることが示唆されている。

5. まとめと次の研究段階の課題

本研究では、絵画のスタイル典型度を計測する新しい実験方法を提案し、さらに、DDMを用いてこの方法の認知科学的妥当性を検証した。この成果を活用することで、従来の手法よりも高い精度で絵画のスタイル典型度を測定でき、また、スタイル典型度と他の心理変数との関係を解明するための実験設計にも寄与できる。

次の研究段階では、絵画のスタイル典型度が、絵画に対する美的評価を含む感性評価にどのような影響を与えるのかについて探究する予定である。また、本実験で

図1 反応時間の予測分布と実測分布の比較の例

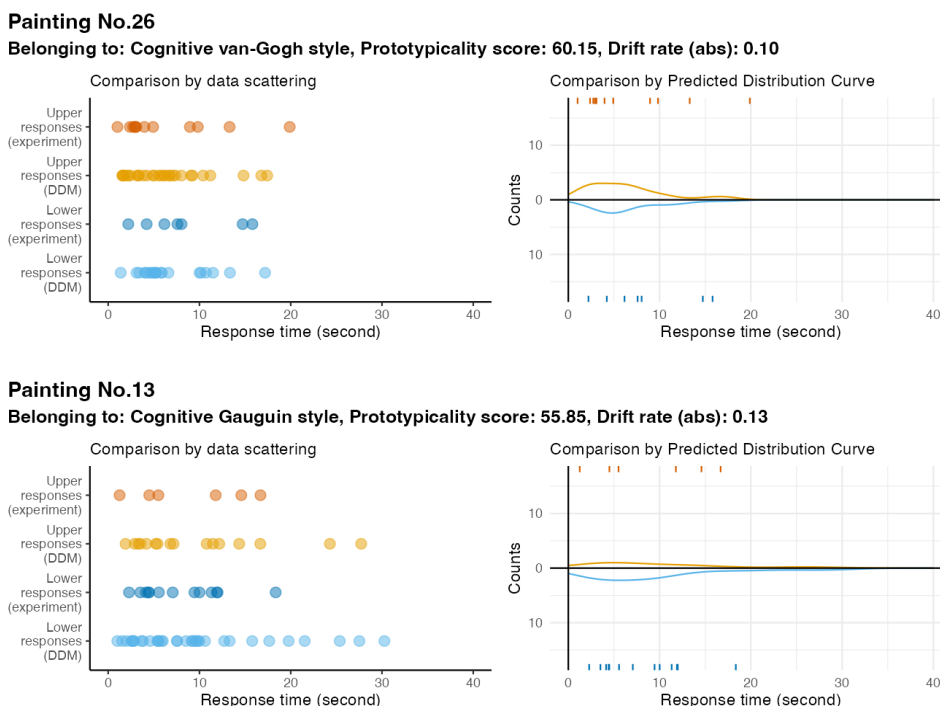
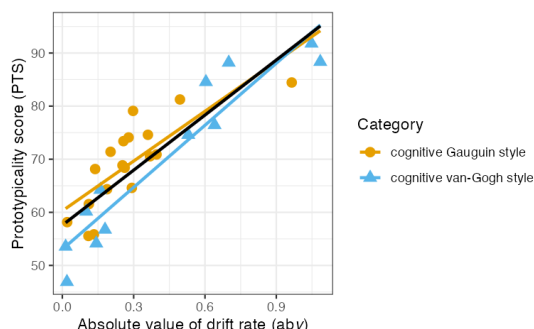


図2 作品のPTSとabvの散布図と相関



は二つのスタイルのみを対象としており、これらのスタイルの選定結果によって実験結果が変動する可能性も否定できない。よって、この計測方法の一般化可能性をさらに評価するためには、今後の研究でより多くのスタイルを対象とした実験を行い、multiple-alternative DDM (Roxin, 2019)を用いて、その実験で測定される典型度の妥当性を検証する必要がある。

謝辞

本研究は早稲田大学特定課題研究助成費（課題番号：2017S-207, 2019E-111）の助成を受けたものです。

文献

Cutting, J. E. (2003). Gustave Caillebotte, French Impressionism, and

mere exposure. *Psychonomic Bulletin & Review*, 10(2), 319–343. <https://doi.org/10.3758/bf03196493>

Farkas, A. (2002). Prototypicality-effect in surrealist paintings. *Empirical Studies of the Arts*, 20(2), 127–136. <https://doi.org/10.2190/UD7Y-GN8P-Q0EV-Q13J>

Hekkert, P. P. M. (1995). *Artful judgements: A psychological inquiry into aesthetic preference for visual patterns* [Thesis technische]. Universiteit Delft.

Lerche, V. & Voss, A. (2019). Experimental validation of the diffusion model based on a slow response time paradigm. *Psychological Research*, 83(6), 1194–1209. <https://doi.org/10.1007/s00426-017-0945-8>

Ogata, S. F., Tawatsuji, Y., & Matsui, T. (2024). Using a drift diffusion model to validate the quantification of style prototypicality as assessed by the viewers of paintings. *Leonardo*, 57(1), 70-78. https://doi.org/10.1162/leon_a_02433

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, 20(4), 873-922. <https://doi.org/10.1162/neco.2008.12-06-420>

Roxin, A. (2019). Drift-diffusion models for multiple-alternative forced-choice decision making. *The Journal of Mathematical Neuroscience*, 9, 5. <https://doi.org/10.1186/s13408-019-0073-4>

Voss, A., Rothermund, K. & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, 32(7), 1206–1220. <https://doi.org/10.3758/bf03196893>