

# ラジカルランダムマイゼーションを用いたフラッシュ 顔変形効果の一般化可能性の探索

益田, 佳卓  
九州大学大学院人間環境学府行動システム専攻

<https://hdl.handle.net/2324/6776839>

---

出版情報 : 九州大学, 2022, 修士, 修士  
バージョン :  
権利関係 :

令和4年度  
修士論文題目

ラジカルランダムマイゼーションを用いた  
フラッシュ顔変形効果の一般化可能性の探索

九州大学大学院 人間環境学府 行動システム専攻  
心理学コース 令和3年度 入学 2HE21202M

益田 佳卓

# 目次

1	序論.....	1
1.1	一般化可能性の危機.....	3
1.2	理論の弱さと頑健な現象の発見.....	7
1.3	ラジカルランダムマイゼーションとメタスタディ.....	9
1.4	フラッシュ顔変形効果.....	15
1.5	本研究の目的.....	17
2	実験1.....	19
2.1	方法.....	19
2.1.1	参加者.....	19
2.1.2	サンプルサイズ及びマイクロ実験設計.....	20
2.1.3	実験計画.....	20
2.1.4	刺激と装置.....	23
2.1.5	顔画像の呈示順番決定方法.....	24
2.1.6	手続き.....	24
2.2	結果.....	28
2.2.1	不気味さについて.....	29
2.2.2	歪みの大きさについて.....	39
2.3	考察.....	48
3	実験2.....	49
3.1	目的と仮説.....	49
3.2	方法.....	50
3.2.1	参加者.....	50
3.2.2	サンプルサイズ設計.....	51
3.2.3	実験計画と分析方法.....	51
3.2.4	刺激と装置.....	52
3.2.5	手続き.....	52
3.3	結果.....	53
3.3.1	不気味さについて.....	53
3.3.2	歪みの大きさについて.....	56
3.4	考察.....	58
4	総合考察.....	59
4.1	本研究の目的と結果の要約.....	59
4.2	ファセットごとの結果の整理と解釈.....	60

4.3	本研究の限界.....	62
4.3.1	実験1・2の共通の限界.....	62
4.3.2	実験1の限界.....	64
4.3.3	実験2の限界.....	65
4.4	今後の展望.....	66
4.4.1	FFDE研究の今後.....	66
4.4.2	ラジカルランダムマイゼーション及びメタスタディの今後.....	67
4.4.3	心理学の再現性問題の今後.....	75
5	結語.....	77
6	引用文献.....	78
7	サプリメンタリ.....	86
7.1	参加者.....	86
7.2	結果.....	87
7.2.1	不気味さについて.....	87
7.2.2	歪みの大きさについて.....	89
7.3	考察.....	91
8	研究倫理.....	91
9	利益相反.....	91
10	オープンプラクティス.....	92
11	謝辞.....	92

## 序論

心理学が科学たりうるためには何が必要なのだろうか。心理学が有用な知見を社会に提供するために求められることには何があるのだろうか。Bem (2011) のいわゆる超能力論文が社会心理学のトップジャーナルである *Journal of Personality and Social Psychology* 誌に掲載されたことを一つのきっかけとして、心理学の知見の中に再現されないものが多く存在することが明らかになり、およそ 10 年が経過した (平石・中村, 2022a; Nosek et al., 2022; Open Science Collaboration, 2015)。再現性は科学の基本原則だが、心理学はいわばその基礎となる部分が揺らいでいるのである。心理学の再現性問題は、その原因や対策を巡って多くの議論がなされてきた (池田・平石, 2016; Shrout & Rodgers, 2018, 植田・益田・佐々木・山田, 印刷中)。これらの一連の流れは信頼性革命 (Credibility Revolution; Vazire, 2018) とも呼ばれる。

心理学の再現性を低下させている原因として、研究者自由度の高さ (Ikeda, Xu, Fuji, Zhu, & Yamada, 2019) や疑わしい研究実践の存在 (Questionable Research Practices: QRPs; John, Loewenstein, & Prelec, 2012), 仮説の事後生成 (Hypothesizing After the Results are Known: HARKing; Kerr, 1988; ただし Rubin (2017) も参照), 出版バイアス (Francis, 2012), 統計的仮説検定の問題 (大久保・岡田, 2012) などが指摘されてきた (詳細は池田・平石, 2016; Shrout & Rodgers, 2018 などを参照されたい)。例えば、実験の計画から論文の執筆・出版に至るまで、研究者が自由に決定できる事柄が多いと (e.g., サンプルサイズ), 意図的に見かけ上の  $p$  値を低め、研究者にとって都合のいい結果が得られたかのように見せかける  $p$ -hacking 等の QRPs を行う余地ができてしま

う (Simmons, Nelson, & Simonsohn, 2011)。これらの QRP が行われると、偽陽性の確率が高くなるため、結果として知見の再現性が低下するのである。

これらの問題に対する解決策として、研究者自由度を縛るアプローチが提案・実践されてきた。代表的なものが、仮説や手続きの事前登録 (pre-registration: プレレジ; 長谷川他, 2021; Nosek, Ebersole, DeHaven, & Mellor, 2018; ただし Rubin (2020) も参照) である。事前登録とは、確証的研究において確認する仮説やその方向、具体的な実験手続きやサンプルサイズとその根拠、分析方法などを第三者機関 (e.g., Open Science Framework: OSF; <https://osf.io/>; Foster & Deardorff, 2017) に記録することを指す。記録した内容は、その日時とともに (原則として) 変更できないようになり、事前登録からの逸脱が明確になるため、HARKing などの QRP を防ぐことが期待されている。しかし、事前登録のみでは全ての QRP を防ぐことはできないことが指摘されており (Ikeda et al., 2019; Yamada, 2018)、近年では事前登録を発展させ出版スタイルに組み込んだ査読付き事前登録 (Registered Reports: 以下、レジレポ; Chambers, 2013; Chambers & Tzavella, 2022) を受け付ける雑誌も増えてきた。レジレポでは、まず一般的な論文における序論と手続きに当たる部分の原稿を執筆し、そこまでの内容について査読を受ける (Stage 1: 第一段階)。査読を通過すれば、原則的採択 (in-principle acceptance: IPA) となり、結果が仮説を支持するかや統計的に有意な結果かに関わらず掲載が担保される。その後、実験や調査などのデータ収集を行い、結果と考察に相当する原稿を執筆し、Stage1 のプロトコル通りに実験が行なわれたかの確認、実験結果の考察等について再び査読を受け (Stage 2: 第二段階)、それを通過すると論文として出版される。これにより、研究者たちが *p*-hacking や HARKing を行なうインセンティブを減らすことができ、通常の前記登録では防ぎきれない QRP や出版バイアスを防止することが期待される。

また、刺激や実験プログラム、ローデータや分析プログラムなどを公開するオープンデータ・マテリアルも推奨され、多く行われるようになってきている (e.g., Munafò et al., 2017; 武藤, 2022)。これらが公開されることにより、研究の透明性が向上するだけでなく、先行研究と同一の刺激・実験プログラムを用いた直接的追試が容易になり、データの二次分析も可能になる利点がある。学術雑誌によっては、どのレベルでオープンサイエンスが行われているかを示すバッジを付与しているものもある (Kidwell et al., 2016)。なお、刺激やローデータ等の公開には OSF 等のサービスが用いられることが多いが、オープンサイエンスは (レジレポ含む) 事前登録を行っていない場合でも可能である。

それでは、事前登録やレジレポ、オープンサイエンスなどを実践していけば、心理学は信頼性を高め、有益な知見を提供できるようになるのだろうか。およそ 10 年に及ぶ議論の中で、心理学が直面する課題は再現性の他にも多くあることが指摘されてきた。例えば、測定の危機 (Questionable Measurement Practices: QMPs; Flake & Fried, 2020) や検証の危機 (Schimmack, 2021) である。これらの指摘はいずれも再現性問題においても重要であると考えられるが、その中でも特に重要な問題提起として、一般化可能性の危機 (Yarkoni, 2022) と理論の危機 (Eronen & Bringmann, 2021; Oberauer & Lewandowsky, 2019) が挙げられる。

### 一般化可能性の危機

一般化可能性の危機とは、心理学の知見は従来主張されてきたほど一般化することができないのではないかという指摘である (Yarkoni, 2022)。定量的な心理学研究では、一般に母集団 (e.g., ヒト一般) から多くの参加者を募り、実験や調査・観察などを通じてデータを収集する。サ

サンプルサイズを十分に大きくすれば個人差の要因の効果は、ランダム化によっていわば「洗い流され」相殺されるため無視できるようになる。これにより、母集団レベルでの推論、つまり標本で得られた結果は母集団でも同様に見られるという一般化が可能になる。それでは、サンプルサイズを大きくすれば結果の一般化可能性は担保されるのだろうか。Yarkoni (2022) は、それだけでは不十分であるとしている。その理由に、未測定の変数の影響を挙げている。言うまでもなく、ヒトの行動や心的過程には様々な要因・変数が影響する。この中には、参加者の個人差や刺激の特性、文化的要因といったものに加え、実験者の教示から実験室の明るさに至るまであらゆるものが含まれる。こうした様々な(未測定なものを含めた)変数のランダム効果を考慮すると、心理学の知見はごく限られた範囲(e.g., その研究の実験状況)でのみ確認されるものになってしまうというというのが Yarkoni (2022) の主張する一般化可能性の危機である(平石・中村, 2022b も参照)。

Yarkoni (2022) の指摘以外にも一般化可能性を制約する要因は様々ある。例えば、心理学研究の多くが、欧米ないし西洋化された文化圏からのサンプルに偏っていることが指摘されている(e.g., Arnett, 2008)。これは WEIRD (Western, Educated, Industrialized, Rich, and Democratic) とも呼ばれ、この傾向は近年でも依然として大きく変化していない(Thalmayer, Toscanelli, & Arnett, 2021)。また、日常場面への心理学研究の知見を適用・応用する際には、生態学的妥当性も重要になってくるだろう。

それでは、一般化可能性が低いとどうなるのだろうか。第一に、手続きや刺激、文脈等が僅かに変化しただけで、知見が確認されなくなる事が挙げられる。これは現象や効果の存在範囲が狭いことを示唆する。そして、知見が限られた状況下でのみ確認されることは検証が行いに

くいため、説明理論の構築を阻害しうる (Eronen & Bringmann, 2021)。これにより理論が構築されにくいため、その現象の理解が進まず、よって理論の構築が進まないという悪循環を生む可能性がある。第二に、心理学の知見 (e.g., 基礎研究) を現実場面に応用し問題解決等を図る際には、その適用が限定的あるいは実質的に適用できなくなる可能性が考えられる。なお、これには効果量の大きさも関係するだろう (practical significance, Torkar et al., 2022)。第三に、一般化可能性が低いと、ある時代のある場所での結果というように、制約の強い記述しかできなくなる。いわばスナップショット的な研究になるのである。このような議論の中で、どの程度まで結果を一般化できると考えられるかを論文中で宣言すべきという提案もなされており、これは一般性制約 (Constraints on Generality: COG; Simons, Shoda, & Lindsay, 2017) と呼ばれている。その中には、理論的・経験的な境界条件や重要と考えられる要因、時代・時間依存性などに言及することが推奨されている。

一般化可能性が低いことは、追試の成否の解釈にも影響を及ぼす (平石・中村, 2022a, 2022b)。再現性問題への注目に伴い、追試の重要性が強調されるようになってきた。追試とは、知見の再現性や頑健性を検証するために、もう一度実験や調査・観察を行い先行研究と同様の知見が得られるかを検討することである。追試には大きく2種類あり (Zwaan, Etz, Lucas, & Donnellan, 2018)、一つは先行研究の仮説や手続きを可能な限り忠実に再現し、同じ結果が得られるかを検討する直接的追試 (direct replication) である。もう一つが、理論から予測される第二・第三の仮説を検証したり、手続きに変更を加え知見の頑健性や一般化可能性を検討したりする概念的追試 (conceptual replication) である。これらはいずれも重要な研究手法であるが、ここで

取り上げるのは追試が失敗したときの結果の解釈である<sup>1</sup>。一般に、直接的追試において、その結果が仮説(ないし先行研究の知見)を支持しなかったとき、追試は失敗したとみなされる。しかし、この結果をそのまま追試が失敗したと解釈してよいのだろうか。直接的追試は、前述のように先行研究をそのままの仮説・手続きで再現することを試みるものである。それでは、そもそも現実的に先行研究と「完全に同一の」追試を行うことは可能なのだろうか。仮に、刺激や手続き、実験環境を先行研究と全く同一にできたとしても。しかし、その場合であっても参加者は以前の研究とは別であることが多い。先行研究と同一の参加者で実験できたとしても、以前の実験に参加したときの記憶や経験があるため、当時とは同一とは言い難い。また、季節や時刻・時代といった時間的要因も異なるかもしれない。このことは、すべての追試は「概念的」追試であり、厳密な意味で完全に先行研究と同一の追試 (exact replication) は不可能であることを示唆している (平石・中村, 2022a, 2022b; Zwaan et al., 2018)。そしてこれは同時に、追試の失敗の解釈が複数可能であることを意味する。一つ目は、素直に対象の現象や効果が存在しない(つまり先行研究の知見が誤り)と解釈することである。二つ目は、手続きや教示が適切ではなかったなどの理由から、実験が適切に行われなかったと解釈することである。そして、三つ目は一般化可能性の問題と解釈することである。これは、全ての追試はどれほど先行研究に忠実に行おうとも僅かな違いが生じるため、その違いによって結果が異なると解釈することにほかならない。このことから、心理学の再現性問題は一般化可能性の問題の部分集合であると考えられる。

---

<sup>1</sup> なお、何をもって追試が「成功」ないし「失敗」したとみなすかは議論が続いている (平石・中村, 2022a, 2022b; Peels & Bouter, 2021)。

ここまで見てきたように、現状の心理学では、手続きをオリジナルの研究と可能な限り揃えた追試であっても、その知見が再現されないことが多いことが明らかになってきた。そして、再現されなかった場合、その結果の解釈を巡って混迷を究めることも示されてきた(平石・中村, 2022a, 2022b)。つまり上記三つの解釈のうち、どの解釈が適切であるかの判断が困難なのである。そして、再現されたとしてもその知見がどれほどの一般化可能性を有するかについての主張は慎重になる必要が指摘されている(Yarkoni, 2022)。このことは、知見の一般化可能性が低く、実験・研究の文脈依存的である可能性を否定できないため、わずかな手続きの違いで再現されなくなる可能性を示唆している。加えて、一般化可能性の問題は、すでに提唱されている研究者自由度の制限や出版バイアスなど再現性問題への対策だけでは解決しにくいと考えられている(平石・中村, 2022a, 2022b)。このことは、仮に再現可能性問題が収束したとしても、心理学にはまだ立ちふさがる課題、一般化可能性の問題が存在するというを示唆している。

### **理論の弱さと頑健な現象の発見**

これら心理学が直面する諸課題の背景には、心理学の理論が弱いことが関係していると考えられる(e.g., Eronen & Bringmann, 2021; Oberauer & Lewandowsky, 2019)。Oberauer & Lewandowsky (2019)は、科学的推論を理論と仮説の関係である理論レベルと、仮説とデータの関係についての実証レベルに区分した。そのうえで、実証レベルの取り組み(e.g., 事前登録)を行ったとしても、理論と仮説のリンクの強さ(理論から適切な仮説が導ける確率)には影響しないうえ、そのリンクが弱い(理論が弱い)と適切な仮説を導くことができないため、偽陽性率が高くなることをシミュレーションで示している。そして、このことが心理学における再現性問題の原因の一

つであるとして、心理学は理論の危機に陥っていると指摘した。その上で、彼らは現象や効果の発見を目指す発見志向型研究ではなく、(特に理論と仮説のリンクが強い時には)理論の検証を目的とする理論検証型研究に移行すべきだと提案している。しかし、現状心理学の多くの理論は弱く、具体的かつ定量的ないし効果量までの予測が行えるものは少ない(e.g.,「ある刺激を1300 ms, ○○という条件で呈示したとき, ヒトは平均的にその呈示時間を1324 ms程度に知覚する」といったように具体的な数値予測ができる時間知覚の理論・モデルは現状存在しない)。また、理論的な検証を行う際には、目的の効果・現象が再現される必要があるが、ここまで見えてきたように、現状ではその確率が低い領域も存在する。このような状況下でどのように強い理論を構築し、検証していくのかについて Oberauer & Lewandowsky (2019) は明確な回答は述べられていない(なお、数理モデリング等の定式化の重要性は強調されている)。

ここで、同様に心理学の理論の危機を指摘している Eronen & Bringmann (2021) は、優れた理論を構築するためには、頑健な現象(データ)が必要であるとしている。すなわち、再現性の高い知見が存在して初めて理論構築が可能になるのである。そして、その知見の存在範囲・適用範囲の経験的な裏付けや影響する要因の明確化も重要である(Scheel, Tiokhin, Isager, & Lakens, 2021)。当然、これらの取り組みはこれまでの心理学でも行われてきた。しかし、再現性問題を鑑みると、それだけでは十分ではなかったと考えられる。その背景には、ヒトの心的機能や行動に関係する要因の多様さ、そしてそれに制約を与える理論も弱いため検討すべき要因を絞り込むことが困難であることが関係していると考えられる。それでは、Eronen & Bringmann (2021) の言う頑健な現象の発見はどうすればよいのだろうか。

一つの方法は、より幅広く経験的な裏付けを行うことであろう。すなわち、当該の現象や効果が頑健であるか、どの程度まで一般化可能なのかを実際確認することである。頑健性の指標にも様々あると考えられ、直接的追試による特定条件下での再現性の高さ、様々な測定手続き・課題においても同様の結果が確認されるか (cf. 概念的追試)、その現象・効果にとって重要でないと考えられる変数・要因の影響を受けない (受けにくい) ことなどがあると考えられる。また、マルチバース分析 (Steenen, Tuerlinckx, Gelman, & Vanpaemel, 2016) や仕様曲線解析 (Simonsohn, Simmons, & Nelson, 2020) 等を用いた統計解析における頑健性の確認も関連する。

特定の条件下での再現性は、研究者自由度の制限やオープンサイエンスなどである程度対策がなされると仮定しても、結果を一般化できる範囲に関しては実際に検討してみるまで経験的な証拠はない。考えられうる手続きの中で、どれほどの範囲で現象・効果が確認されるか、すなわち一般化可能性の探索は、理論の構築や検証において経験的な証拠が理論に対して制約を課すという点で重要な意味を持つ (Scheel et al., 2021)。しかし、これまで一般化可能性を体系的に検討する方法はおそらくあまり普及しておらず、それゆえ検討が進まなかったという側面もあるだろう。そこで、近年それを効率的に検討するための手法が提案された。それが、Baribault et al. (2018) が提唱するラジカルランダムマイゼーションとメタスタディである。

### **ラジカルランダムマイゼーションとメタスタディ**

ラジカルランダムマイゼーション (Radical Randomization) とメタスタディ (Metastudies) とは、理論の頑健性や一般化可能性を検討するための新しい実験方法である (Baribault et al., 2018)。この実験方法の中核となる特徴は、考えられうる変数を全てランダムマイズすることで、一定の刺激

を使用するのではなく、事前に定義した集合からランダムに選ぶことにある。Baribault et al. (2018) は、逆向マスクングの実験を例に、このアイデアを説明している。逆向マスクングとは、時間的に連続して呈示される2つの刺激のうち、後続する刺激が先行する刺激の知覚を阻害する現象である(熊田, 2019)。逆向マスクングの一般的な手続きでは、ターゲット刺激が呈示された直後にマスク刺激が呈示される。このとき、オリジナルの研究ではマスク刺激は「&&&」であったとする。それでは、この一連の手続きにおいて逆向マスクングが生じるためには、マスク刺激が「&&&」である必要があるのだろうか、「%%」では逆向マスクングは生じなくなるのだろうか。この実験において核となる手続きは、ターゲット刺激の直後に(なにかしらの)マスク刺激が呈示されることである。そのため、マスク刺激が「&&&」であることは重要ではなく、「%%」でも逆向マスクングは生じると考えられる(Baribault et al., 2018)。換言すれば、マスク刺激の種類という特徴は研究者によって「恣意的に」選ばれているのである。ここで、このような恣意的に選ばれている様々な変数に変化を加えても、現象・効果が確認されるのであれば、これらの特徴については頑健であると推測することができる。

ラジカルランダムマイゼーションでは、このように研究目的において不可欠な(そして変えてはいけない)特徴以外を事前に定義した範囲でランダムマイズすることで、目的の現象・効果の頑健性や一般化可能な範囲を検討する。この時、研究目的において、理論上・仮説上不可欠な独立変数(現象・効果を定義するために必要な独立変数と言い換えても良い)を基本独立変数(elementary independent variables)と呼ぶ。マスクングの例では、マスク刺激の有無が基本独立変数に相当する。そして、研究目的にとっては重要ではないと考えられる要因、ないし今回頑健性を検討する種々の要因はファセット(facet)と呼ばれる。これには、マスク刺激の呈示時間や

画面の背景色 (e.g., 灰色, 白色, 黒色) などが含まれる (当然これら以外でも良い)。ラジカルランダムマイゼーションでは, このような「恣意的に」選ばれた特徴を複数個, ファセットとして組み込み, その水準をランダムに決定し実験を繰り返し行う。ファセットは, その値がランダムに選ばれる独立変数とも言える。そして複数のファセットからなる, 実験計画法におけるセルに対応する概念を マイクロ実験 (microexperiment<sup>2</sup>) と呼ぶ。ファセットの値は水準 (levels, 実験計画法における条件・condition に相当) と呼ばれ, マイクロ実験の開始時に, 研究者が決定した任意の分布からランダムに決定される。なお, それぞれのマイクロ実験内では, 各ファセットの水準は一定である。各ファセットとその水準が抽出される分布 (ないし範囲) から, 研究者が当該研究において目的の効果・現象の存在を確認することを目指す範囲である, 意図一般化母域 (universe of intended generalization) が定義される。なお, 意図一般化母域は研究者が任意に決定する。このように複数のファセットを含む研究をメタスタディと呼ぶ。

ラジカルランダムマイゼーションの中核となるアイデアは, ファセットの値に合理的変動域 (reasonable variation) を持たせることである。これは, 考えられうる中で可能なファセットの組み合わせからなる母集団である, 手続き空間 (method space) の中の幅広い範囲において効果の存在を検討するためである<sup>3</sup>。言い換えれば, ラジカルランダムマイゼーションとは, 手続き空間 (実際には意図一般化母域) からマイクロ実験をランダムサンプリングすることに相当する。

---

<sup>2</sup> DeKay, Rubinchik, Li, & De Boeck (2022) ではマイクロスタディ (microstudies) と呼称されている。

<sup>3</sup> 意図一般化母域は手続き空間の部分集合であると考えられる。

それでは、なぜこのような実験方法を取る必要があるのだろうか。1つ目は、一般化可能性との関連である。Baribault et al. (2018) は、心理学者はごく少数の条件や刺激から広い範囲の条件・刺激へ一般化して考える傾向があり、実験の結果は、ある特定の実験状況のみのものだと考えないと述べている。例えば、ある手続きによって記憶が促進されたとしよう。その場合、心理学者はその実験で用いた刺激でのみ、記憶促進効果がみられ、他の刺激に変更した時には効果が確認されなくなるとは考えないということである。しかし、Yarkoni (2022) の一般化可能性の危機の指摘を踏まえると、結果の一般化は慎重に行う必要があるだろう。適切に一般化可能な範囲を主張するためには、様々な刺激や条件、要因を実際に検討し、その結果をもとに行うことが理想的であると考えられる。しかし、現実問題として広範な要因や条件・刺激の全パターンを調べるのは非常に困難である。そこで、Baribault et al. (2018) は、一般化可能性を効率的に検討するために、手続き空間からマイクロ実験としてその部分空間をランダムサンプリングすることを提案しているのである。これは、母集団 (e.g., ヒト一般) 全てを調べるのが (事実上) 不可能であるため、そこから標本 (実験参加者) をランダムサンプリングし実験を行い、その結果から母集団の特徴を推測することと本質的に同じである。以前から刺激を母集団からサンプリングしたものとして扱い、そのランダム効果を考慮すべきだという指摘はされており (e.g., Monin & Oppenheimer, 2014; Judd, Westfall, & Kenny, 2012; Wells & Windschitl, 1999), ラジカルランダムマイゼーションはこの流れの中に位置づけることができると考えられる。また、ラジカルランダムマイゼーションに類似した試みは、モデルの頑健性のテスト方法 (Weisberg, 2013 松王訳 2017) においても提案されている。

2つ目の理由は、厳密な直接的追試 (exact replication) の不可能性がある。前述のように、どれほど配慮したとしてもオリジナルの研究と厳密な意味で全く同じ実験を行うことはできない。言い換えれば、元の研究とは異なる部分が必ず存在する。その違いの中には、研究目的や仮説にとって重要ではない特徴も含まれる (e.g., 画面の背景色)。ラジカルランダムマイゼーションは、このような将来の追試で生じるかもしれない潜在的で偶発的な (実験デザイン上の) 変動を模したものであるため、結論をより強固にすると主張されている (Baribault et al., 2018)。

Baribault et al. (2018) はメタスタディの分析方法として、グローバル検定 (Global Tests)、レベル2 異質性・調整分析 (Level-2 Heterogeneity and Moderation)、計画的メタ分析 (Planned Meta-Analysis) の3つを提案している。グローバル検定は、固定効果モデルに相当する分析である。十分にサンプルサイズが大きければ、参加者の個人差のランダム効果はランダム化によって「洗い流され」無視できる。同様に、ファセットのランダム効果もサンプルサイズ (この場合はマイクロ実験) が大きければ無視することができる。ただし、この分析はマイクロ実験間で結果が比較的均質である場合のみ有効であるとされている。グローバル検定が適切ではない場合は、ファセットが切片を含む基本独立変数と交互作用を持つことができる階層線形モデルを用いた分析を行う。いわゆるランダム効果モデルである。これをレベル2 異質性・調整分析と呼ぶ。これにより、現象や効果が各ファセットの範囲で (どのように) 変化するかを検討できる。効果量  $\beta$  に対するファセットからのレベル2分散を  $\alpha$  としたとき、 $\alpha$  は効果量の異質性を定量化する指標となる。この  $\alpha$  を  $\beta$  で割った値  $\rho$  は変動係数 (coefficient of variation) と呼ばれ頑健性の指標として解釈される。 $\rho$  が小さい時 (Baribault et al. (2018) は 1/3 や 1/4 としている)、そのファセットについては頑健であると解釈することができる。また、実施されたマイクロ実験を通常

のメタ分析における一つの実験とみなして行う計画的メタ分析も提案されている。一般的なメタ分析は論文等で刊行された研究を対象に行うため、分析対象の出版バイアスなどの問題がある。しかし、メタスタディにおける計画的メタ分析では、分析対象は全てその研究で実施したマイクロ実験であるため(すなわち出版された研究を対象としたものではないため)、出版バイアスの懸念がなく実施可能とされている (Baribault et al., 2018)。また、マイクロ実験間での差異も定義されているなどのメリットもある。なお、メタスタディにおけるメタ分析については DeKay et al. (2022) も参照されたい。

このようにラジカルランダマイゼーション・メタスタディを用いた研究を行うことで、多数の要因(正確にはファセット)やその水準の実験であっても、目的の現象・効果の境界条件や理論の適用範囲を効率的に探ることができる。これは、一般化可能な範囲を論文中に明記する上でも有用である(cf. 一般性制約: COG)。これは、その記述に経験的な裏付けを持たせることができるようになるためである(当然、検討した範囲で境界条件がみられないのであれば、それも重要な知見である)。そして、このことは理論構築にも役立つと考えられる。ラジカルランダマイゼーション・メタスタディの結果、頑健な現象であると分かれば、それを足がかりに理論的検討を進めることができ (Eronen & Bringmann, 2021)、境界条件が見つければ、それは理論構築に対する経験からの制約条件となる(境界条件よりも広い範囲で現象の存在を予測するような理論は経験的事実に照らして不適切であると推測することができる)。

このように有用な実験手法であると考えられるラジカルランダマイゼーション・メタスタディは、2018年に提案されてからおよそ4-5年経過したが、本稿執筆時ではこの手法を用いた研究はほとんど実施されていない。(提案者である Baribault et al., 2018を除いて)これらの手法をもち

いた唯一の研究としては、DeKay et al. (2022) が挙げられる。彼らはフレーミング効果についてラジカルランダマイゼーションとメタスタディを用いて検討している。そのため、知覚・認知心理学実験における実践は (Baribault et al., 2018 を除いて出版物としては) 未だ存在しない。このためか、ラジカルランダマイゼーション・メタスタディの実践にかかる諸問題や検討課題、それらの解決策についての議論はほとんど行われておらず、これが手法普及のハードルとなっている可能性がある。そこで本研究では、フラッシュ顔変形効果の実験を題材に、ラジカルランダマイゼーション・メタスタディの有用性や課題などを検討することを目的とした<sup>4</sup>。

## フラッシュ顔変形効果

周辺視野に目の位置を揃えた顔刺激を、間隔を空けず瞬間的に連続呈示すると、顔の形が歪んだりグロテスクに知覚されたりすることが報告されている。この錯視はフラッシュ顔変形効果<sup>5</sup>と呼ばれている (Flashed Face Distortion Effect: 以下, FFDE; Tangen, Murphy, & Thompson, 2011)。この錯視は、目の位置が揃っていない場合や、顔刺激間にブランク (何も呈示しない瞬間) を挟むと弱まることが知られている (Tangen et al., 2011)。周辺視は中心視に比べ視力が低いが、Bowden, Whitaker, & Dunn (2019) は、この顔刺激が呈示される視野の違いが FFDE の強さに及ぼす影響を検討した。実験では中心視で刺激を観察する条件、周辺視で観察する条件、周辺視での刺激呈示を模してぼかした (ブラーをかけた) 刺激を中心視で観察する条件の 3 条件を比

---

<sup>4</sup> なお、FFDE を対象とした理由は、筆者の興味やオンライン実験・ラジカルランダマイゼーションの実施可能性などである。

<sup>5</sup> この錯視のデモンストレーションは以下の URL から確認できる (<https://youtu.be/wM6lGNhPujE>)。

較した。その結果、顔刺激が周辺視に呈示された時に FFDE は強く生じ、中心視に呈示された場合には FFDE がほとんど生じなかった (Bowden et al., 2019)。また、ぼかした刺激を中心視で観察した条件は、ぼかしていない刺激を中心視で観察する場合よりは FFDE が有意に強く生じたが、周辺視で観察する場合よりは有意に弱かった。このことから、FFDE の生起には顔刺激が周辺視野で呈示されることが重要であると考えられる。一方で、刺激の大きさや化粧の有無 (Balas & Pearson, 2019)、色やコントラスト (鈴木・氏家・高橋, 2022) は FFDE の強さに影響しにくいことが報告されている。

これらは比較的低次な知覚処理と FFDE の関係を検討したものであるが、顔の全体処理のような高次な処理との関連も示唆されている。鈴木他 (2022) は、呈示する顔刺激の部位が FFDE に及ぼす影響を、顔全体を呈示する条件、顔の上半分のみ呈示する条件、目のみが無い条件、目のみが呈示される条件、輪郭がない条件の計 5 条件で検討した。その結果、顔全体が呈示されるとき、最も FFDE が強かったことを報告している。また、目のみが無い条件や目のみが呈示される条件は、顔全体が呈示されるより FFDE は弱かった。一方で、輪郭が無い条件と顔全体が呈示される条件の間には有意な差は見られなかった。このことから、FFDE には目の呈示が重要であるものの (Tangen et al., 2011)、目のみでは生じないことが示唆された。また、顔全体が呈示されているときに最も歪みが強く知覚されたことから、顔の全体処理が FFDE に関与している可能性も示唆された。顔の全体処理の重要性は、顔刺激がいわゆる他人種やサル、ポジネガ反転、刺激を倒立させた時に FFDE が弱まることから示唆されている (Utz & Carbon, 2015)。ただし、顔刺激の倒立については、Tangen et al. (2011) 及び Balas & Pearson (2019) は顔刺激の倒立させた場合でも FFDE の強さは大きく変化しないと報告している。

FFDE のメカニズムについては未だ不明な点が多く、先行研究でも見解が一致しないものもある。このことから、FFDE に影響を及ぼす要因の特定や、その影響がみられる (ないし FFDE が生じる) 範囲を明確にすることは、その生起メカニズムを検討する上でも重要であると考えられる。ここで、ある現象に影響する要因や、その現象の生起条件 (範囲) を探るための方法としてはラジカルランダムマイゼーション及びメタスタディが有用であろう (Baribault et al., 2018)。

## 本研究の目的

そこで本研究は、FFDE に影響する要因と FFDE が生じる範囲の探索を通じて、ラジカルランダムマイゼーションの有用性や課題を検討した。本研究の主たる目的はラジカルランダムマイゼーション・メタスタディの実施とその有用性の検討<sup>6</sup>、今後議論すべき論点の整理であり、FFDE に対する検討はあくまで二次的な目的である。

本研究では、ラジカルランダムマイゼーション・メタスタディを知覚・認知心理学実験に組み込むための一つ方法として、以下の流れを実施・提案する。第一段階として、頑健で再現可能な知見を発見、ないし目的の効果・現象に影響する要因 (ファセット) を探索するための手法としてラジカルランダムマイゼーション・メタスタディを用いた実験を行う。これは本研究の実験 1 に相当する。この段階の主たる目的は、様々なファセットを組み込み、その影響を検討することで、次に検討すべき要因を探索することである。すなわち、複数の要因の影響を効率的に探索するための手法にラジカルランダムマイゼーションを用いるという提案である。一般的な実験計画法に比べ、参加者は一定回数のマイクロ実験を行うだけでよいため、要因 (ファセット) や

---

<sup>6</sup> なお、本研究はその中でもラジカルランダムマイゼーションの側面を中心に実施・検討した。

水準が多い実験であっても参加者負担を一定に保ったまま、実験を実施できる利点がある。ただし、この段階は探索的研究であるため、その結果を確証的なものと解釈することは適切ではない。そこで第二段階として、実験計画法を用いた確証実験を行う。これは本研究の実験2に相当する。この段階では、前段階で確認された結果 (e.g., ある現象に影響する要因の発見) が再現されるかを実験的に改めて検証する。そして、本研究では時間的な制約から行えなかったが、第三段階として再度ラジカルランダマイゼーション・メタスタディを行い第二段階の知見の再現性・一般化可能性を検討する。これは、Baribault et al. (2018) が提唱したラジカルランダマイゼーション・メタスタディの使用方法である。そして、理想的にはこれら第一から第三段階を繰り返し実施し知見の一般化できる範囲を明確化していくことが望ましいと考えられる。以下では、ラジカルランダマイゼーションを知覚・認知心理学研究に組み込むための流れを FFDE の検討を通じて実践した結果を報告する。

## 実験 1

実験 1 では、ラジカルランダムマイゼーションとメタスタディを用いて、様々なファセットと FFDE の関係を探索的に検討した。これにより、各ファセットが FFDE の強さに及ぼす影響と今回検討した意図一般化母域において、FFDE がどれだけの範囲で確認されるかを検討することができる。なお本実験は探索的研究であるため、各ファセットが FFDE に及ぼす影響については、具体的な仮説は立てなかった。本実験は、OSF に行った事前登録に基づいて実施された (<https://osf.io/x4kfs>)。

## 方法

### 参加者

実験にはインフォームドコンセントを取得した上で、150 名が参加した。参加者の募集は Yahoo! クラウドソーシング (<https://crowdsourcing.yahoo.co.jp/>) を用い、オンライン上で実験を行った。参加者には報酬として 120 ポイント (日本円で 120 円相当) が支払われた。なお、この金額は実験時間 5 分あたり 20 ポイントとして計算した (実験 1 は実施時間がおよそ 30 分とみなし報酬額を算出した)。

事前登録した除外基準に基づき、全回答のうち 90% 以上で同じ選択肢を選んでいた参加者 2 名は課題を適切に行っていないと判断し除外した。また、一般に普及しているリフレッシュレートが 60 Hz のモニターでは、1 フレームあたりの平均時間は基本的には 16.7 ms 前後になる。そのため、約 2 倍の 32 ms 以上になっている試行は、適切な呈示時間で刺激が呈示されていないと判断し除外した。この基準で参加者単位では 8 名が除外された。18 歳未満の参加者はい

なかった。分析には、140名(女性25名、男性113名、無回答2名、平均年齢45.4歳、標準偏差10.2歳)、1367個のマイクロ実験のデータが用いられた。

## サンプルサイズ及びマイクロ実験設計

サンプルサイズは、日本かつオンラインでFFDE研究を行った鈴木他(2022; <https://osf.io/uwqfs>)を参考に人数を決定した。従って、目標サンプルサイズは100名であった。ただし、欠損値や除外があると仮定し、目標サンプルサイズを満たせるよう150名を募集した。

参加者は一人当たり、10ブロック(すなわち10マイクロ実験)を行った。オンライン上(クラウドソーシング)で実施する関係から、実験実施時間(約20-30分)を考慮し一人あたりのブロック数を決定した。実験1全体でのマイクロ実験数はサンプルサイズ×1人あたりのマイクロ実験数で算出される。すなわち、本実験全体での目標マイクロ実験は1000-1500個であった。なお、本実験の目標マイクロ実験数は統計的な根拠に基づき決定されたものではない。

## 実験計画

本実験は、ラジカルランダムマイゼーション及びメタスタディを用いた研究デザインである(詳細はBaribault et al., 2018も参照)。一般的な実験計画法における要因(ないし独立変数)に相当するものはファセットと呼ばれる。本実験のファセットと水準をTable 1に示した。ファセットとその水準はオンライン実験で実施可能であることを前提に選定した。顔刺激の特徴に関するファセットは、顔刺激を作成する際に用いたAIで指定可能なパラメータの一部である。これらの

ファセットの水準も同様に AI で設定可能なものを用いた<sup>7</sup>。顔刺激 1 枚あたりの呈示時間と顔刺激の呈示回数については、鈴木他 (2022) における FFDE が生じるまでの潜時と、FFDE を知覚した累積時間を参考に決定した。鈴木他 (2022) は、本研究と同様に顔全体を刺激として呈示した場合、潜時がおおよそ 3500-4000 ms、累積時間もおおよそ 3500-4000 ms であったと報告している。

これをもとに、一連の顔刺激の呈示時間 (顔刺激 1 枚あたりの呈示時間と顔刺激の呈示回数の乗算で算出) がこの範囲よりも広くなるよう設定した。すなわち、本研究における最も短い一連の顔刺激の呈示時間を 3000 ms (150 ms × 20 回) と、鈴木他 (2022) の潜時よりも短くなるよう設定した。また、最も長い一連の顔刺激の呈示時間を 18000 ms (300 ms × 60 回) と、鈴木他 (2022) の潜時 + 累積時間を大きく見積もった場合の 8000 ms (潜時 4000 ms + 累積時間 4000 ms) よりも長くなるよう設定した。顔刺激 1 枚あたりの呈示時間の水準の間隔については、オンライン実験で呈示可能な時間精度を考慮し 50 ms 刻みとした (Bridges, Pitiot, MacAskill, & Peirce, 2020)。

従属変数は、FFDE 刺激を見た後に行う一連の顔についての不気味さの評定値と顔の歪みの大きさの評定値であった。Balas & Pearson (2019) 及び鈴木他 (2022) と同様に、不気味さ・顔の歪みの大きさは、いずれも「全く○○ではなかった」から「とても○○だった」の 7 件法のリッカート尺度を用いて回答した。

---

<sup>7</sup> なお、「顔刺激の向き」に関しては、顔画像作成時の制約から、顔向きをカテゴリカルに指定した。具体的な角度 (e.g., 右向きに 5°) を指定することはできなかったため、全ての刺激で顔向きの角度が厳密に揃っているわけではない。同様に、「顔刺激の年齢」についても、大まかな区分であり、具体的な年齢を指定しているわけではない。

Table 1  
 実験1のファセット一覧とその水準

ファセット	水準	鈴木他 (2022)
顔刺激の性別	女性, 男性	女性
顔刺激の年齢	成人, 子供	成人
顔刺激の向き	左向き, 正面, 右向き	正面
顔刺激の表情	喜び, 中立	中立
顔刺激の回転角度	0° ~ 350° (10°刻み)	0°
顔刺激1枚あたりの呈示時間	150 ms ~ 350 ms (50 ms 刻み)	250 ms
一連の顔刺激の呈示回数	20回 ~ 60回 (1回刻み)	40回

注) 回転角度は右回りで定義された。例えば, 90° のとき顔刺激は髪の毛 (頭頂) が右側にある状態である。

## 刺激と装置

参加者は各自のパーソナルコンピュータからオンラインで実験に参加した。刺激の呈示及び実験の制御は jsPsych 7.3 (de Leeuw, 2015) 及び jspsych-psychophysics プラグイン (Kuroki, 2021) を用いた。画面の大きさや視距離は統制しなかった。

顔刺激は鈴木他 (2022) と同様に AI (<https://generated.photos/>) で作成した 1440 枚を用いた。顔画像は 24 通りの組み合わせがあり、それぞれ 60 枚用意した (24 通り × 60 枚 = 1440 枚)。顔画像の組み合わせは、顔刺激の性別 (男性・女性, 2 水準) × 顔刺激の年齢 (成人・子供, 2 水準) × 顔刺激の向き (左向き・正面・右向き, 3 水準) × 顔刺激の表情 (喜び・中立, 2 水準) である。なお、AI での作成時、上記の項目以外のパラメータはデフォルトで作成した (e.g., Eye color)。また、背景は透明 (透過) とし、眼鏡をかけておらず、かつ髭が濃くないものを使用した。

画面の背景色は灰色で、その中央に黒色の注視点が呈示された。顔刺激は注視点の左右に注視点と重なることなく呈示された。なお、本実験はオンラインで実施したため、表示される刺激の大きさは参加者のデバイスや観察距離に依存して異なる<sup>8</sup>。顔画像は一辺が 256 ピクセルであり、1.5 倍に拡大され呈示された。注視点と左右それぞれの顔画像との距離は 360 ピクセルであった。顔画像が呈示されている間も注視点は呈示され続けた。また、顔刺激は時間的な間隔を置かず連続的に呈示された。

---

<sup>8</sup> なお、顔刺激の大きさは FFDE の強さに影響しないことが報告されている (Balas & Pearson, 2019)。

## 顔画像の呈示順番決定方法

使用する画像は各ブロック (すなわちマイクロ実験) 開始前に決定された条件 (e.g., 男性, 子供, 右向き, 喜び) の 60 枚のうち, 呈示回数分使用された。ここでは顔刺激の呈示回数が 41 回だった場合の例を記述したが, 呈示回数が他の回数や偶数であった場合も同様である。

まず, 各試行開始時に該当ブロックの条件 (e.g., 男性, 子供, 右向き, 喜び) の 60 枚の顔画像をランダムな順に並べ替える。次に, その 60 枚のうち前半の「呈示回数 + (呈示回数 % 2)」枚を用いる (=この場合 42 枚)。なお, 「呈示回数 % 2」は呈示回数を 2 で割った余りのことであり, 呈示回数が偶数なら 0 に, 奇数であれば 1 となる。続いて, 前段階の 42 枚を半分 (この場合 21 枚ずつ) に分割し, 前半半分 (21 枚) と後半「 $(42 / 2) - (\text{呈示回数} \% 2)$ 」枚 (=20 枚) に (プログラム内部的に) 分ける。そして, 注視点の左側に呈示される顔画像は前段階で選んだもの (この場合計 41 枚) を前半・後半の順で呈示し, 右側に呈示される顔画像は前段階で分けた前半と後半を入れ替えたもの (後半・前半の順, この場合計 41 枚) を呈示した。

## 手続き

参加者は実験内容の説明文を読んだ後, インフォームドコンセントを取得し実験を行った。まず, 自認する性別と年齢を回答すると, ブラウザはフルスクリーンに移行した。その後, 参加者は手続きについての教示を読んだ上で練習試行を 3 試行行った。この際, 画面中心の注視点を見ることが教示された。また, 練習試行・本番試行ともに, 注視点を見るようにするため, 注視点呈示される前に毎試行マウスのカーソルを画面の端に移動させるよう教示した。練習試行の刺激の呈示条件は, 顔刺激が「女性, 成人, 正面, 中立表情, 正立 (回転角度 0°)」であ

り、顔刺激 1 枚あたりの呈示時間は 250 ms、一連の顔刺激の呈示回数は 40 回であった。この条件は、鈴木他 (2022) で用いられたものを参考に決定した。なお、条件以外の刺激の呈示方法は本番試行と同一であった。参加者は、FFDE のデモを見た後に 7 件法で不気味さの評定と 2 択で顔が歪んだかを回答した。このとき、3 試行のうち 1 度以上顔が歪んだと回答した参加者のみ本番試行へと移り、3 試行とも歪まなかったと回答した参加者はこの時点で実験を終了した。

その後、参加者は本番試行を行った (Figure 1)。まず、各ブロック (マイクロ実験) 開始時に全てのファセットの水準の値をそれぞれ一様分布からランダムサンプリングし決定した。ここでは、仮に 1 ブロック目における各ファセットの水準が「男性、成人、右向き、中立表情、30°、顔刺激 1 枚あたりの呈示時間が 300 ms、一連の顔刺激の呈示回数が 52 回」になったとして説明する。水準の決定は各ブロックの開始前のみ行われ、ブロック内では一貫して同じ水準が用いられた。すなわち、今回の例では 1 ブロック目における全ての試行は、各ファセットの水準が「男性、成人、右向き、中立表情、30°、顔刺激 1 枚あたりの呈示時間が 300 ms、一連の顔刺激の呈示回数が 52 回」である。その後、ブロック内の各試行が始まる前に、前述の顔刺激の呈示順番決定方法に従い、呈示する顔刺激とその順番が決定された。試行が始まると、まず灰色の背景の画面中央に黒い注視点が 1000 ms 呈示された。その後、顔刺激が注視点の左右にブロック開始時に決定された回数 (この例では 52 回)、呈示された。この間も画面中央には注視点が呈示され続けた。顔刺激の呈示が終わると、1000 ms 注視点のみが呈示され、その後参加者は一連の顔刺激についての印象をリッカート尺度で回答した。項目は「どの程度、不気味に感じました

か？」について、「全く不気味ではなかった」から「とても不気味だった」の7件法<sup>9</sup>で回答し、「どの程度、顔が物理的に歪みましたか？」についても同様に「全く歪まなかった」から「とても歪んだ」の7件法<sup>10</sup>で回答した。これを1ブロックあたり5試行繰り返した。1ブロック目が終了すると、2ブロック目が開始される前に、再び全ファセットの水準をランダムに決定し、その条件で2ブロック目を行った<sup>11</sup>。これを繰り返し、参加者は全10ブロックを行った。実験時間はおよそ20-30分であった。

---

<sup>9</sup> 全く不気味ではなかった - 不気味ではなかった - あまり不気味ではなかった - どちらでもない - やや不気味だった - 不気味だった - とても不気味だった, の7件法

<sup>10</sup> 全く歪まなかった - 歪まなかった - あまり歪まなかった - どちらでもない - やや歪んだ - 歪んだ - とても歪んだ, の7件法

<sup>11</sup> なお、水準の決定の際のランダムサンプリングは復元抽出で行われた。そのため、全く同一の条件が複数ブロックで連続することもありうる。

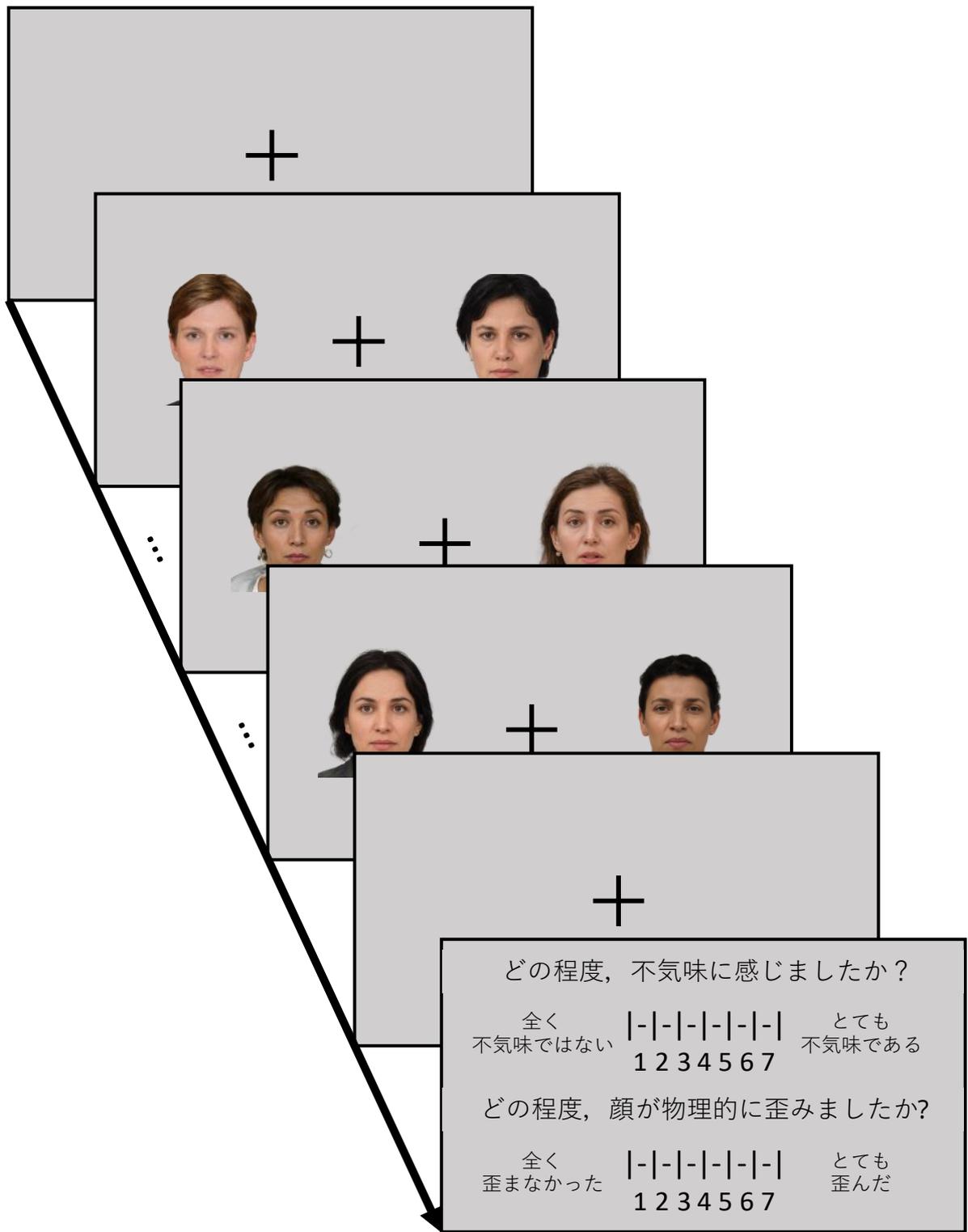


Figure 1. 実験 1 の本番試行における試行の流れ (ただし、不気味さ・歪みの大きさの評定画面については、図としての可読性を優先し実際の画面とは多少異なる)。

## 結果

本研究の分析は後続の実験含め、いずれも Jupyter Notebook 6.4.6 (Kluyver et al., 2016) と R 4.1.0 (R Core Team, 2021) を用いた。分析には、*tidyverse* 1.3.1 パッケージ (Wickham et al., 2019) を、Hedges' *g* 及びその信頼区間の算出は *effectsize* 0.6.0.1 パッケージ (Ben-Shachar, Lüdtke, & Makowski, 2020) を用いた。一般化イータ 2 乗及びその信頼区間の算出は、*anovakun* 4.8.7 (井関, 2022) を用いた。信頼区間の算出方法はブートストラップ法を使用した。相関係数とその信頼区間の算出は、R の組み込み関数である *cor.test* 関数を用いた。練習試行 3 試行の回答は事前登録に従い、分析には使用しなかった。これは後続の実験でも同様であった。

参加者の不気味さ・歪みの大きさの評定値は、「全く○○ではなかった」を 1, 「とても○○だった」を 7 として各段階 1 刻みで数値化した。ファセットの効果を検討するため、これらの評定値をマイクロ実験単位で算出した後、各ファセットの水準ごとに平均値を算出した。評定値は数値が大きいほど、FFDE による一連の顔刺激の不気味さ・歪みの大きさを強く知覚していたことを表す。

実験 1 の分析は Baribault et al. (2018) におけるグローバル検定に相当する。不気味さ・歪みの大きさの評定値ごとに、ファセットを一つずつ基本独立変数として扱い分析することで、それぞれのファセットが FFDE の強さに及ぼす影響を検討した。本実験は探索的研究であるため、有意性などの判断は行わなかった。ファセット間での 1 次の交互作用については OSF にあるサプリメントファイルを参照されたい (<https://osf.io/845g3>)。2 次以上の交互作用については、その組み合わせが多いこと、解釈可能性、各セルに対応するマイクロ実験の数が少なくなることを

考慮し分析しなかった。なお、実験1のデータはOSF上 (<https://osf.io/h5eyq>)で公開されているため、第三者による二次分析も可能である。

## 不気味さについて

マイクロ実験単位での、不気味さ評定値の分布を Figure 2 に示した。箱ひげ図より、評定値が3後半から5中盤に多くのマイクロ実験が集中していることが読み取れる。次にファセットごとの不気味さ評定値の平均値 ( $M$ )と標準偏差 ( $SD$ )、ファセットの効果量を算出した。顔刺激の性別について、女性条件の評定値は  $M = 4.62$ ,  $SD = 1.44$ , 男性条件は  $M = 4.41$ ,  $SD = 1.42$  であり、条件間の差は Hedges'  $g = 0.14$ , 95%CI [0.03, 0.25]であった (Figure 3)。顔刺激の年齢について、成人条件では  $M = 4.54$ ,  $SD = 1.44$ , 子供条件では  $M = 4.49$ ,  $SD = 1.42$  であり、条件間の差は Hedges'  $g = 0.03$ , 95%CI [-0.08, 0.14]であった (Figure 4)。顔刺激の向きについて、左向き条件では  $M = 4.43$ ,  $SD = 1.59$ , 正面条件では  $M = 4.49$ ,  $SD = 1.60$ , 右向き条件では  $M = 4.60$ ,  $SD = 1.55$  であった (Figure 5)。顔向きファセットの効果量は、 $\eta^2 = .002$ , 95%CI [.00, .01] であった。顔刺激の表情について、喜び条件では  $M = 4.46$ ,  $SD = 1.46$ , 中立表情条件では  $M = 4.58$ ,  $SD = 1.41$  であり、条件間の差は Hedges'  $g = -0.08$ , 95%CI [-0.19, 0.03]であった (Figure 6)。顔刺激の回転角度について、不気味さの評定値との相関係数は  $r = .03$ , 95%CI [-.02, .09]であった。なお、このファセットは水準数が多いため、詳細は Figure 7 を参照されたい。

顔刺激1枚あたりの呈示時間については、150 ms 条件では  $M = 3.93$ ,  $SD = 1.43$ , 200 ms 条件では  $M = 4.22$ ,  $SD = 1.47$ , 250 ms 条件では  $M = 4.63$ ,  $SD = 1.37$ , 300 ms 条件では  $M = 4.95$ ,  $SD = 1.19$ , 350 ms 条件では  $M = 4.89$ ,  $SD = 1.39$  であった (Figure 8)。不気味さ評定値との相関は  $r$

= .26, 95%CI [.21, .31]であった。不気味さの評定値と呈示回数との相関は  $r = .13$ , 95%CI [.08, .18]

であった。このファセットも水準数が多いため、詳細は Figure 9 を参照されたい。

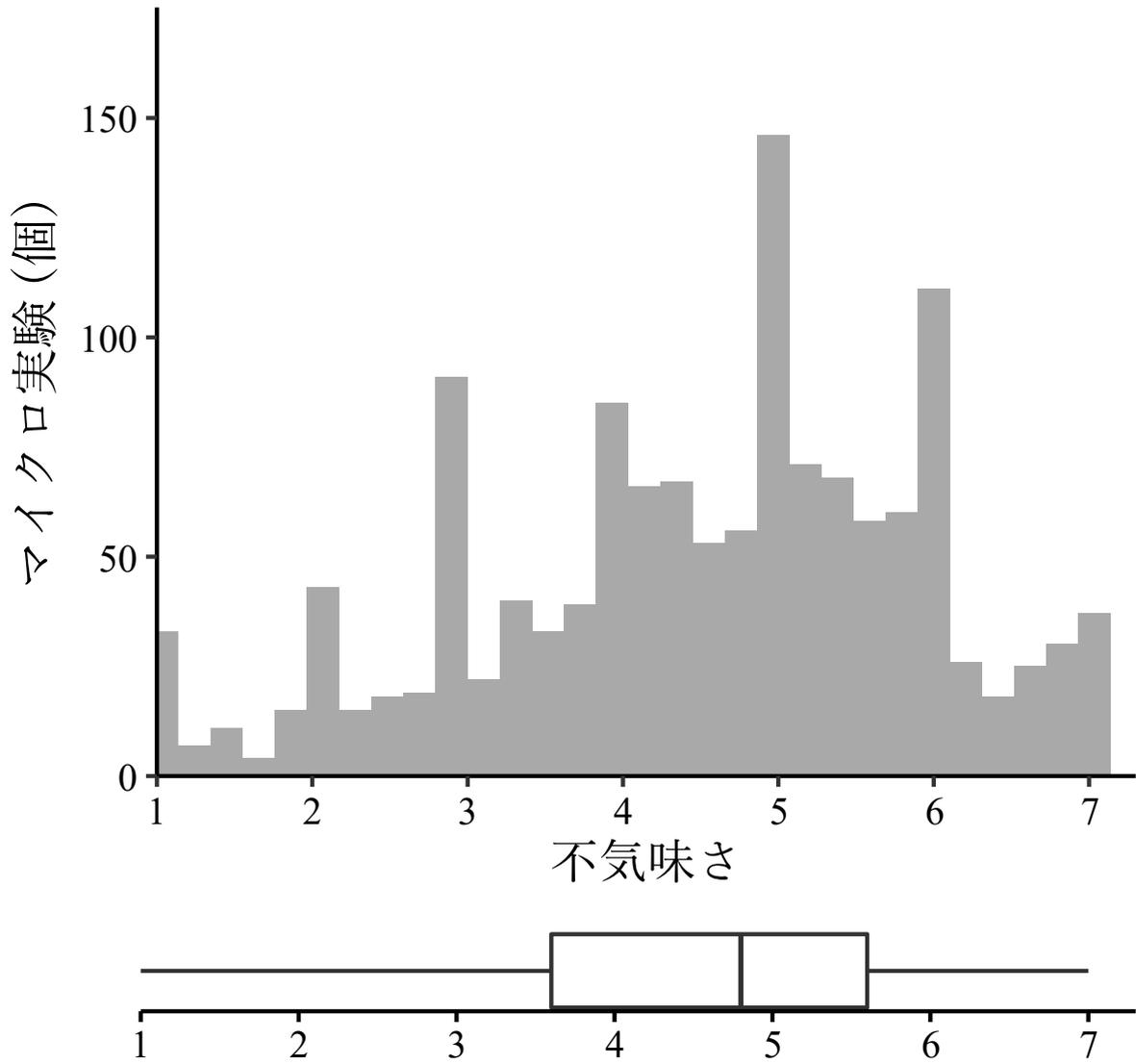


Figure 2. 実験 1 における不気味の評定値ごとのマイクロ実験の分布と箱ひげ図。

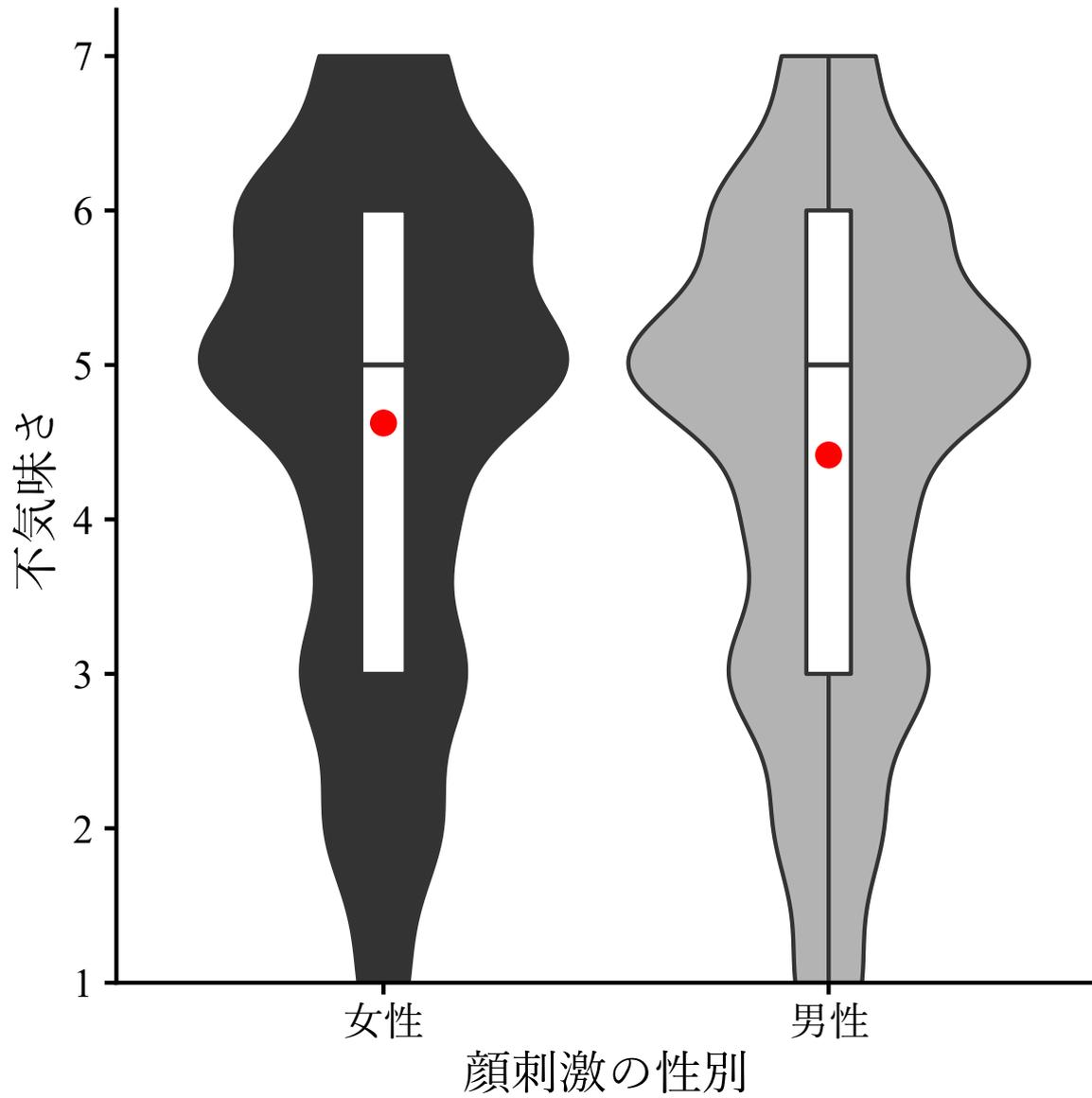


Figure 3. 実験 1 における顔刺激の性別ごとの不気味さ評定値のバイオリンプロットと箱ひげ図。

赤い丸は平均値を表す。

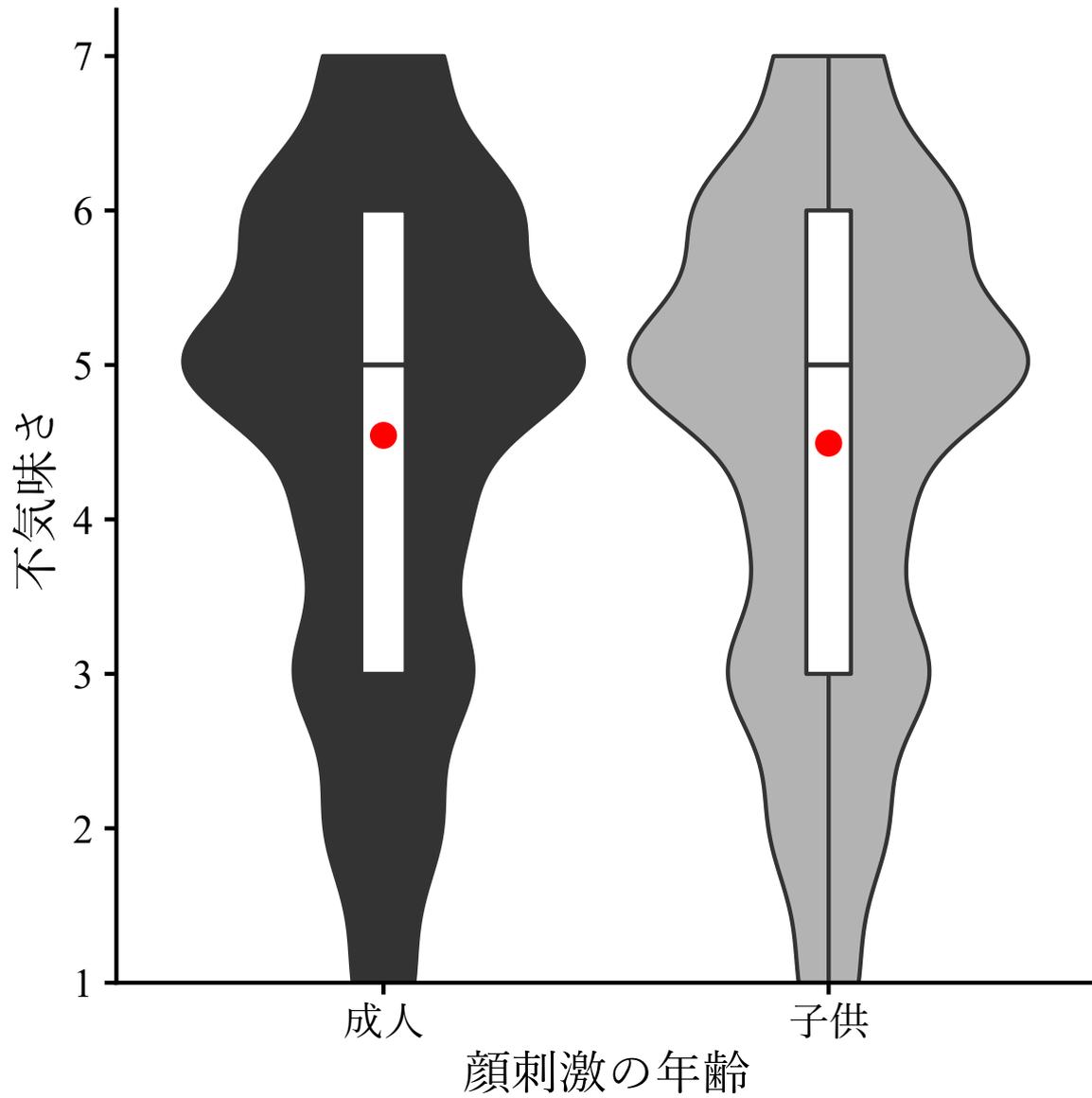


Figure 4. 実験 1 における顔刺激の年齢ごとの不気味さ評定値のバイオリンプロットと箱ひげ図。

赤い丸は平均値を表す。

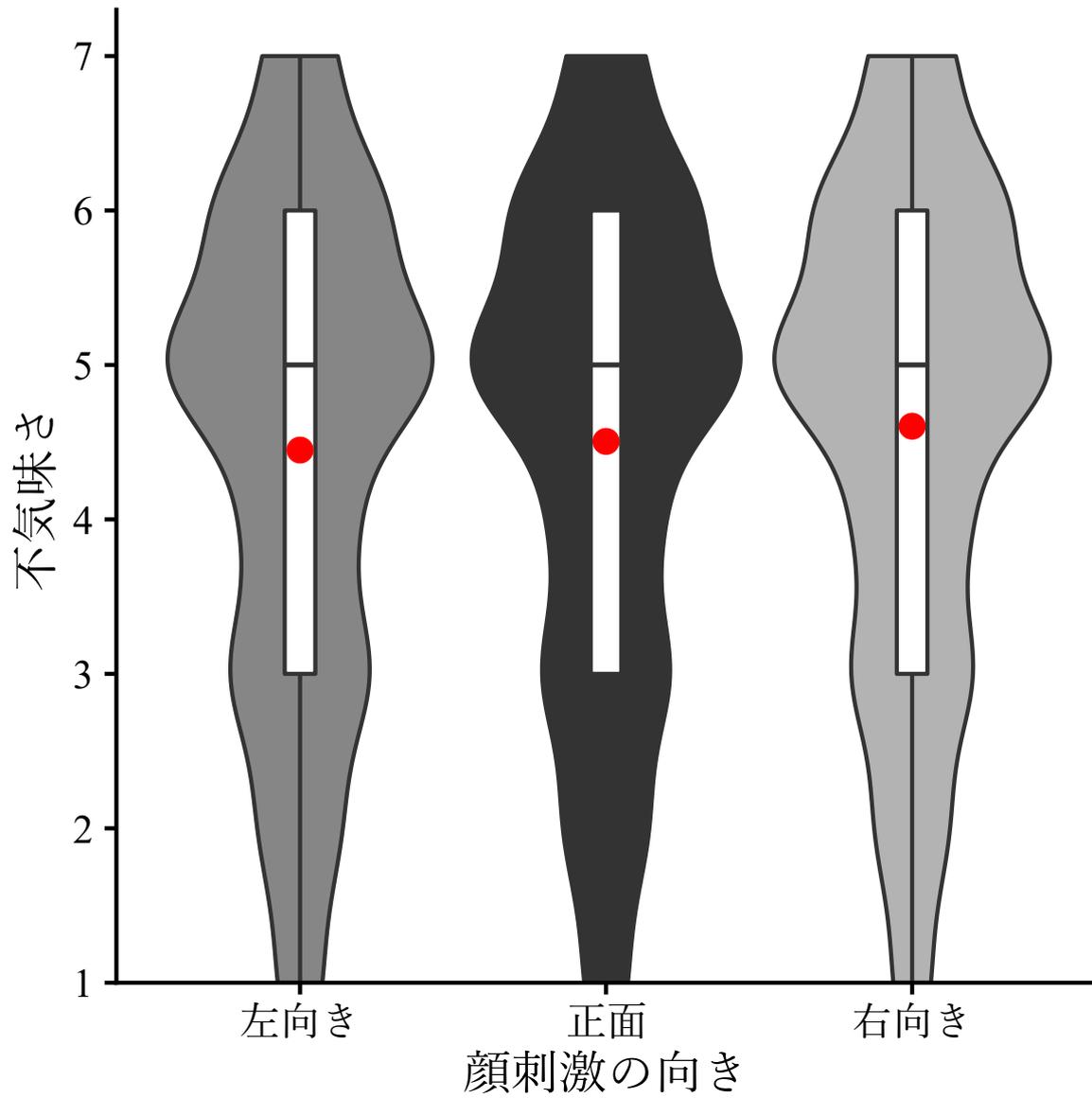


Figure 5. 実験 1 における顔刺激の向きごとの不気味さ評定値のバイオリンプロットと箱ひげ図。

赤い丸は平均値を表す。

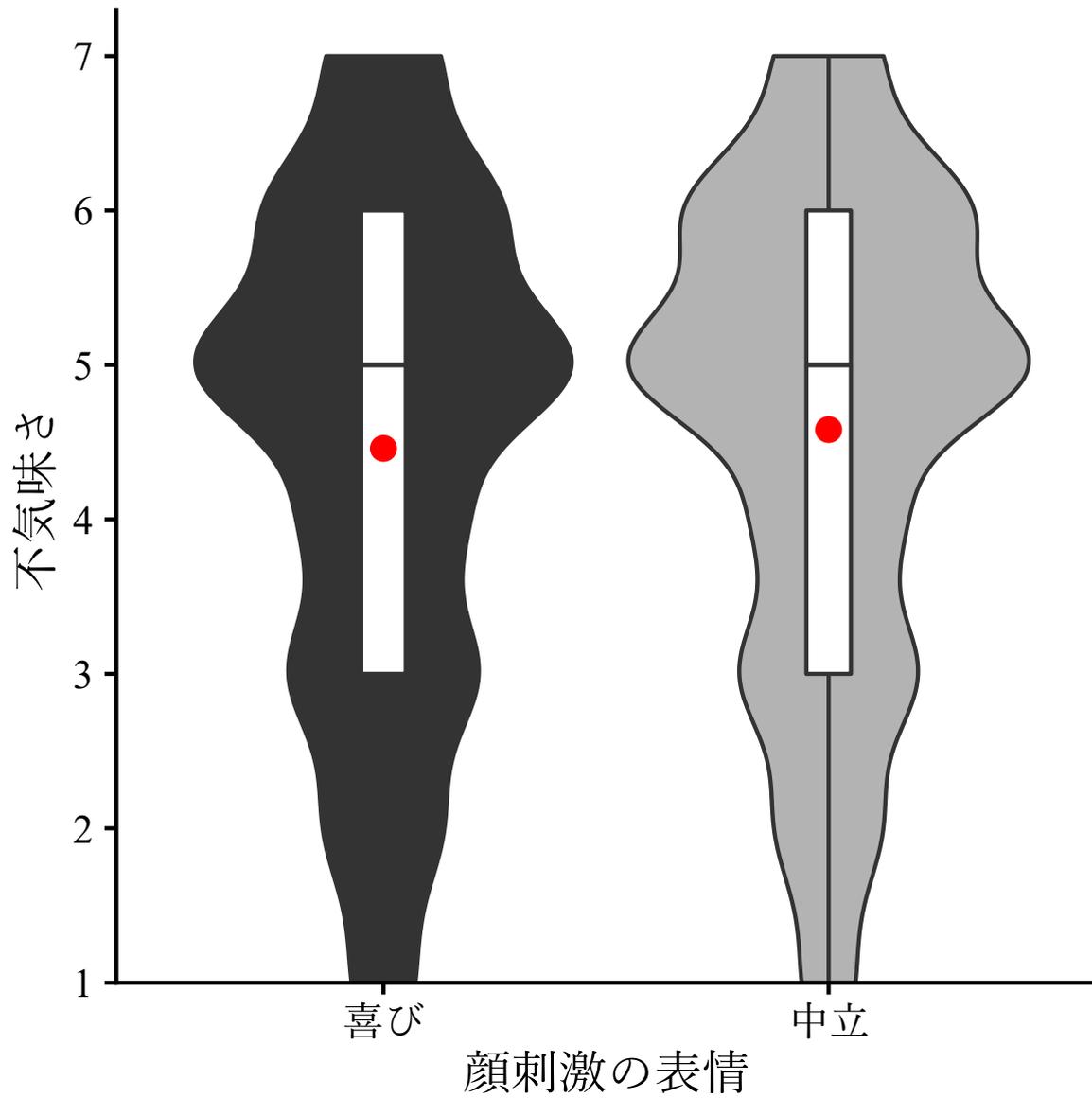


Figure 6. 実験 1 における顔刺激の表情ごとの不気味さ評定値のバイオリンプロットと箱ひげ図。

赤い丸は平均値を表す。

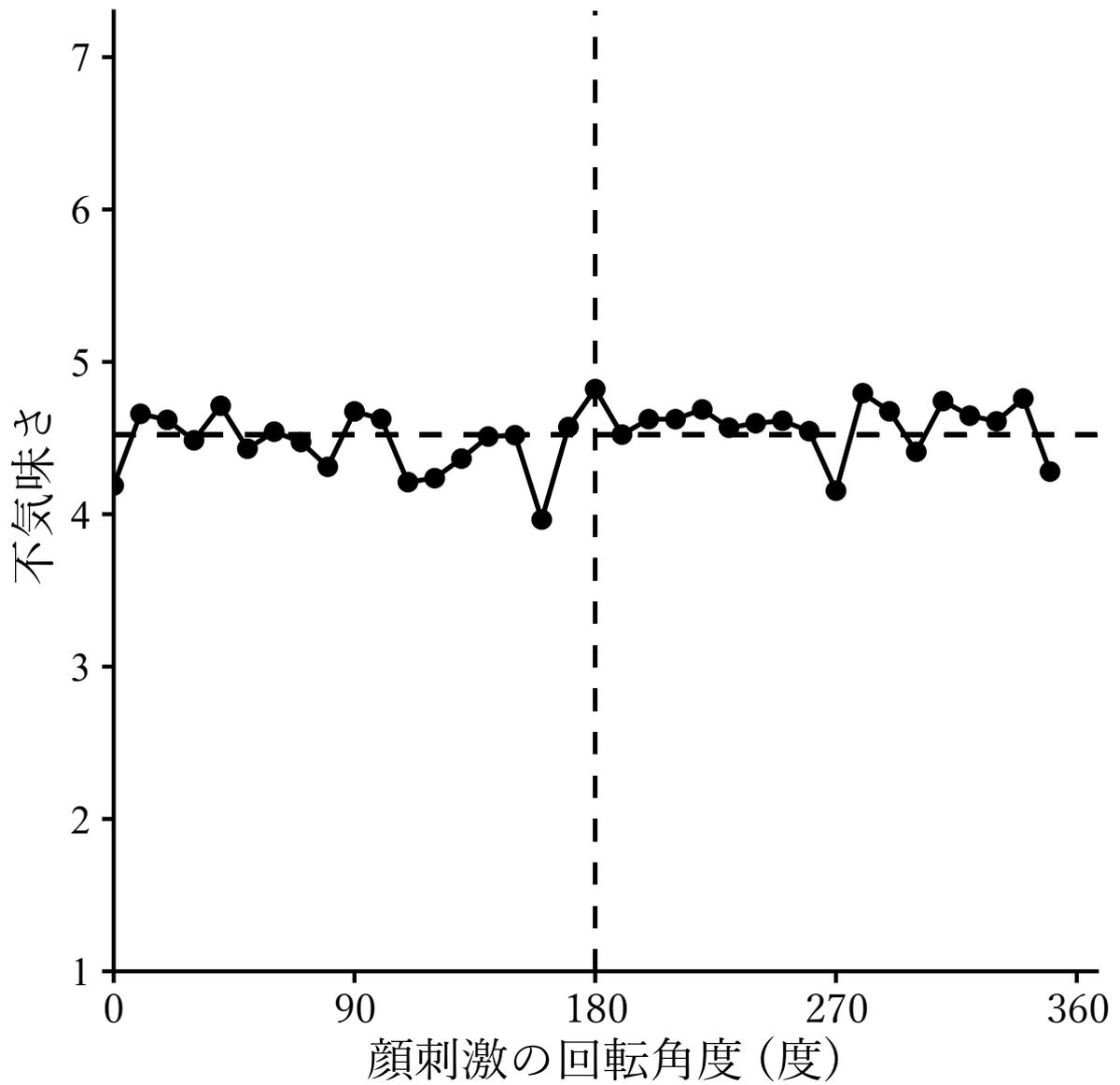


Figure 7. 実験 1 における顔刺激の回転角度と不気味さ評定値の折れ線グラフ。水平方向の破線は不気味さ評定値の平均値を表し、垂直方向の破線は顔刺激の回転角度が 180°であることを示す。

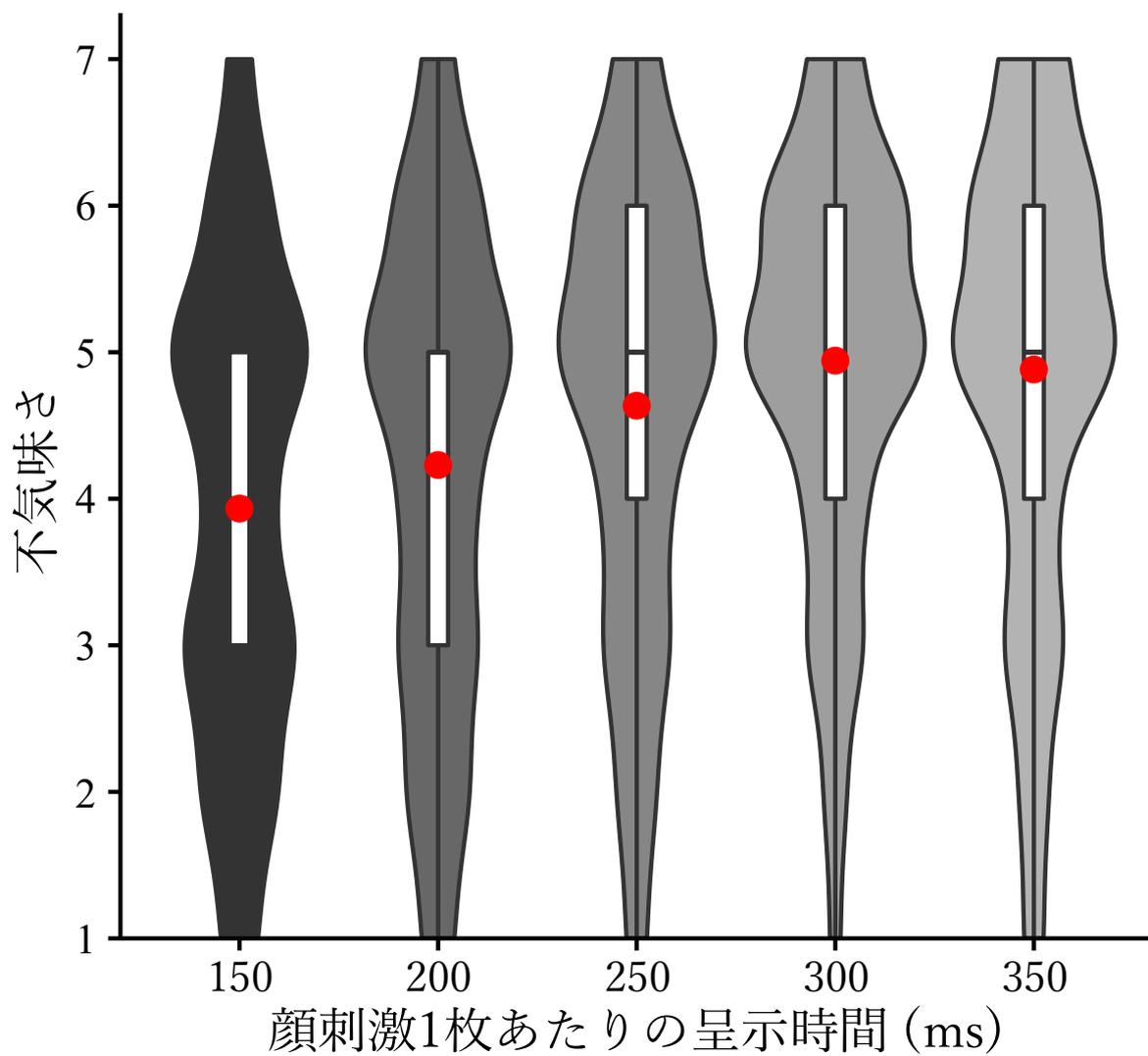


Figure 8. 実験 1 における顔刺激 1 枚あたりの呈示時間ごとの不気味さ評定値のバイオリンプロット

と箱ひげ図。赤い丸は平均値を表す。

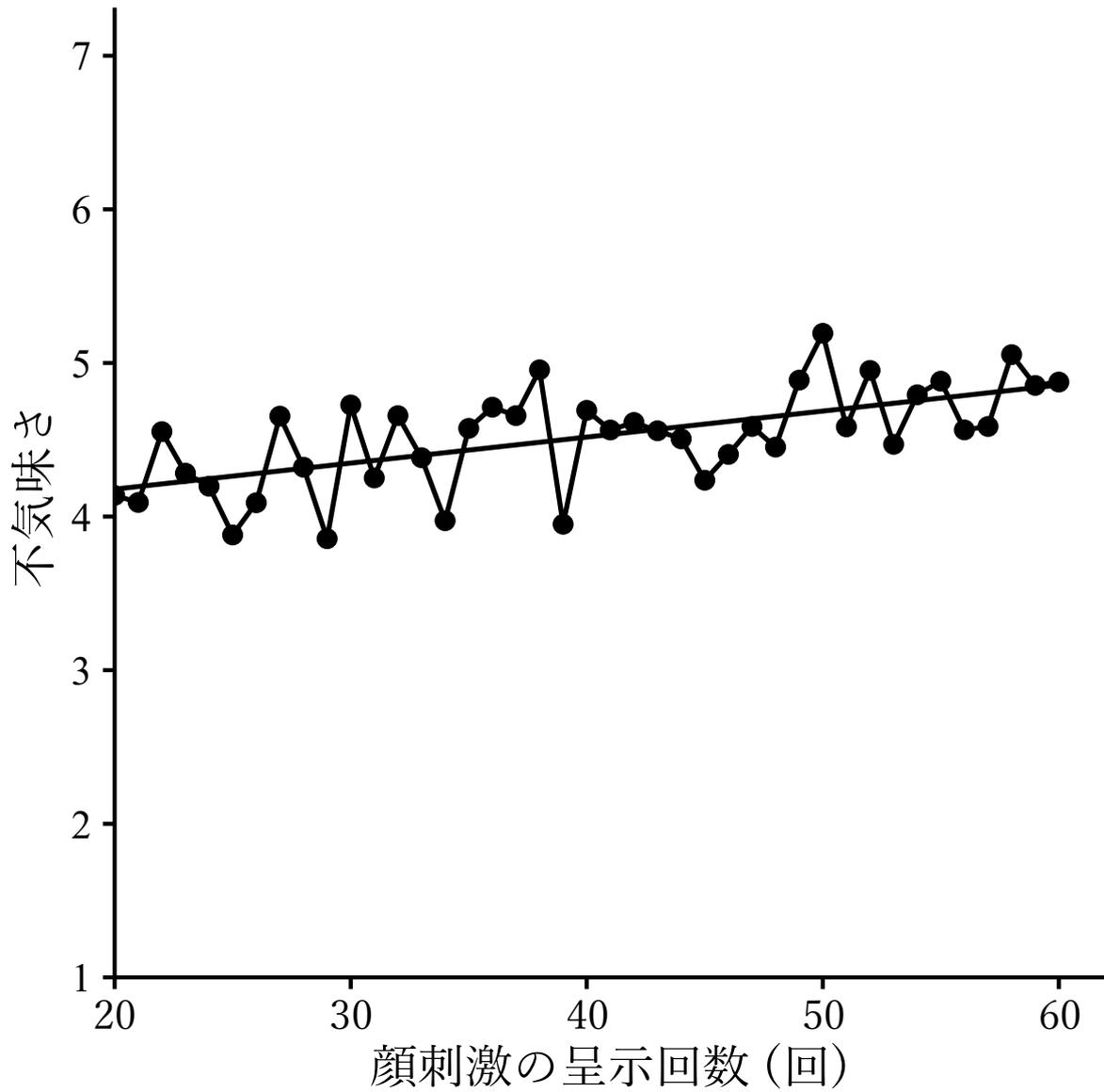


Figure 9. 実験 1 における顔刺激の呈示回数と不気味さ評定値の折れ線グラフ。実線は回帰直線を表す。

## 歪みの大きさについて

マイクロ実験単位での、歪みの大きさの評定値の分布を Figure 10 に示した。箱ひげ図より、評定値が 4 前半から 5 後半に多くのマイクロ実験が集中していることが読み取れる。次にファセットごとの歪みの大きさの評定値の平均値 ( $M$ ) と標準偏差 ( $SD$ )、ファセットの効果量を算出した。顔刺激の性別については、女性条件の評定値は  $M = 4.93$ ,  $SD = 1.29$ , 男性条件は  $M = 4.74$ ,  $SD = 1.30$  であり、条件間での差は Hedges'  $g = 0.15$ , 95%CI [0.04, 0.25]であった (Figure 11)。顔刺激の年齢について、成人条件では  $M = 4.87$ ,  $SD = 1.33$ , 子供条件では  $M = 4.80$ ,  $SD = 1.26$  であり、条件間での差は Hedges'  $g = 0.05$ , 95%CI [-0.05, 0.16]であった (Figure 12)。顔刺激の向きについて、左向き条件では  $M = 4.76$ ,  $SD = 1.49$ , 正面条件では  $M = 4.80$ ,  $SD = 1.49$ , 右向き条件では  $M = 4.91$ ,  $SD = 1.45$  であった (Figure 13)。顔向きファセットの効果量は、 $\eta^2_G = .002$ , 95%CI [.00, .01] であった。顔刺激の表情について、喜び条件では  $M = 4.80$ ,  $SD = 1.31$ , 中立表情条件では  $M = 4.86$ ,  $SD = 1.28$  であり、条件間での差は Hedges'  $g = -0.04$ , 95%CI [-0.15, 0.06] であった (Figure 14)。顔刺激の回転角度について、歪みの大きさの評定値との相関係数は  $r = -.02$ , 95%CI [-.07, .03] であった。なお、このファセットは水準数が多いため、詳細は Figure 15 を参照されたい。

顔刺激 1 枚あたりの呈示時間については、150 ms 条件では  $M = 4.31$ ,  $SD = 1.38$ , 200 ms 条件では  $M = 4.56$ ,  $SD = 1.33$ , 250 ms 条件では  $M = 4.92$ ,  $SD = 1.22$ , 300 ms 条件では  $M = 5.22$ ,  $SD = 1.04$ , 350 ms 条件では  $M = 5.17$ ,  $SD = 1.22$  であった (Figure 16)。歪みの大きさ評定値との相関は  $r = .26$ , 95%CI [.21, .31] であった。歪みの大きさ評定値と呈示回数の相関は  $r = .18$ , 95%CI [.13, .23] であった。このファセットも水準数が多いため、詳細は Figure 17 を参照されたい。

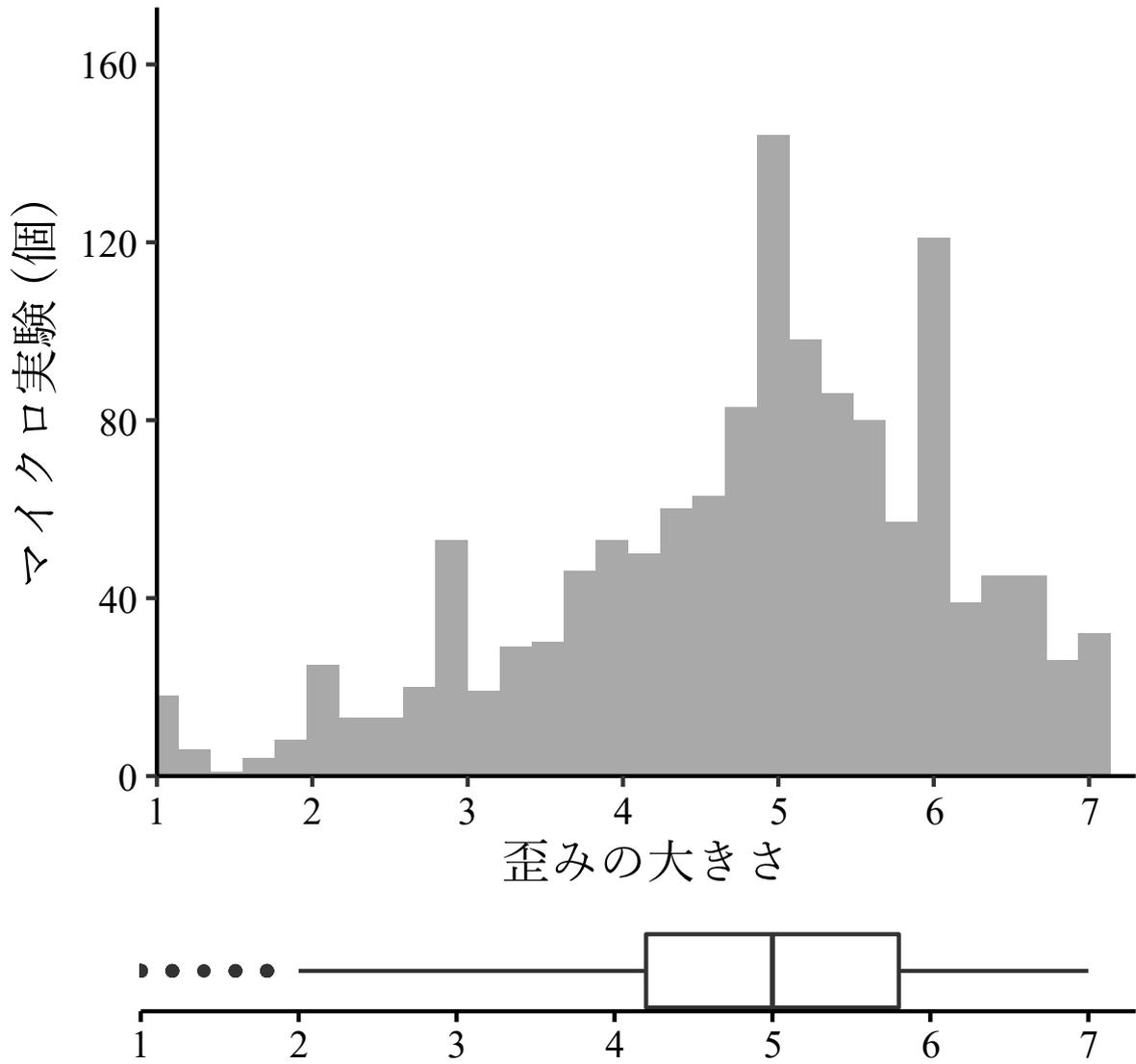


Figure 10. 実験 1 における歪みの大きさの評定値ごとのマイクロ実験の分布と箱ひげ図。

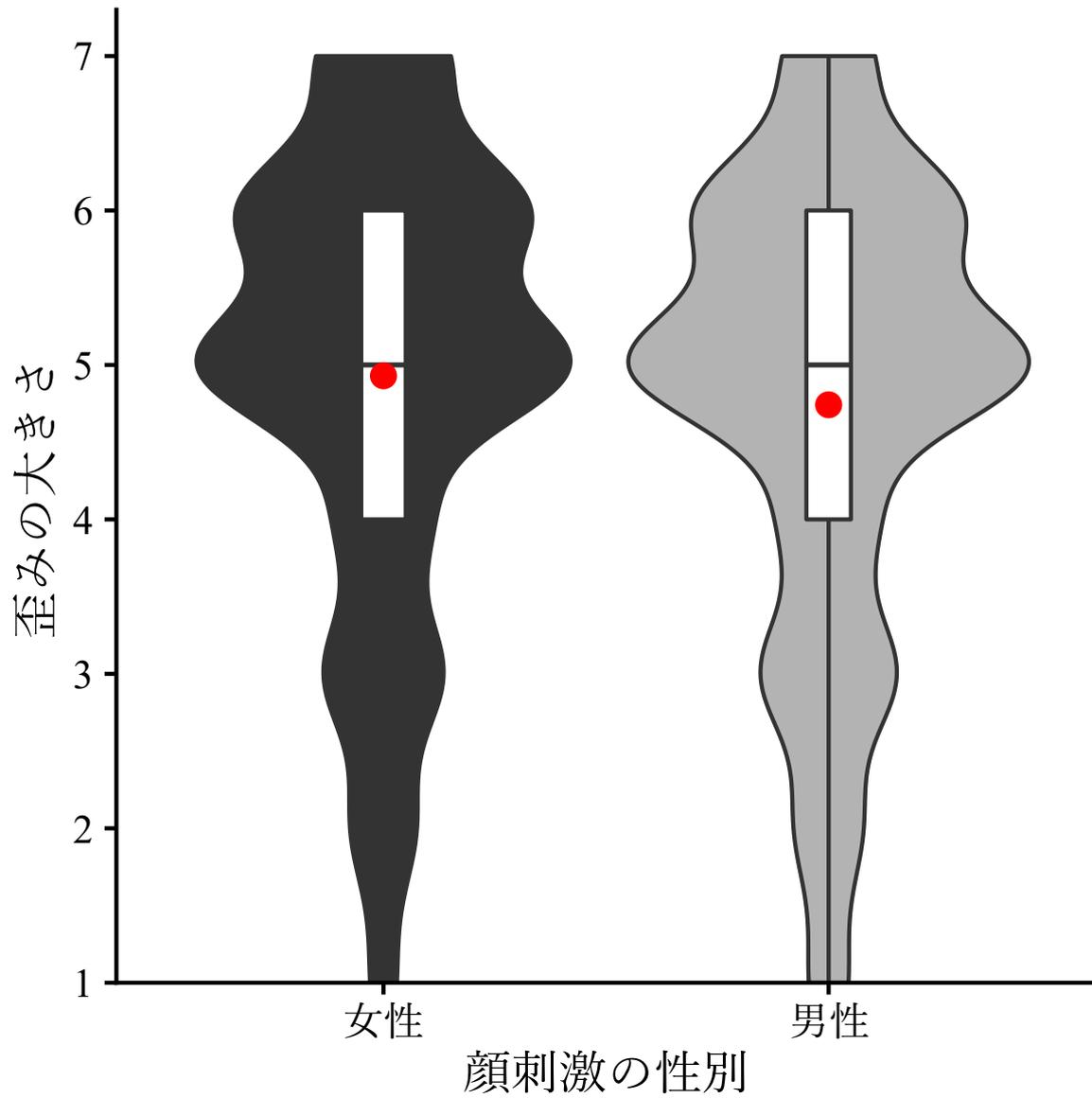


Figure 11. 実験 1 における顔刺激の性別ごとの歪みの大きさ評定値のバイオリンプロットと箱ひげ図。赤い丸は平均値を表す。

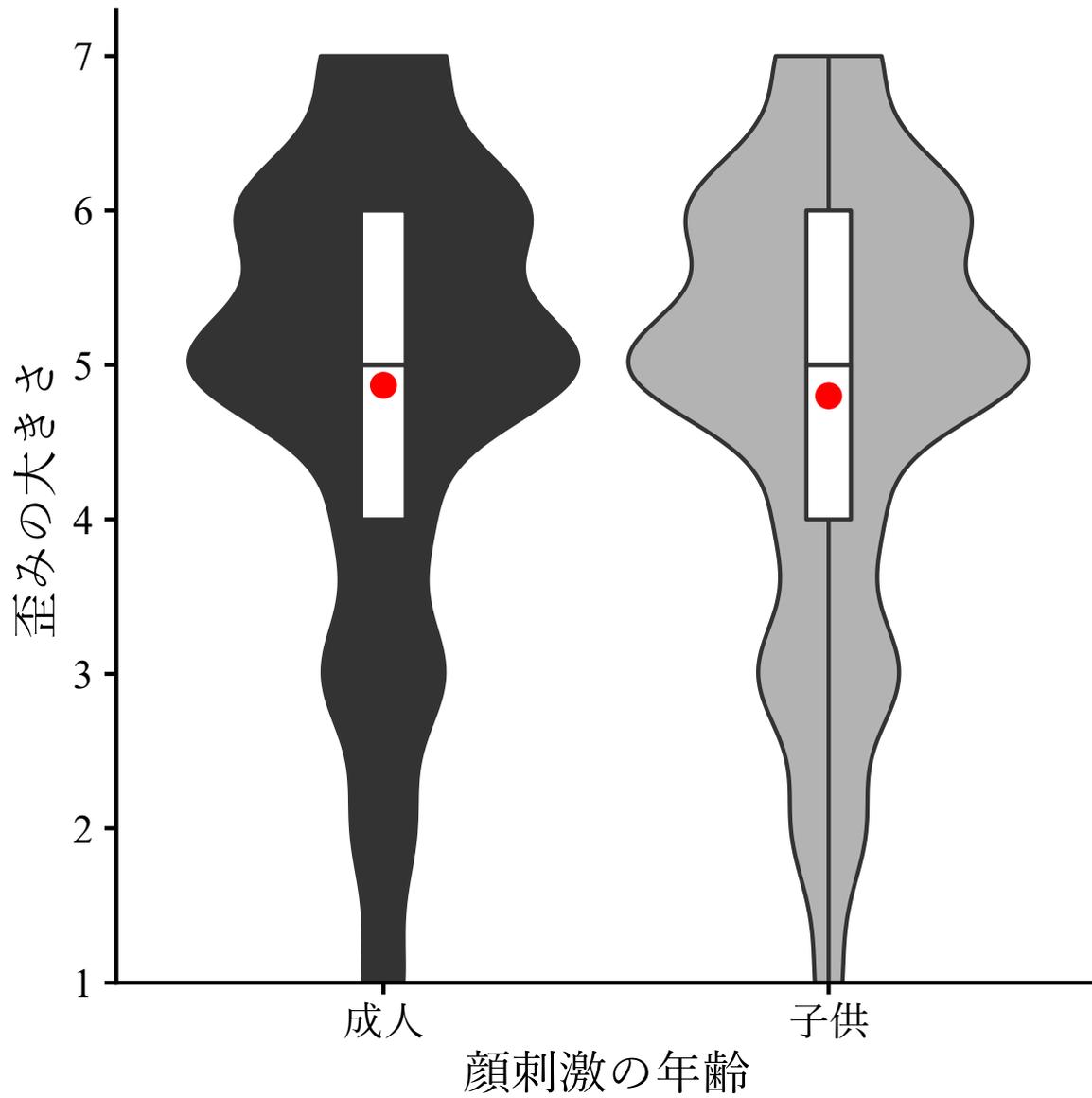


Figure 12. 実験 1 における顔刺激の年齢ごとの歪みの大きさ評定値のバイオリンプロットと箱ひげ図。赤い丸は平均値を表す。

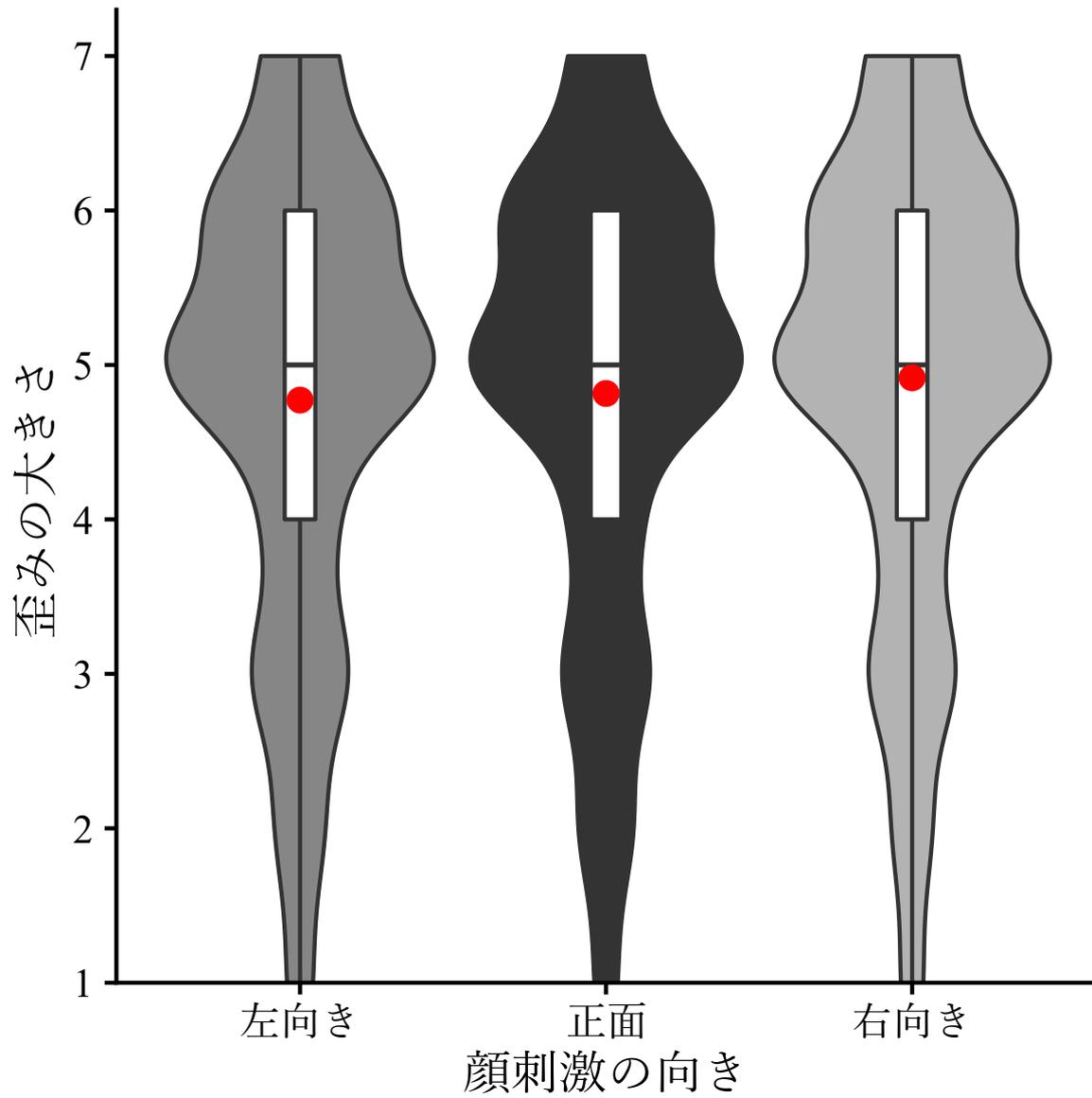


Figure 13. 実験 1 における顔刺激の向きごとの歪みの大きさ評定値のバイオリンプロットと箱ひげ図。赤い丸は平均値を表す。

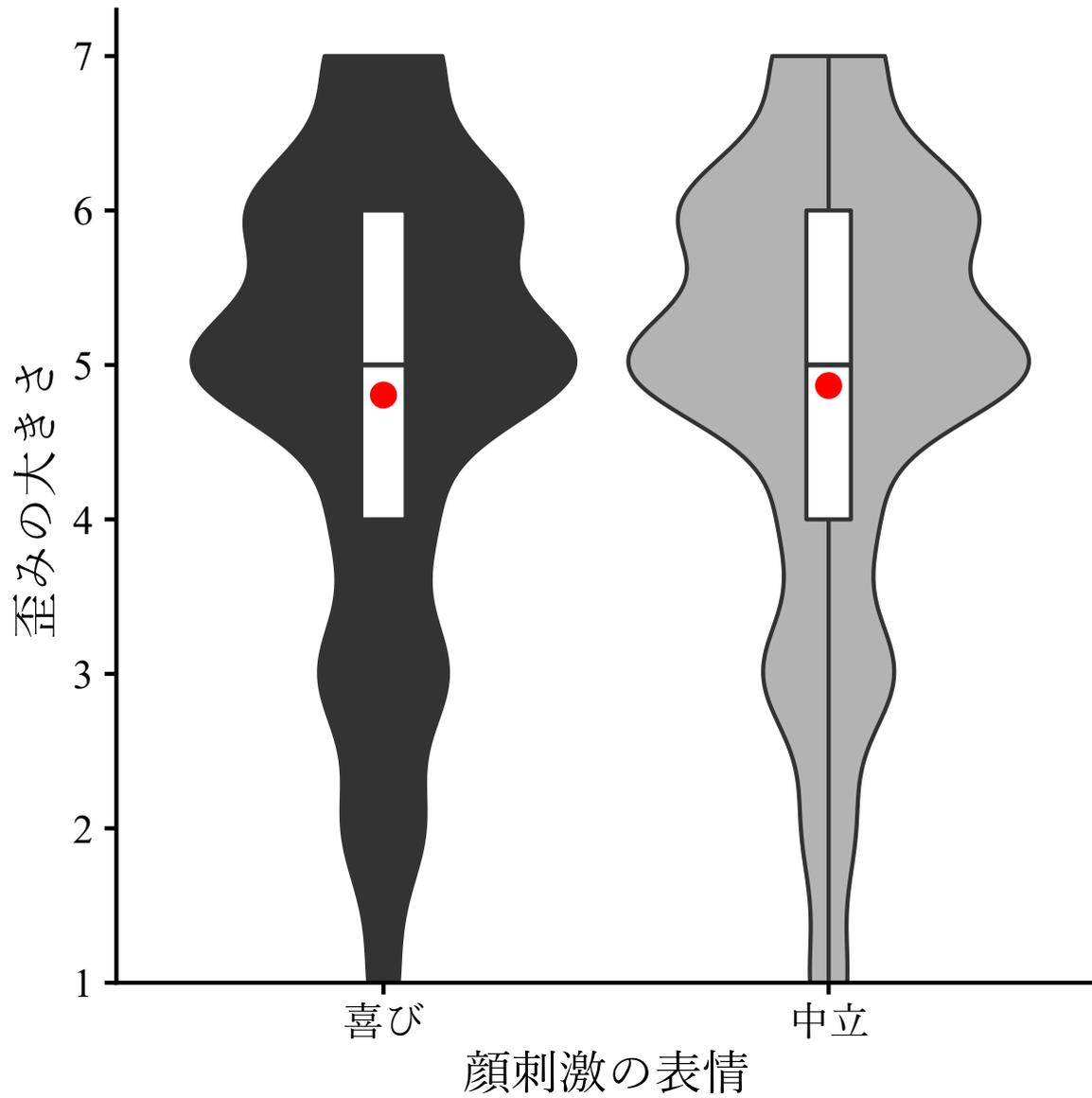


Figure 14. 実験 1 における顔刺激の表情ごとの歪みの大きさ評定値のバイオリンプロットと箱ひげ図。赤い丸は平均値を表す。

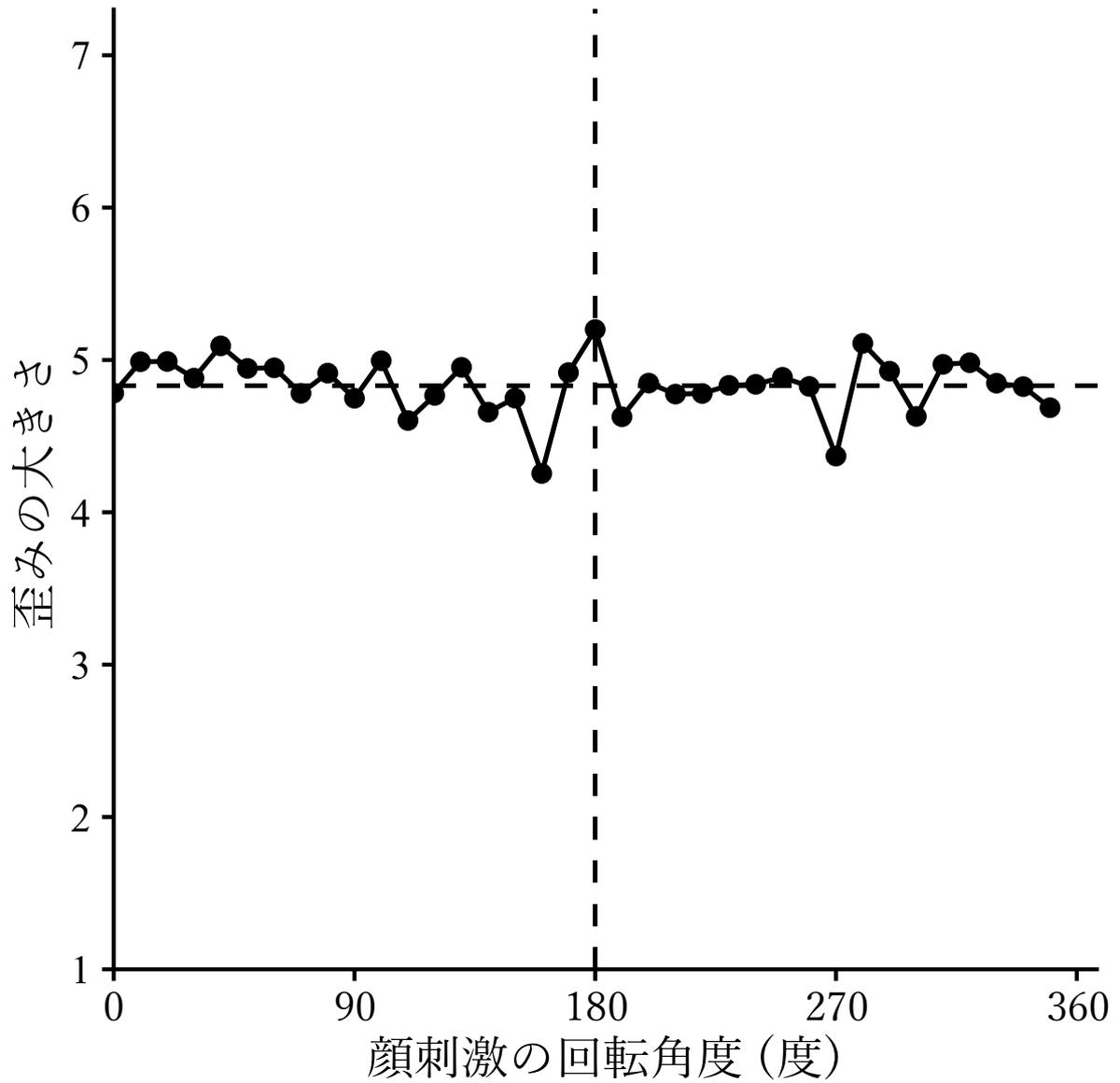


Figure 15. 実験 1 における顔刺激の回転角度と歪みの大きさ評定値の折れ線グラフ。水平方向の破線は歪みの大きさ評定値の平均値を表し、垂直方向の破線は顔刺激の回転角度が 180°であることを示す。

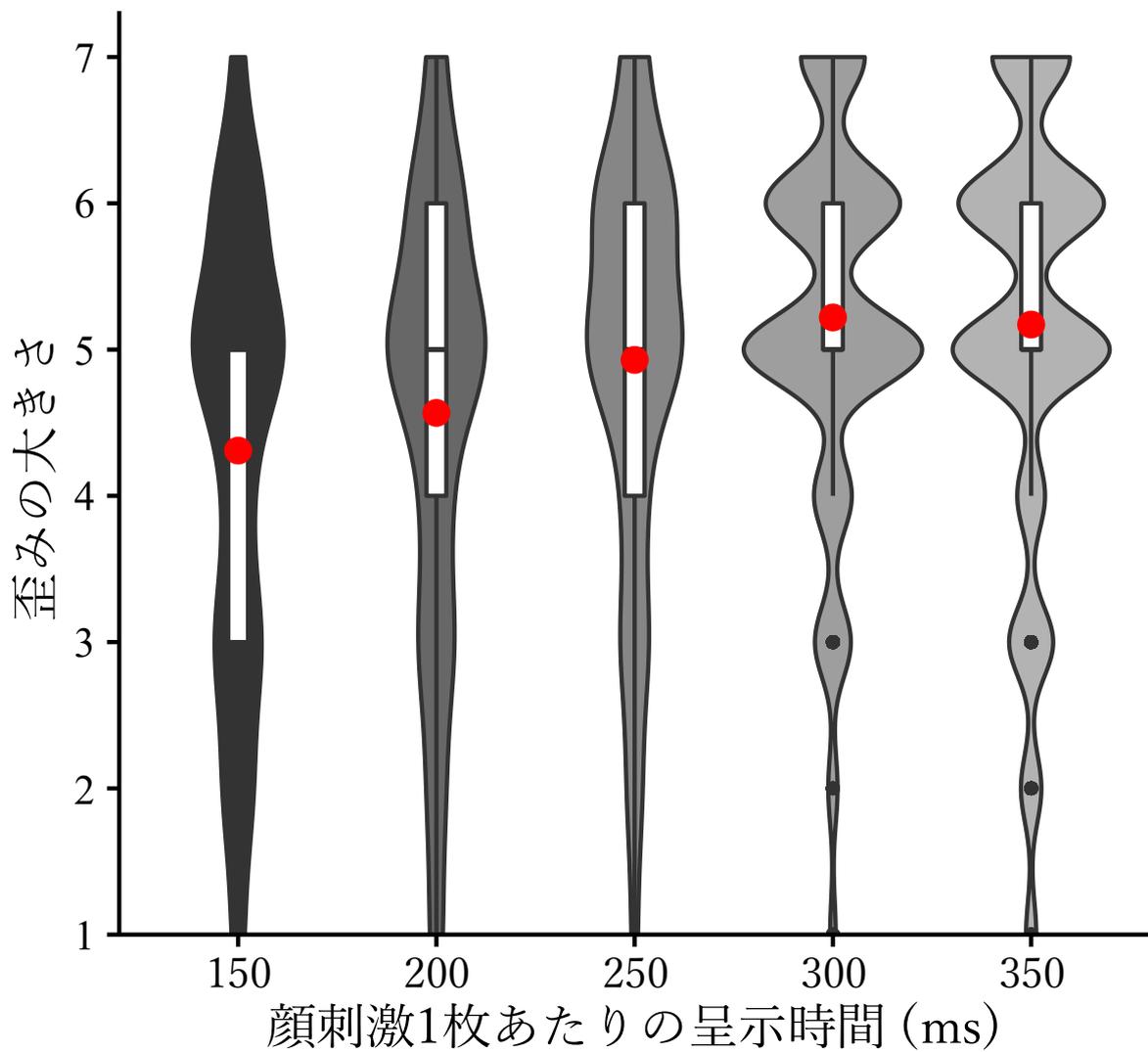


Figure 16. 実験 1 における顔刺激 1 枚あたりの呈示時間ごとの歪みの大きさ評定値のバイオリン

プロットと箱ひげ図。赤い丸は平均値を表す。

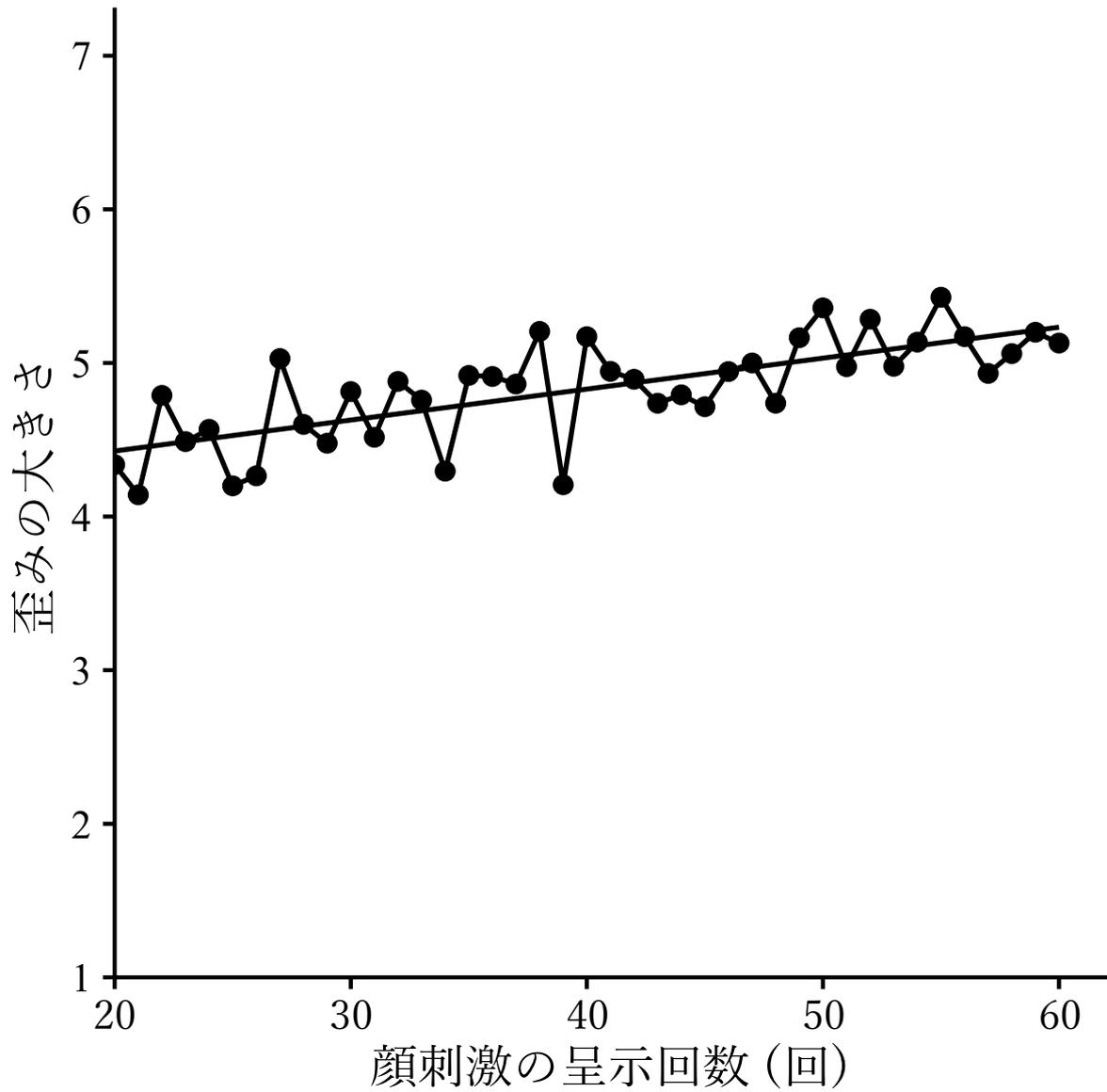


Figure 17. 実験 1 における顔刺激の呈示回数と歪みの大きさ評定値の折れ線グラフ。実線は回帰直線を表す。

## 考察

実験1の目的は、ラジカルランダムマイゼーションを用いて、様々なファセットとFFDEの関係を探ることであった。その結果、算出された効果量及びバイオリンプロットによるデータの分布を確認すると、FFDEの強さ(不気味さ・歪みの大きさ)は顔向きなどの顔刺激の特徴によって大きく変化したとは言えなかった。一方、顔刺激の呈示方法(e.g., 顔刺激1枚あたりの呈示時間)は、比較的影響が大きかった。これらの結果から、FFDEは顔刺激の特徴に対して頑健である可能性が示唆された。また、不気味さの結果と歪みの大きさの結果は概ね同様の傾向であった。

Utz & Carbon (2015)は、顔刺激が倒立しているとき正立時に比べ有意にFFDEが弱まることを報告している。一方で、本研究の結果は顔の倒立がFFDEに影響するとは言えないと報告しているTangen et al. (2011)及びBalas & Pearson (2019)を支持するものであった。また、顔刺激1枚あたりの呈示時間が短いほどFFDEが弱まったことは、Balas & Pearson (2019)の結果と一致する。これらについては総合考察にて検討する。顔刺激の呈示回数が少ないほどFFDEが弱まったことを合わせると、この錯視の生起には低次の視覚処理が重要である可能性が考えられる。

しかし、本実験は探索的研究であり、この結果から確証的な判断をすることは適切ではないと考えられる。そこで、続く実験2では、本実験で効果が比較的大きかった顔刺激1枚あたりの呈示時間と顔刺激の呈示回数がFFDEに与える影響を改めて検討した。実験2においても、これらの要因について本実験と同様の結果が確認されれば、ラジカルランダムマイゼーションを用いた探索的研究が有効である可能性が示唆されると考えられる。

## 実験 2

### 目的と仮説

実験 1 の結果、顔刺激 1 枚あたりの呈示時間が短いときは長いときに比べ FFDE が弱まる可能性が示唆された。また、一連の顔刺激呈示回数についても呈示回数が少ないときは多いときより FFDE が弱まる可能性が示唆された。しかし、実験 1 は探索的な研究であり、この結果から確証的な判断をすることは適切ではない。そこで、実験 2 ではこれら 2 つの結果について、同様の結果が得られるか実験計画法を用い改めて検討した。

本実験で検証する仮説は以下の通りである。実験 1 の結果に基づくと、顔刺激 1 枚あたりの呈示時間が短い条件 (150 ms) は長い条件 (300 ms) より、FFDE が弱まると考えられる。すなわち、不気味さや歪みの評定値が、短い条件のほうが長い条件より小さくなると予測される。また、一連の顔刺激の呈示回数については、それが少ない条件 (20 回) は多い条件 (60 回) より、FFDE が弱まると考えられる。すなわち、不気味さや歪みの評定値が少ない条件のほうが多い条件より小さくなると予測される。これら 2 要因の交互作用については、実験 1 ではほとんど見られなかったため、有意ではないと予測した。本実験は、OSF に行った事前登録に基づいて実施された (<https://osf.io/x25v6>)。

## 方法

### 参加者

実験2にはインフォームドコンセントを取得した上で、99名が参加した。しかしこの内、前半の50名は実験プログラムのミスで条件設定が適切に行えていなかった(この条件設定が適切ではなかった参加者の結果についてはサプリメンタリを参照)。そのため、実験プログラムを修正し再度50名の参加者を募集した(なお、参加したのは49名であった)。参加者の募集はYahoo!クラウドソーシング(<https://crowdsourcing.yahoo.co.jp/>)を用い、オンライン上で実験を行った。参加者には報酬として100ポイント(日本円で100円相当)が支払われた。なお、この金額は実験時間5分あたり20ポイントとして計算した(実験2は実施時間がおよそ25分<sup>12</sup>とみなし報酬額を算出した)。

事前登録した除外基準に基づき、全回答のうち90%以上で同じ選択肢を選んでいた参加者2名は課題を適切に行っていないと判断し除外した。また、1フレームあたりの平均時間が32ms以上になっている試行を除外した。この基準で2名が除外された。18歳未満の参加者はいなかった。分析には45名(女性6名、男性39名、平均年齢44.8歳、標準偏差10.6歳)のデータが用いられた。

---

<sup>12</sup> 事前登録では実験時間の目安を20分と表記していたが、これは25分の誤りである。なお、この変更は仮説や結果の解釈には影響しない。

## サンプルサイズ設計

顔刺激 1 枚あたりの呈示時間及び、一連の顔刺激の呈示回数それぞれの主効果の効果量を中程度 ( $f = 0.25$ ) と見積もり、検定力分析の結果からサンプルサイズを決定した。G\*Power 3.1.9.7 (Faul, Erdfelder, Lang, & Buchner, 2007) を使用し、効果量  $f = 0.25$ 、有意水準  $\alpha = .05$ 、検定力  $1 - \beta = .80$  とし、反復測定間の相関と球面性修正項はそれぞれ G\*Power のデフォルト値である 0.5 と 1 を使用した。その結果、必要サンプルサイズは 34 名であった。欠損値や除外があると仮定し、目標サンプルサイズを満たすため 50 名を募集した (本研究の参加者募集にはクラウドソーシングを用いた。そのため募集人数は多少前後する可能性が考えられる。50 名を超えた場合、除外されなかった参加者のデータは全て分析に用いることとした)。また、分析対象 (除外後) の参加者数が 34 名に達しない場合は、有効な参加者が 34 名に達するまで参加者を募集することとした。仮に、除外がなく 50 名全員のデータが分析に用いられる場合は、効果量  $f = 0.25$ 、有意水準  $\alpha = .05$  の場合、検定力  $1 - \beta$  は約 .93 である。

## 実験計画と分析方法

顔刺激 1 枚あたりの呈示時間 (150 ms · 300 ms, 2 水準) 及び一連の顔刺激の呈示回数 (20 回 · 60 回, 2 水準) をそれぞれ独立変数とする 2 要因参加者内計画であった。従属変数は、FFDE 刺激を見た後に行う、その一連の顔についての評定値であった。不気味さと歪みの大きさについては実験 1 と同様に、7 件法リッカート尺度で回答した。そして、2 つの従属変数 (不気味さ、歪みの大きさの評定値) ごとに、対応のある 2 要因の分散分析を行った。

## 刺激と装置

参加者は各自のパーソナルコンピュータからオンラインで実験に参加した。刺激の呈示や実験の制御は jsPsych 7.3 (de Leeuw, 2015) 及び jspsych-psychophysics プラグイン (Kuroki, 2021) を用いた。画面の大きさや視距離は統制しなかった。

顔刺激は、実験 1 で用いたもののうち、「女性、成人、正面、中立表情」の 60 枚を用いた。また、顔刺激は練習試行と本番試行で共通であった。これら以外の刺激や顔刺激の呈示順番決定アルゴリズムは実験 1 と同様であった。

## 手続き

練習試行や本番試行への参加条件、教示は実験 1 と同様であった。本番試行は、全 60 試行で構成された (顔刺激 1 枚あたりの呈示時間 2 水準  $\times$  一連の顔刺激の呈示回数 2 水準  $\times$  繰り返し回数 15 回)。これらはランダムな順番で実施された。まず灰色の背景の画面中央に黒い注視点が 1000 ms 呈示された。その後、顔刺激 (女性、成人、正面、中立表情、正立) が注視点の左右に 20 回または 60 回呈示された。この間も画面中心には注視点が呈示され続けた。顔刺激の呈示が終わると、1000 ms 注視点のみが呈示され、その後参加者は一連の顔刺激についての印象をリッカート尺度で回答した。回答方法は実験 1 と同様であった。実験時間はおよそ 15-25 分であった。

## 結果

参加者ごとかつ条件ごとに不気味さ、歪みの大きさの評定値の平均値を算出した。評定値が大きいほど、一連の顔刺激の不気味さ・歪みの大きさを強く知覚していたことを表す。これらの分析の際、除外は事前に登録した基準 (<https://osf.io/x25v6>) に基づいて行われた。ただし、1フレームあたりの平均時間が 32 ms 以上になっている試行を除外した際、1名の参加者の 150 ms・60回条件の全てのデータが除外された。対応のある分析を行うため、この参加者の全条件のデータを分析対象から除外した。この除外は事前登録からの逸脱である。その後、不気味さ・歪みの大きさごとに、*anovakun* 4.8.7 (井関, 2022) を用いて対応のある 2 要因の分散分析を行った。分析環境については、*effectsize* パッケージと *cor.test* 関数を使用しなかったこと以外は実験 1 と同一であった。

### 不気味さについて

条件ごとの不気味さの評定値の平均値 ( $M$ ) と標準偏差 ( $SD$ ) は、150 ms・20回条件では  $M = 3.94$ ,  $SD = 1.21$ , 150 ms・60回条件では  $M = 4.83$ ,  $SD = 1.11$ , 300 ms・20回条件では  $M = 4.53$ ,  $SD = 1.18$ , 300 ms・60回条件では  $M = 5.30$ ,  $SD = 1.19$  であった (Figure 18)。顔刺激 1 枚あたりの呈示時間と一連の顔刺激の呈示回数が FFDE によって生じる不気味さに与える影響を検討するため、それぞれを要因とする対応のある 2 要因分散分析を行った。球面性の仮定が成立しているかを検討する Mendoza の多標本球面性検定は、2 要因とも参加者内要因かつ 2 水準であったため行われなかった。分散分析の結果、顔刺激 1 枚あたりの呈示時間の主効果及び、一連の顔刺激の呈示回数的主効果が有意であった (それぞれ,  $F(1, 43) = 31.5$ ,  $p < .001$ ,  $\eta^2 = .05$ , 95% CI

[.02, .10],  $MSe = 0.39$ ;  $F(1, 43) = 65.8$ ,  $p < .001$ ,  $\eta_G^2 = .11$ , 95%CI [.05, .18],  $MSe = 0.45$ )。これらの2  
要因の交互作用は有意ではなかった( $F(1, 43) = 1.84$ ,  $p = .18$ ,  $\eta_G^2 < .001$ , 95%CI [.000, .004],  $MSe =$   
0.09)。従って、顔刺激1枚あたりの呈示時間が150 msの条件は300 msの条件より不気味さの評  
定値が有意に低いことが示された。また、一連の顔刺激の呈示回数が20回の条件は60回の条件  
より評定値が有意に低いことが示された。

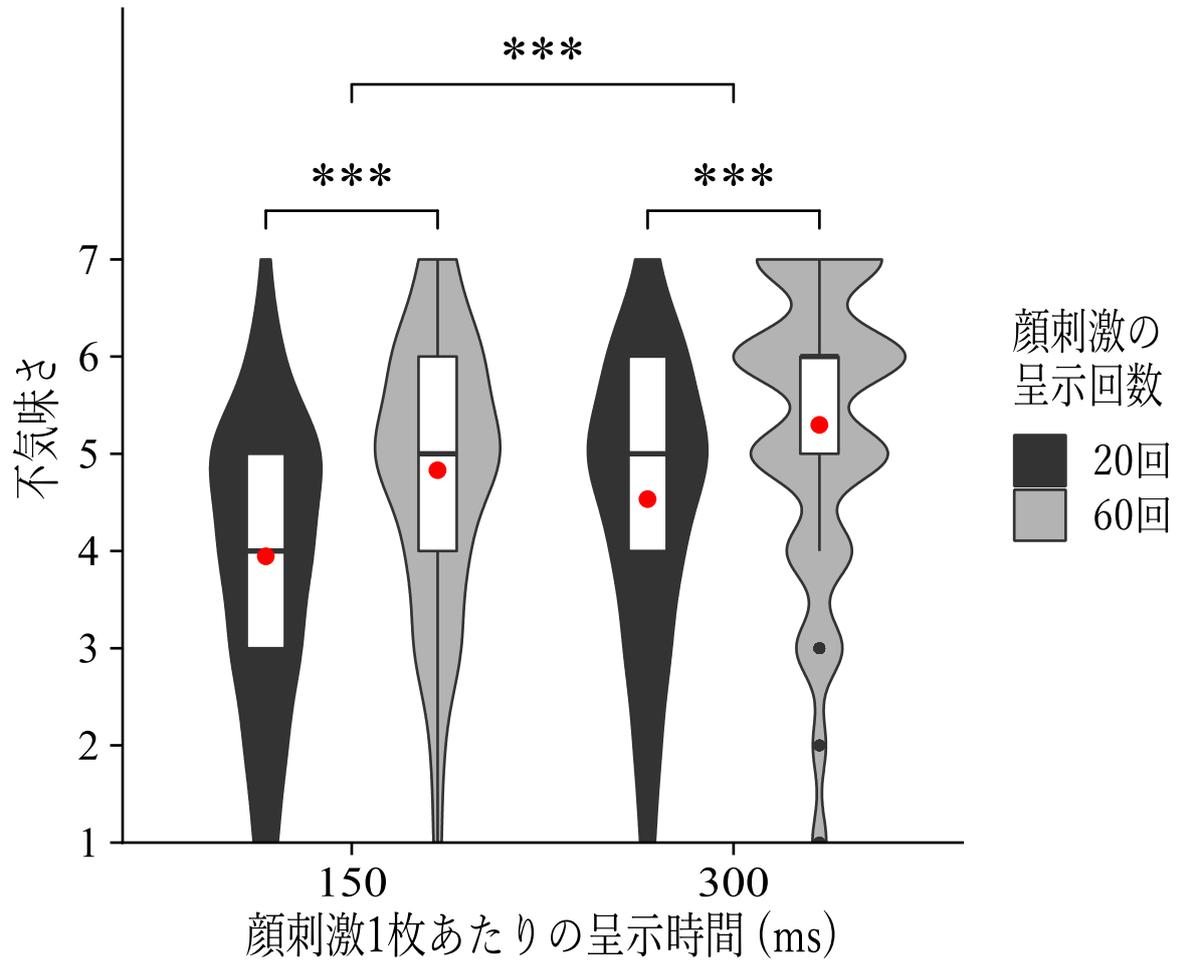


Figure 18. 実験 2 における条件ごとの不気味さ評定値。赤い丸は平均値を、\*\*\*は  $p < .001$  を表す。

## 歪みの大きさについて

条件ごとの歪みの大きさの評定値の平均値 ( $M$ )と標準偏差 ( $SD$ ) は、150 ms・20回条件では  $M = 4.24$ ,  $SD = 1.12$ , 150 ms・60回条件では  $M = 5.19$ ,  $SD = 0.87$ , 300 ms・20回条件では  $M = 4.79$ ,  $SD = 1.06$ , 300 ms・60回条件では  $M = 5.61$ ,  $SD = 0.92$  であった (Figure 19)。顔刺激1枚あたりの呈示時間と一連の顔刺激の呈示回数がFFDEによって生じる歪みの大きさに与える影響を検討するため、それぞれを要因とする対応のある2要因分散分析を行った。球面性の仮定が成立しているかを検討するMendozaの多標本球面性検定は、2要因とも参加者内要因かつ2水準であったため行われなかった。分散分析の結果、顔刺激1枚あたりの呈示時間の主効果及び、一連の顔刺激の呈示回数的主効果が有意であった(それぞれ、 $F(1, 43) = 22.3, p < .001, \eta_G^2 = .06, 95\%CI [.02, .11], MSe = 0.47$ ;  $F(1, 43) = 86.6, p < .001, \eta_G^2 = .17, 95\%CI [.11, .23], MSe = 0.40$ )。これらの要因間での交互作用は有意ではなかった( $F(1, 43) = 2.17, p = .15, \eta_G^2 = .001, 95\%CI [.00, .01], MSe = 0.09$ )。従って、顔刺激1枚あたりの呈示時間が150 msの条件は300 msの条件より歪みの評定値が有意に低いことが示された。また、一連の顔刺激の呈示回数が20回の条件は60回の条件より評定値が有意に低いことが示された。

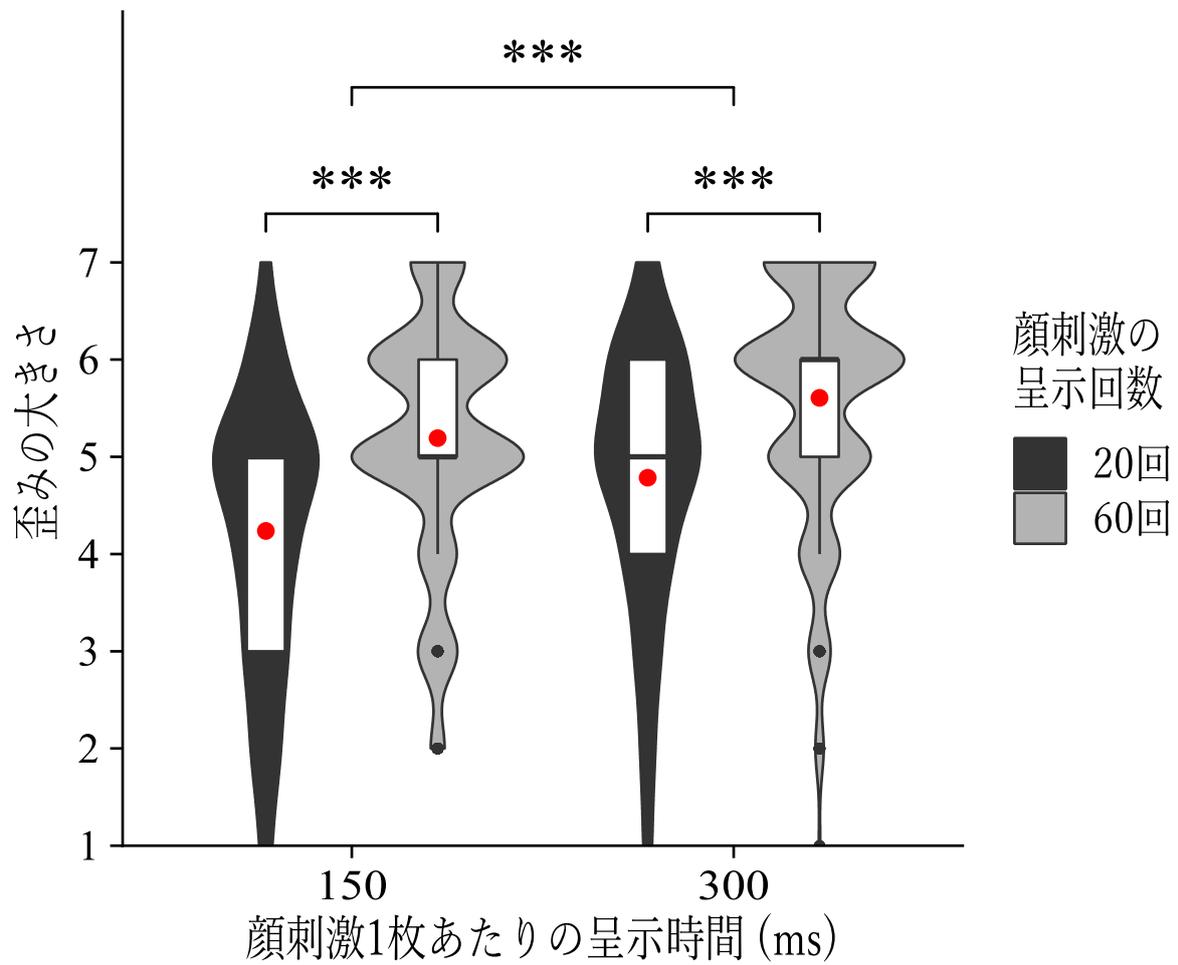


Figure 19. 実験 2 における条件ごとの歪みの大きさ評定値。赤い丸は平均値を、\*\*\*は  $p < .001$  を表す。

## 考察

実験2の目的は、実験計画法を用いて顔刺激1枚あたりの呈示時間と一連の顔刺激の呈示回数がFFDEの強さに及ぼす影響を再確認することであった。実験1の結果に基づき、1枚あたりの呈示時間が短い、呈示回数が少ないほど不気味さ・歪みの大きさの評定値が小さくなると予測した。その結果、顔刺激1枚あたりの呈示時間の主効果が有意であり、呈示時間が短いとFFDEが弱くなることが示唆された。また、顔刺激の呈示回数の主効果も有意であり、呈示回数が少ないとFFDEが弱くなることが示唆された。従って、本実験の仮説は支持され実験1の結果が再現された。また、不気味さの結果と歪みの大きさの結果は概ね同様の傾向であった。

この結果は、顔刺激1枚あたりの呈示時間が短い場合、長い場合に比べFFDEが弱まるという実験1及びBalas & Pearson (2019)を支持するものである。同様に、顔刺激の呈示回数の影響も実験1の結果を再現した。これらの結果は、ラジカルランダムマイゼーションを用いた探索的研究が、現象や効果に影響を及ぼす要因を検討する手法として有効である可能性を示唆する。

## 総合考察

### 本研究の目的と結果の要約

本研究の目的は、顔刺激を瞬間的に連続呈示することで顔が歪んで知覚される錯視である FFDE を例に、ラジカルランダムマイゼーションを用いた研究のデモンストレーションを行い、その有用性の検討と今後議論すべき点を整理することであった。実験 1 では、顔刺激の特徴や刺激呈示の時空間特性を含む計 7 つのファセットから刺激のパラメータをランダムに決定し、それらのファセットが FFDE の強さに及ぼす影響を探索的に検討した。その結果、一連の顔刺激に対する不気味さ・歪みの大きさの評定値は、顔刺激の年齢や顔向きといった刺激特徴によってほとんど変化しなかった。すなわち、FFDE は顔刺激の特徴に関しては、頑健である可能性が示唆された。それに対し、顔刺激 1 枚あたりの呈示時間や顔刺激の呈示回数は、それらが短く・少なくなるほど FFDE が弱まることが示唆された。しかし実験 1 は探索的研究であり、この結果から確証的な判断をすることは適切ではないと考えられる。そこで続く実験 2 では、実験計画法を用いて顔刺激 1 枚あたりの呈示時間と顔刺激の呈示回数が FFDE に及ぼす影響を再検討した。その結果、実験 1 と同様に顔刺激 1 枚あたりの呈示時間や顔刺激の呈示回数が短く・少なくなるほど FFDE が弱まることが示唆された。すなわち実験 1 の結果が再現された。このことから、FFDE は顔刺激の特徴については比較的頑健であり<sup>13</sup>、刺激がどのように呈示されるかという特徴には影響を受けやすい現象であると考えられる。また、実験 2 において実験 1 の結果が再現されたこ

---

<sup>13</sup> 繰り返すが、あくまでこれは探索的な研究の結果であるため、確証的に解釈することには慎重になるべきである。

とは、ラジカルランダマイゼーションを用いて探索的研究を行うことの有効性を示唆していると考えられる。

## ファセットごとの結果の整理と解釈

本研究の主たる目的は、FFDEではなくラジカルランダマイゼーションの検討である。そのため、FFDEについての考察は全てのファセットについては行わず、先行研究との比較程度に留める。

顔認知において全体処理は重要な特性である。例えば、刺激を180°倒立させるだけで顔の認知は阻害される(e.g., サッチャー錯視; Thompson, 1980)。これまでの研究で、FFDEの生起には顔の全体処理が重要であることが指摘されてきた(e.g., 鈴木他, 2022; ただし, Balas & Pearson, 2019も参照)。このことから、顔刺激が倒立ないしそれに近い角度で呈示されたときは、顔の全体処理が阻害され、FFDEが弱くなる可能性が考えられる。実際、顔が倒立しているとき、FFDEが弱まることが報告されている<sup>14</sup>(Utz & Carbon, 2015; ただし, Tangen et al. (2011)も参照)。しかし、本研究の結果はFFDEの強さは顔刺激の回転角度によって概ね変化しなかった(Figure 7, Figure 15)。このことは、FFDEが顔の全体処理に大きく依存していない可能性を示唆する(Balas & Pearson, 2019)。ただし、実験1は探索的な研究による結果であるため、FFDEと顔刺激の回転角度の関係については改めて検討することが必要である。

---

<sup>14</sup> また、Balas & Pearson (2019)でも倒立刺激でFFDEが弱まる傾向は報告されているが、ベイズファクターは倒立効果に対して弱い支持のみであり判断は保留している。

顔刺激 1 枚あたりの呈示時間については、それが短いとき FFDE が弱まることを報告している Balas & Pearson (2019) 及び Utz & Carbon (2015) を支持するものであった。FFDE が最も強く生じる最適周波数はいくつか報告されており、Tangen et al. (2011) では 4-5 Hz すなわち顔刺激 1 枚あたりの呈示時間が 200-250 ms とされている。この 4 Hz という時間周波数は、その後の研究でも多く用いられている (Bowden et al., 2019; 鈴木他, 2022; Utz & Carbon, 2015; Wen & Kung, 2014)。Utz & Carbon (2015) では、1 Hz (1 枚あたり 1000 ms) ・ 4 Hz (1 枚あたり 250 ms) ・ 10 Hz (1 枚あたり 100 ms) で検討を行っており、10 Hz の時最もグロテスクさの評定が弱く、1 Hz の時最も評定が高いと報告している。本研究の実験 1 においても 1 枚あたり 150 ms (約 6.67 Hz) では、不気味さの評定値の平均値が 3.93 と FFDE が生じていなかった。このことから、FFDE は”Flash”と命名されてはいるものの、この錯視が生じるためには顔刺激 1 枚あたり、ある程度 (およそ 200 ms 以上) の呈示時間が必要であると考えられる。

他方、一連の顔刺激の呈示回数の影響については、これまでほとんど検討されてきていない。類似した研究としては、顔刺激の呈示回数は一定に保った上で、顔刺激の繰り返しの影響が検討されている。Wen & Kung (2014) は、2 つの顔が交互に繰り返される条件、3 つの顔が交互に呈示される条件、繰り返しの無い条件と照明が異なる条件 (統制条件) の 4 水準で FFDE の強さを比較している。その結果、繰り返しが少ないほど FFDE 強くなることを報告している。このことは、顔が異なる顔に切り替わる回数が FFDE の強さには重要である可能性を示唆している。本研究の結果は、Wen & Kung (2014) の延長線上に位置づけることができると考えられる。本研

究では、注視点の左右それぞれでは顔刺激は繰り返し呈示されることはなかった<sup>15</sup>。実験 1 及び実験 2 の結果を踏まえると、同一の顔刺激が繰り返し呈示されない状況においては、新しい顔に切り替わる回数の多い、すなわち呈示回数が多いほど FFDE が強くなったと考えられる。

このことは、一連の顔刺激の呈示時間の結果からも確認できる。実験 2 において、150 ms・60 回条件 (一連の呈示時間は 9000 ms) のほうが、300 ms・20 回条件 (6000 ms) よりも不気味さ・歪みの大きさの評定値が高かった。そこで、これらの条件間に差があるのかを検討するために探索的に対応のある *t* 検定を行った。その結果、不気味さの評定値、歪みの大きさの評定値ともに、150 ms・60 回条件のほうが、300 ms・20 回条件より有意に大きかった (それぞれ  $t(43) = 2.81, p = .01, \text{Hedges' } g = 0.42, 95\% \text{ CI } [0.11, 0.73]; t(43) = 3.76, p < .001, \text{Hedges' } g = 0.56, 95\% \text{ CI } [0.24, 0.88]$ )。このことは、一連の顔刺激の呈示時間が長ければ FFDE が強く知覚されるわけではなく、定期的に顔が切り替わることが重要であることを示唆していると考えられる。

## 本研究の限界

### 実験 1・2 の共通の限界

本研究は実験 1、実験 2 ともにオンラインで実施した。そのため、対面実験と比べ刺激呈示が正確な時間精度で行えていかなかった可能性がある (Bridges et al., 2020)。ただし、本研究では 1 フレームあたりの平均時間が通常考えられる時間の 2 倍以上になっている試行は除外して分析した。また、実験 1 と実験 2 (加えてサプリメントに記載の実験) において顔刺激 1 枚あたりの呈

---

<sup>15</sup> なお、左右全体で考えると 1 度繰り返されている。すなわち、右側 (左側) で呈示された顔はその後左側 (右側) で呈示された。

示時間の結果が同様の傾向を示していることから、オンラインでの実施が実験結果に与えた影響は大きくないと考えられる。

本研究はオンライン実施であることを考慮し、刺激が呈示されるディスプレイの大きさや視距離、部屋・ディスプレイの明るさなどは統制しなかった。顔刺激の大きさについては、FFDEに影響しないことが報告されているが (Balas & Pearson, 2019), ディスプレイからの視距離が参加者ごとに異なると、それに依存して画面中央からの偏心度も変わると考えられる。偏心度が大きいほど、FFDEは強く知覚されるため (Balas & Pearson, 2019; Bowden et al., 2019), この点においては本研究の結果の解釈は慎重になる必要があると考えられる。現在では、クレジットカードや盲点を用いて視距離の統制や推定を行う方法が提案されており (Li, Joo, Yeatman, & Reinecke, 2020), 今後はこれらの方法を用いて視距離を統制したオンライン実験を行うことも有用であろう。

顔刺激について、本研究では鈴木他 (2022) と同様に AI (<https://generated.photos/>) を用いて作成したものを使用した。また、実験2の顔刺激は、刺激準備の時間等の限界から、実験1で用いたものの中から条件に一致する60枚用いた (条件については実験2の手続きセクション参照)。このことは顔刺激についての一般化可能性を制限する可能性がある。また、顔刺激の表情については、事前に評定などを行い選定・分類したわけではなく、前述のAIのEmotionパラメータを「Joy」または「Neutral」に指定したうえで、実験者の主観に基づいて使用された。そのため、表情の妥当性が担保されていない可能性がある。この2点についても結果の解釈は注意が必要であろう。また、作成した顔刺激はいずれもいわゆる他人種 (白人や黒人など日本人ではない顔) であり、かつ複数の人種が混在していた (当該AIでいわゆる日本人を選択的に生成すること

は、刺激作成時はできなかった。なお、アジア人という区分は存在した)。Utz & Carbon (2015) は、顔刺激と参加者の人種が異なると、FFDE が有意に弱まることを報告している。このことは、顔刺激の人種を統制する必要がある可能性を示唆している。なお、本研究で用いた顔刺激については OSF (<https://osf.io/f9g6s>) にて全て公開されているため、第三者による刺激の妥当性等の検討は可能である。

## 実験 1 の限界

一つは探索的研究ということである。特に、実験 2 で確認していないファセットの効果の解釈は慎重になる必要がある。例えば、顔刺激の年齢の影響について、実験 1 ではほとんど効果が見られなかったが、これは同等性検定 (Lakens, 2017; Lakens, Scheel, & Isager, 2018) やベイズファクター (e.g., 岡田, 2018) によって差がない (ないし、実質的に差がない) ことが示された結果ではない。特に効果が小さかったファセットの効果の解釈は慎重になるべきであり、差がないことを示すには再度確認研究を行い、前述の指標などを用いた分析に基づく判断を行うことが求められる。

次にラジカルランダマイゼーションに関する限界について述べたい。第一に、参加者は全ファセット・水準からなるセル (マイクロ実験に相当) で実験を行ったわけではないことが挙げられる。換言すれば、参加者一人当たりが本実験で考えられうるすべての組み合わせのマイクロ実験を実施したわけではないということである。これは、個人差が大きいような現象・効果を対象にラジカルランダマイゼーションを行った場合、その参加者が割り当てられたマイクロ実験の効果適切に解釈できない可能性があることを示唆している。第二に実験 1 では、分析者の

技術等の限界から Baribault et al. (2018) が提案する分析方法のうちグローバル検定に相当する分析のみを行った事が挙げられる。これは、参加者やマイクロ実験、各ファセットのランダム効果をいわば無視した分析である。これらのランダム効果が大きい場合は、標準誤差が過小に推定されているため、第一種の誤りの確率が高い可能性がある。なお、ローデータは OSF (<https://osf.io/h5eyq>) から入手できるため、第三者による二次分析も可能である。分析に関連して、交互作用の検討についても限界があると考えられる。詳細は今後の展望にて述べるが、マイクロ実験の数がファセット・水準に比して少ない場合、ラジカルランダムマイゼーションを用いた交互作用の検討は難しい可能性がある (DeKay et al., 2022)。これは、その条件に該当するマイクロ実験を行った参加者が少なすぎる、場合によっては存在しない可能性があるためである。本研究の目的は、ファセットが FFDE に及ぼす主効果の検討が中心であったため、問題にはならないと思われるが、交互作用の検討が目的の研究をラジカルランダムマイゼーションで行う場合には研究全体のマイクロ実験数が十分になるように、サンプルサイズや参加者一人当たりのマイクロ実験数を決定する必要があるだろう。

## 実験 2 の限界

実験 2 の条件設定では、顔刺激 1 枚あたりの呈示時間、顔刺激の呈示回数それぞれ効果と、一連の顔刺激の呈示時間 (2 要因の乗算で算出可能) を切り分けることができない。そのため、今後は、一連の顔刺激の呈示時間を統制したうえで、これら 2 要因の影響を確認する必要があると考えられる。なお、探索的分析ではあるが 150 ms・60 回条件 (一連の呈示時間は 9000 ms) のほうが、300 ms・20 回条件 (6000 ms) よりも不気味さ・歪みの大きさの評定値のいずれも有意に高

くなっていること及びそれぞれの要因の効果量から考えると、一連の呈示時間よりも顔刺激を何回呈示するかの方が FFDE への影響が大きい可能性が考えられる。これらは今後の検討課題であろう。

## 今後の展望

### FFDE 研究の今後

本研究は Balas & Pearson (2019) 及び鈴木他 (2022) と同様に FFDE の強さ (不気味さ・歪みの大きさ) を 7 件法のリッカート尺度を用いた主観評定で測定した。しかし、主観評定は様々な要因の影響を受け易いため、客観的指標を用いた測定も同時に行うことが望ましいと考えられる。鈴木他 (2022) は客観的指標による測定の試みとして、歪みを知覚するまでの時間 (潜時) と歪みを知覚した累積時間を測定に用いている。これらの指標は、FFDE の時間的な変化を明らかにできる点で有用であると考えられる。実際、本研究における一連の顔刺激の呈示時間はこれらの指標の結果を参考に決定した。本研究や Balas & Pearson (2019)、鈴木他 (2022) で用いられた主観評定は一連の顔刺激が呈示され終わった後行われた。そのため、刺激呈示中の情報を持っておらず、参加者の FFDE の知覚が時間的に変化するかは測定できない。今後は、客観的指標と時間的な変化を測定できる指標、主観評定を組み合わせた FFDE の時間特性も検討することが望まれる。

実験 1 では (探索的研究の結果とはいえ) 顔の回転角度の影響が、ほとんど見られなかった。他方、実験 1・実験 2 の結果から、顔刺激の呈示方法が FFDE の強さに重要であることが示唆された。これらの結果は、FFDE において顔の全体処理などの高次な視覚処理の重要性に疑問を呈

するものであると考えられる (Balas & Pearson, 2019)。FFDE が顔特有の現象なのかは議論が続いており (e.g., Balas & Pearson, 2019; Utz & Carbon, 2015), 現状見解の一致をみていない。顔特有の現象なのかを検討するため, 今後は顔のパーツは用いた上でその位置関係をランダムにするスクランブルフェイスや, カリカチュア, 非顔刺激 (e.g., 幾何学図形) などの刺激を用いた検討が必要であろう。

また, 実験 1 では顔刺激の特性 (e.g., 性別) は FFDE の強さに与える影響は小さかった。このことは (少なくとも顔刺激であれば) FFDE がかなり頑健に生じうることを示唆している。未だ不明な点の多い FFDE のメカニズムを探るうえでは, FFDE を強める要因だけでなく, 弱める要因 (e.g., 顔刺激の間にブランクを挟む) を見つけることも有用なアプローチであろう。

### **ラジカルランダマイゼーション及びメタスタディの今後**

ラジカルランダマイゼーションとメタスタディは本稿執筆時で提唱から 4-5 年経過しており (Baribault et al., 2018), 提唱元の論文は 120 回を超える引用がされている (2023 年 1 月 18 日現在, Google Scholar 調べ)。しかし, その方法論や分析方法などについて詳細に議論している論文は少なく (数少ない例外に DeKay et al., 2022), 実際にこの手法を用いた研究は筆者が知る限り DeKay et al. (2022) のみである。そこで, 本項では実際にラジカルランダマイゼーションを用いた実験を行った経験を踏まえ, 今後議論すべき点をまとめたい。論点を整理することで, ラジ

カルランダマイゼーション及びメタスタディを用いた研究や、これらに対する議論の活発化を期待する<sup>16</sup>。

ラジカルランダマイゼーション及びメタスタディを用いた研究を行うにあたり、障害となりうる点は、実現可能性に関する議論と実験詳細の決定法及びその正当化に関する議論に大きく分けられると考えられる<sup>17</sup>。まずは、実現可能性にかかる論点を整理したい。

### 実施可能性に関する議論

ラジカルランダマイゼーション及びメタスタディを用いた実験方法では、参加者1人当たり複数回のマイクロ実験を実施することが特徴である。このことは、複数回のマイクロ実験から構成される実験全体の実施時間が一定であれば、1回のマイクロ実験あたりは、ある程度短い時間で完了できる必要があることを示唆している。例えば、参加者1人当たり15回マイクロ実験を行い、かつ実験全体の実験時間が(参加者の集中力やモチベーション等の都合から)1時間程度が限界の場合、各マイクロ実験の実施時間はおおよそ4分前後である必要がある。ここで、知覚・認知心理学実験では1条件あたり複数回(場合によっては数十回)繰り返し測定することが一般的である。これは、1条件あたりの繰り返し回数を減らすことが実験目的上困難な研究(e.g., 恒常法を用いた閾値測定)であれば、1試行あたりの実施時間が相当に短く無い限り、ラジカルランダマイゼーションの実施は困難であることを意味する。本研究の場合、実験1における1試

---

<sup>16</sup> なお、今後行われる全ての研究をラジカルランダマイゼーション・メタスタディを用いた実験方法で行う必要はないと思われる(Baribault et al., 2018; ただし, DeKay et al., 2022も参照)。ここでの議論は、あくまで今後これらの方法を用いた研究を行いたい場合のための議論である。

<sup>17</sup> ただし、これらは相互に関連しあっており、完全に切り分けることは難しいと思われる。

行あたりの実施時間は平均的に 10 秒であり、繰り返し回数は 5 回であった。これらは鈴木他 (2022) を参考に決定したが、参加者内でのばらつきが大きい実験では繰り返し回数を少なくすることは困難であろう。このような場合、1 条件あたりの繰り返し回数は減らさず、参加者 1 名あたりが行うマイクロ実験の数を減らすことで実験全体での実施時間を調整可能である。しかし、このような方略をとると、ラジカルランダマイゼーション本来の目的である手続き空間の幅広い探索 (マイクロ実験によるサンプリング) には制約がかかると考えられる。現状では、このように 1 試行あたりの時間がかかる場合や 1 条件あたりの繰り返し回数が必要な実験を、どのようにラジカルランダマイゼーションで行うかについては議論が全く行われていない。ラジカルランダマイゼーション・メタスタディを用いた研究を普及させるためには、この課題について統計的側面も含めた議論が必要であると考えられる。

実施時間に関する議論は、実験の実施形式とも関連すると考えられる。新型コロナウイルス感染症 (COVID -19) の流行などを受け、近年ではオンラインでの実験も増えている。オンライン実験の最大の利点は、いつでもどこでも実験を行える (ないし参加できる) ことであろう。クラウドソーシングサービス等で参加者を募集すれば、対面実験よりも迅速かつ多くの参加者を集めることも可能である。他方、実施時間が長すぎる場合、脱落者や除外が多くなるなど適切なデータが得られない場合も多い。そのため、オンライン実験でラジカルランダマイゼーションを行う場合、研究全体の実施時間が対面時よりも短いことが望ましいと考えられる。なお、対面とオンラインの両方で実験を行い、それらをファセットとして分析に組み込むことも可能である。対面実験とオンライン実験で研究全体の実験時間や参加者一人あたりのマイクロ実験数を変えることも可能だが、これが統計的・分析的に妥当かについては議論が必要である。

## 実験の詳細の決定とその正当化に関する議論

再現性問題に関する議論を受け、実験実施前にサンプルサイズ設計を行うことが推奨されている (e.g., Lakens, 2022; 大久保・岡田, 2012)。ラジカルランダマイゼーション・メタスタディを用いた研究においてもこれは同様であると考えられるが、その手法については Baribault et al. (2018) では言及されていない。本研究では行わなかったが、階層線形モデルを用いた分析におけるサンプルサイズ設計については、村井・橋本 (2017) や宇佐美 (2011) で扱われている。また、DeKay et al. (2022) は、同じサンプルサイズであっても条件数が少なくかつ条件ごとの参加者が多い研究に比べ、メタスタディを用いた研究のほうが統計的な推定精度と検定力が高くなることを報告している。ここで、Baribault et al. (2018) はメタスタディの結果、すなわちそれぞれのマイクロ実験を (通常の研究における) 個別の実験とみなし、メタ分析を行う分析方法も提案している。この場合のサンプルサイズ設計については、議論が行われていない。

また、特に議論が必要であると考えられるものに、研究全体のマイクロ実験設計が挙げられる。ラジカルランダマイゼーション・メタスタディにおける分析の単位がマイクロ実験であることを考慮すると、サンプルサイズ設計同様にマイクロ実験設計は重要であると考えられる。これには統計的側面だけではなく、前述の実施可能性に関する議論も関連する。参加者が行うマイクロ実験数が一定であると仮定すれば、研究全体のマイクロ実験数は、参加者数 × 参加者一人あたりのマイクロ実験数で算出される。極端な例を出せば、研究全体のマイクロ実験の目標数を 1000 としたとき、10 名の参加者に一人あたり 100 マイクロ実験を実施する場合と 100 名の参加者に 10 マイクロ実験を実施する場合とどちらが適切 (ないし効果的) なのかということである。対面実験など一人の参加者に長時間の実験に参加を依頼できる、ないし複数回同一の実

験に参加ができる場合は前者の方略も可能であろう。知覚などの個人差が比較的小さい領域であれば、この方略が適切であるかもしれない。一方で、個人差が大きい場合や、オンラインで実施する場合は一人あたりのマイクロ実験を増やすより、多くの参加者に少ない回数のマイクロ実験を実施してもらうことが有用(ないし現実的)であるように思われる。

いずれにせよ、研究全体のマイクロ実験数を決定するためには、サンプルサイズ設計と一人あたりのマイクロ実験数の決定が必要であり、そのための議論は現時点では十分にされているとは言い難い状況である。また、サンプルサイズ・マイクロ実験数は、参加者の負担(e.g., 実験時間)や参加者募集の容易さ以外にも、統計的側面を考慮する必要があると考えられる。

それは意図一般化母域との関係である。研究者が目的の効果・現象が存在するのかを調べたい範囲が広い場合、すなわち、多様なファセットと広い水準を設定した場合、一般に異質性が高くなり検定力が低下する(DeKay et al., 2022)。そのため、広い意図一般化母域を高い検定力で探索するためには、その広さに応じたサンプル(この場合はマイクロ実験)が必要であると考えられる。ファセット数や水準の範囲(正確にはこれらによって決定される、意図一般化母域の広さ)と必要とされるサンプルサイズ・マイクロ実験数の関連については議論が無いのが現状である。

この問題は、同時にファセット及びその水準をどのように決定・正当化すればよいのかという問題でもある。これまでのラジカルランダマイゼーション・メタスタディを用いた研究は、本研究も含めファセットの選定は研究実施者が任意に選んでいる(Baribault et al., 2018; DeKay et al., 2022)。このとき、どのファセットを対象とすべきか、その数や水準の範囲をどのように決定・正当化するかは議論が行われていない。対象とすべきファセットは目的とする効果や現象

の背景にある理論に基づいて決定することで、選定理由を正当化することは可能であろう。しかし、1つの実験の中にいくつのファセットを組み込むべきかについては悩ましい問題である。ラジカルランダムマイゼーション本来の目的に照らせば、多くのファセットを組み込み可能な限り広く手続き空間を探索することが望ましいだろう (Baribault et al., 2018)。すなわち、(実施可能な範囲で)多くのファセットを採用することが目的に適していると考えられる<sup>18</sup>。

しかし、これだけでは水準の設定という問題は残る。可能な限り広い意図一般化母域を探索することが望ましいとした場合、ファセットと水準の範囲はどのように決定するのがよいのだろうか。極端な例で言い換えるならば、少ないファセットと広い水準によって探索する意図一般化母域を担保することが望ましいのか、それとも多いファセット・狭い水準を採用するのかという問題である。前述のように様々な変動を考慮した実験計画という目的を重要視するならば、後者が望ましいと考えられる。また、水準の範囲を非常に広く取ることに現実的な意味があるのかは検討する必要があるだろう。本研究の例で言えば、FFDEの一般化可能性を検討する際に、顔刺激1枚あたりの呈示時間を100msから10分までの範囲で調べることに、ほとんど意味が無いと考えられる。なぜなら、1枚あたりの呈示時間を10分にしたときにFFDEが生じるとは到底考えにくいからである(もはや”Flash”ではない)。ここまで極端なことはないとしても、

---

<sup>18</sup> ただし、どれほど多くのファセットを組み込んだとしても、当該現象や効果に関連するもの・しないもの「全て」のファセットを扱うことは不可能である。そのため、ラジカルランダムマイゼーションを用いた実験を実施したからといって、広範な一般化可能性が担保されるわけではない。例えば、実験実施の季節なども結果に影響する可能性は否定できないが、それをファセットに組み込むことは、現実問題として困難である。そもそも、そのような変数を検討する必要があるかについても議論が必要であろう。ラジカルランダムマイゼーションにおいて重要なことは、1つの実験で検討できる意図一般化母域が一般的な実験計画法の実験と比べて広いということであり、実際に検討していない範囲にまで結果を一般化できるかについては、ラジカルランダムマイゼーションも他の実験方法と同様に捉える必要がある。

水準の範囲については、まずは心理学者ないしコミュニティの「常識」の範囲で設定すればよいと考えられる。実験の結果、検討した範囲では境界条件が見られないのであれば、より広い範囲で探索すればよい。以上から、(当然ある程度の幅は持たせた上で)多くのファセット・現実的な水準の範囲で検討することが有用ではないかと考えられる。いずれにせよ、サンプルサイズ、マイクロ実験の数、ファセットの数、水準の範囲の4要素の関係性については効果量や検定力など統計的な側面も含めたさらなる議論が必要である。

また、一般化可能な範囲をどのように定義するかについても、追試の成功をどのように判断するかという議論と同様に、議論が必要であると考えられる (DeKay et al., 2022)。

### **ラジカルランダマイゼーション・メタスタディを実施するメリットはあるのか**

今後議論すべき点があることは、必ずしも手法のデメリットではない。それでは、ラジカルランダマイゼーション・メタスタディを用いるメリットには何があるのだろうか。

第一に実験1のような探索的研究としての有用性である。ラジカルランダマイゼーション及びメタスタディを用いることで、一般的な実験計画法を用いた実験よりも多くの要因(ないしファセット)・水準を1つの実験で検討可能になる。実際、実験1のように(参加者内)要因・水準数が多い実験の全条件を一人の参加者が実施することは参加者負担や実験時間を考えると現実的ではない。しかし、ラジカルランダマイゼーションを用いることで、参加者はこの全ての要因(ファセット)・水準の組み合わせ(つまりマイクロ実験)の中から一定の回数を行うだけで済む。すなわち、要因や水準が多い実験であっても参加者の負担を増やしすぎず実施できる。

また、この際、主効果に相当する影響のみならず交互作用の検討も可能である (cf.

<https://osf.io/845g3>)。ただし、(研究者が解釈可能かどうかはともかく) 高次の交互作用の確認が目的の場合は、水準数を減らすか、研究全体のマイクロ実験数を増やす必要があるだろう。このとき、通常の実験計画法を上回る効率で実験が可能なのは今後の検討が必要である<sup>19</sup>。

ラジカルランダムマイゼーションでは、ファセット・水準を多くした実験が行われる。このとき、ファセット・水準が多い(ないし、ランダムに手続き空間から実験をサンプリングする)ことは統制を放棄しているのではないかという指摘がされるかもしれない。しかし、この指摘は当たらないと考えられる。むしろ、水準を事前に明確に定義した上で実験手続きの変動を考慮している点では統制に重きをおいていると捉えることも可能である。重要なことは、ファセット・水準などの実験計画をどのように選定・設定するかであり、これは一般的な実験計画法となんら変わらない。

第二に単一の研究室でも実施可能であることが挙げられる (DeKay et al., 2022)。Baribault et al. (2018) はメタスタディは多くの研究室で共同で行う研究スタイル (e.g., ManyLabs プロジェクト; Ebersole et al., 2016, 2020; Klein et al., 2014, 2018, 2022) に向いているとしているとしたうえで、6 つの研究室で共同で行った実験を用いてメタスタディのデモンストレーションをしている。

Many Labs のような大規模追試プロジェクトでは、複数の (国の) 研究室で協力し、多くの参加者を募ることが一般的である。この形式での実施は、プロジェクトの規模が大きいがゆえに、多くの時間やコストがかかる。一方で、メタスタディは本研究や DeKay et al. (2022) のように単一

---

<sup>19</sup> しかしながら、ラジカルランダムマイゼーション・メタスタディの主たる目的は探索的研究であったとしても、目的の効果・現象が確認される範囲を実験的に検討することであろう。その点で、高次の交互作用を対象とすることは、この目的からやや逸脱している可能性はある。

の研究室でも行うことができる。大規模追試の後には、その結果についてメタ分析を実施することも多いが、メタスタディでも同様にメタ分析を行うことも可能である (Baribault et al., 2018)。一般的にメタ分析が行われるまで研究が蓄積されるには時間を要するが、メタスタディを行うことで研究の少ない領域であっても、国際的な大規模追試などを行わずともメタ分析を実施できる。

ラジカルランダマイゼーションとメタスタディを用いる最大のメリットは、Baribault et al. (2018) が主張するように、一般化可能性に留意しつつ再現性問題を意識した研究の流れを実施できることにある。ラジカルランダマイゼーション・メタスタディでは、類似した実験を複数回行い<sup>20</sup>、実験手続きに研究者が意図的に変動を持たせることで、再現性だけでなく一般化可能性についても同時に検討している。また、メタスタディを用いた研究では多くのファセットを組み込むため、データの持つ情報量が多いと考えられる。このことは、二次分析を行う上でも有益であると考えられる。

## 心理学の再現性問題の今後

最後に、今後の心理学研究において、再現性や一般化可能性を担保するための展望を述べたい。科学が科学たるためには、知見が再現されることはその根底をなす重要な条件である。心理学が科学を標榜するのであれば、蓄積されてきた知見が再現されることは科学であるための

---

<sup>20</sup> ただし、心理学の再現性問題が明らかになる一つの発端になった Bem (2011) でも 9つの実験が行われ、そのうち 8つで超能力の存在を示唆する統計的に有意な結果が報告されている。すなわち、類似した実験が繰り返し実施され、かつ同様の結果が得られていることだけでは、必ずしも再現性や一般化可能性が担保されるわけではないことには留意されたい。

必要条件であると考えられる。10年以上に渡る再現性問題に関する議論や対策によって、その解決のための道筋が徐々に見え始めてきている。これからは、その提案を実行していく段階にあると考えられる (Edlund, Cuccolo, Irgens, Wagge, & Zlokovich, 2022; Nosek et al., 2022)<sup>21</sup>。

それでは、「再現される」だけで十分なのだろうか。再現性を担保するのはいわば前提であって、今後はそれに加えて一般化可能性ないし適用範囲・境界条件を見極め報告することが重要であると考えられる。経験に裏付けられた理論の適用範囲や現象・効果の境界条件を把握することは、その後の研究の進展や現実場面の課題に対して心理学の知見を応用する上でも重要になると考えられる。

これまでの議論で再現性には、方法の再現可能性・結果の再現可能性・推論の再現可能性の3つの側面があることが指摘されている (Goodman, Fanelli, & Ioannidis, 2016)。この指摘は換言すれば、この3つの過程のそれぞれで再現性に影響しうる余地があるということにほかならない。そのため、これらの側面のそれぞれで頑健性を高め検証していくことが今後の心理学研究に求められるかもしれない。方法・結果についての頑健性を確認するための手法としては、本研究で見てきたようにラジカルランダムマイゼーション・メタスタディ (Baribault et al., 2018) が有用であると考えられる。結果の再現可能性に関連して、分析の頑健性を検討するための手法としては、マルチバース分析 (Steenen et al., 2016) や仕様曲線解析 (Simonsohn et al., 2020) などが挙げられる。推論の頑健性については、ある知見に対して異なる見解を持つ研究者とともに研究を行う敵対的共同研究 (Latham, Erez, & Locke, 1988) のように、複数の研究者ないし研究グループが独

---

<sup>21</sup> ただし、これらの対策は再現性を「高める」ものではなく、「低下させることを防ぐ」ものであると考えられる。

立で考察を執筆することが有用かもしれない。これらの手法を事前登録やオープンサイエンスといった、研究の透明性を高める手法と組み合わせ実施することが望まれる。

## 結語

本研究は FFDE に影響する要因を検討することを通じて、ラジカルランダムマイゼーション・メタスタディの有用性の検討や今後議論すべき論点を整理した。その結果、ラジカルランダムマイゼーションを探索的研究として用いることの有用性が示唆された。一方で、この手法を用いた研究を行う上では、まだ議論すべき点が多いことも示された。

一般化可能性の探究は知識の拡張でもある。今後は、結果が再現されるかだけでなく、どの範囲まで一般化可能なのか、どの程度頑健なのかを合わせて検討・報告することが必要であろう。そして、そのための手法の一つとしてラジカルランダムマイゼーション・メタスタディは有用であると考えられる。本稿が、今後これらの手法を用いた研究を行う上での一助となれば幸いである。

## 引用文献

- Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, 63(7), 602–614. <https://doi.org/10.1037/0003-066X.63.7.602>
- Balas, B., & Pearson, H. (2019). The Flashed Face Distortion Effect Does Not Depend on Face-Specific Mechanisms. *Scientific Reports*, 9(1), 1-11. <https://doi.org/10.1038/s41598-018-37991-9>
- Bem, D. J. (2011). Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of personality and social psychology*, 100(3), 407. <https://doi.org/10.1037/a0021524>
- Ben-Shachar MS, Lüdtke D, Makowski D (2020). “effectsize: Estimation of Effect Size Indices and Standardized Parameters.” *Journal of Open Source Software*, 5(56), 2815. <https://doi.org/10.21105/joss.02815>
- Baribault, B., Donkin, C., Little, D. R., Trueblood, J. S., Oravecz, Z., Van Ravenzwaaij, D., ... & Vandekerckhove, J. (2018). Metastudies for robust tests of theory. *Proceedings of the National Academy of Sciences*, 115(11), 2607-2612. <https://doi.org/10.1073/pnas.1708285114>
- Bowden, J., Whitaker, D., & Dunn, M. J. (2019). The role of peripheral vision in the flashed face distortion effect. *Perception*, 48(1), 93-101. <https://doi.org/10.1177/0301006618817419>
- Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: comparing a range of experiment generators, both lab-based and online. *PeerJ*, 8, e9414. <https://doi.org/10.7717/peerj.9414>
- Chambers, C. D. (2013) Registered Reports: a new publishing initiative at Cortex. *Cortex*, 49, 609–610. <https://doi.org/10.1016/j.cortex.2012.12.016>
- Chambers, C.D., Tzavella, L. (2022). The past, present and future of Registered Reports. *Nature Human Behavior*, 6, 29-42. <https://doi.org/10.1038/s41562-021-01193-7>
- DeKay, M. L., Rubinchik, N., Li, Z., & De Boeck, P. (2022). Accelerating psychological science with metastudies: A demonstration using the risky-choice framing effect. *Perspectives on Psychological Science*, 17(6), 1704-1736. <https://doi.org/10.1177/17456916221079611>
- Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., ... & Nosek, B. A. (2016). Many Labs 3: Evaluating participant pool quality across the academic

- semester via replication. *Journal of Experimental Social Psychology*, 67, 68-82.  
<https://doi.org/10.1016/j.jesp.2015.10.012>
- Ebersole, C. R., Mathur, M. B., Baranski, E., Bart-Plange, D. J., Buttrick, N. R., Chartier, C. R., ... & Szecsi, P. (2020). Many Labs 5: Testing pre-data-collection peer review as an intervention to increase replicability. *Advances in Methods and Practices in Psychological Science*, 3(3), 309-331.  
<https://doi.org/10.1177/2515245920958687>
- Edlund, J. E., Cuccolo, K., Irgens, M. S., Wagge, J. R., & Zlokovich, M. S. (2022). Saving science through replication studies. *Perspectives on Psychological Science*, 17(1), 216-225.  
<https://doi.org/10.1177/1745691620984385>
- Eronen, M. I., & Bringmann, L. F. (2021). The theory crisis in psychology: How to move forward. *Perspectives on Psychological Science*, 16(4), 779-788. <https://doi.org/10.1177/1745691620970586>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G\* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2), 175-191. <https://doi.org/10.3758/BF03193146>
- Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, 3(4), 456-465. <https://doi.org/10.1177/2515245920952393>
- Foster, E. D., & Deardorff, A. (2017). Open Science Framework (OSF). *Journal of the Medical Library Association: JMLA*, 105(2), 203–206. <https://doi.org/10.5195/jmla.2017.88>
- Francis, G. (2012). Too good to be true: Publication bias in two prominent studies from experimental psychology. *Psychonomic bulletin & review*, 19(2), 151-156. <https://doi.org/10.3758/s13423-012-0227-9>
- Goodman, S. N., Fanelli, D., & Ioannidis, J. P. (2016). What does research reproducibility mean?. *Science translational medicine*, 8(341), 341ps12-341ps12. <https://doi.org/10.1126/scitranslmed.aaf5027>
- 長谷川 龍樹・多田 奏恵・米満 文哉・池田 鮎美・山田 祐樹・高橋 康介・近藤 洋史 (2021). 実証的研究の事前登録の現状と実践 心理学研究, 92(3), 188-196.  
<https://doi.org/10.4992/jjpsy.92.20217>

平石 界・中村 大輝 (2022). 心理学における再現性危機の 10 年 —危機は克服されたのか, 克服され得るのか— 科学哲学, 54(2), 27-50. [https://doi.org/10.4216/jpsj.54.2\\_27](https://doi.org/10.4216/jpsj.54.2_27)

平石 界・中村 大輝 (2022b). 心理学における再現性危機の 10 年 危機は克服されたのか, 克服され得るのか (非短縮版) *PsyArXiv*. <https://doi.org/10.31234/osf.io/r72vt>

Ikeda, A., Xu, H., Fuji, N., Zhu, S., & Yamada, Y. (2019). Questionable research practices following pre-registration. 心理学評論, 62(3), 281-295. [https://doi.org/10.24602/sjpr.62.3\\_281](https://doi.org/10.24602/sjpr.62.3_281)

池田 功毅・平石 界 (2016). 心理学における再現可能性危機：問題の構造と解決策 心理学評論, 59, 3-14. [https://doi.org/10.24602/sjpr.59.1\\_3](https://doi.org/10.24602/sjpr.59.1_3)

井関 龍太 (2022). ANOVA 君. 井関龍太のページ. Retrieved December 4, 2022, from <http://riseki.php.xdomain.jp/index.php?ANOVA%E5%90%9B>

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological science*, 23(5), 524-532. <https://doi.org/10.1177/0956797611430953>

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: a new and comprehensive solution to a pervasive but largely ignored problem. *Journal of personality and social psychology*, 103(1), 54. <https://doi.org/10.1037/a0028347>

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and social psychology review*, 2(3), 196-217. [https://doi.org/10.1207/s15327957pspr0203\\_4](https://doi.org/10.1207/s15327957pspr0203_4)

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L. S., ... & Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS Biology*, 14(5), e1002456. <https://doi.org/10.1371/journal.pbio.1002456>

Klein, R. A., Cook, C. L., Ebersole, C. R., Vitiello, C., Nosek, B. A., Hilgard, J., ... & Ratliff, K. A. (2022). Many Labs 4: Failure to replicate mortality salience effect with and without original author involvement. *Collabra: Psychology*, 8(1), 35271. <https://doi.org/10.1525/collabra.35271>

- Klein, R. A., Ratliff, K. A., Vianello, M., Adams Jr, R. B., Bahník, Š., Bernstein, M. J., ... & Nosek, B. A. (2014). Investigating variation in replicability. *Social psychology*, 45(3), 142–152.  
<https://doi.org/10.1027/1864-9335/a000178>
- Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams Jr, R. B., Alper, S., ... & Sowden, W. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, 1(4), 443-490.  
<https://doi.org/10.1177/2515245918810225>
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., ... & Willing, C. (2016). Jupyter Notebooks-a publishing format for reproducible computational workflows (Vol. 2016, pp. 87-90).
- 熊田孝恒. (2019). 心理学的現象から見た脳のフィードフォワード処理とフィードバック処理: 視覚マスキング, プライミングと注意から. *日本神経回路学会誌*, 26(3), 51-61.  
<https://doi.org/10.3902/jnns.26.51>
- Kuroki, D. (2021). A new jsPsych plugin for psychophysics, providing accurate display duration and stimulus onset asynchrony. *Behavior Research Methods*, 53(1), 301-310.  
<https://doi.org/10.3758/s13428-020-01445-w>
- Lakens, D. (2017). Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses. *Social Psychological and Personality Science*, 8(4), 355–362.  
<https://doi.org/10.1177/1948550617697177>
- Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, 8(1), 33267.  
<https://doi.org/10.1525/collabra.33267>
- Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, 1(2), 259-269.  
<https://doi.org/10.1177/2515245918770963>
- Latham, G. P., Erez, M., & Locke, E. A. (1988). Resolving scientific disputes by the joint design of crucial experiments by the antagonists: Application to the Erez–Latham dispute regarding participation in goal setting. *Journal of Applied Psychology*, 73(4), 753–772. <https://doi.org/10.1037/0021-9010.73.4.753>

- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior research methods*, 47(1), 1-12. <https://doi.org/10.3758/s13428-014-0458-y>
- Li, Q., Joo, S. J., Yeatman, J. D., & Reinecke, K. (2020). Controlling for Participants' Viewing Distance in Large-Scale, Psychophysical Online Experiments Using a Virtual Chinrest. *Scientific Reports*, 10(1), 1-11. <https://doi.org/10.1038/s41598-019-57204-1>
- Monin, B., & Oppenheimer, D. M. (2014). The limits of direct replications and the virtues of stimulus sampling. *Social Psychology*, 45(4), 299–300.
- Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Percie du Sert, N., ... & Ioannidis, J. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1), 1-9. <https://doi.org/10.1038/s41562-016-0021>
- 村井潤一郎・橋本貴充 (2017). 心理学のためのサンプルサイズ設計入門 講談社.
- 武藤拓之 (2022). データ・マテリアル・分析スクリプトのオープン化が拓く心理学の未来 科学, 92(9), 800-805. <https://doi.org/10.31219/osf.io/qw43h> [出版前原稿, オープンアクセス]
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600-2606. <https://doi.org/10.1073/pnas.1708274114>
- Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., ... & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual review of psychology*, 73, 719-748. <https://doi.org/10.1146/annurev-psych-020821-114157>
- 岡田謙介 (2018). ベイズファクターによる心理学的仮説・モデルの評価 心理学評論, 61(1), 101-115. [https://doi.org/10.24602/sjpr.61.1\\_101](https://doi.org/10.24602/sjpr.61.1_101)
- Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic bulletin & review*, 26(5), 1596-1618. <https://doi.org/10.3758/s13423-019-01645-2>
- 大久保街垂・岡田謙介 (2012). 伝えるための心理統計—効果量・信頼区間・検定力 勁草書房.
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349, aac4716. <https://doi.org/10.1126/science.aac4716>

- Peels, R., & Bouter, L. (2021). Replication and trustworthiness. *Accountability in Research*, 1-11.  
<https://doi.org/10.1080/08989621.2021.1963708>
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Rubin, M. (2017). When does HARKing hurt? Identifying when different types of undisclosed post hoc hypothesizing harm scientific progress. *Review of General Psychology*, 21(4), 308-320.  
<https://doi.org/10.1037/gpr0000128>
- Rubin, M. (2020). Does preregistration improve the credibility of research findings? *The Quantitative Methods for Psychology*, 16 (4), 376–390. <https://doi.org/10.20982/tqmp.16.4.p376>
- Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2021). Why hypothesis testers should spend less time testing hypotheses. *Perspectives on Psychological Science*, 16(4), 744-755.  
<https://doi.org/10.1177/1745691620966795>
- Schimmack, U. (2021). The validation crisis in psychology. *Meta-Psychology*, 5.  
<https://doi.org/10.15626/MP.2019.1645>
- Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual review of psychology*, 69, 487-510.  
<https://doi.org/10.1146/annurev-psych-122216-011845>
- Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, 12(6), 1123-1128.  
<https://doi.org/10.1177/1745691617708630>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, 13(2), 1359-1366. <https://doi.org/10.1177/0956797611417632>
- Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, 4(11), 1208-1214. <https://doi.org/10.1038/s41562-020-0912-z>
- Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5), 702-712.  
<https://doi.org/10.1177/1745691616658637>

- 鈴木 萌々香・氏家 悠太・高橋 康介. (2022). 連続提示顔の変形効果に影響する顔部位の検討. 日本  
認知心理学会発表論文集 日本認知心理学会第 19 回大会 (pp. 15-15). 日本認知心理学会.  
[https://doi.org/10.14875/cogpsy.2021.0\\_15](https://doi.org/10.14875/cogpsy.2021.0_15)
- Tangen, J. M., Murphy, S. C., & Thompson, M. B. (2011). Flashed face distortion effect: Grotesque faces  
from relative spaces. *Perception*, 40(5), 628-630. <https://doi.org/10.1068/p6968>
- Thalmayer, A. G., Toscanelli, C., & Arnett, J. J. (2021). The neglected 95% revisited: Is American  
psychology becoming less American? *American Psychologist*, 76(1), 116-129.  
<https://doi.org/10.1037/amp0000622>
- Thompson, P. (1980). Margaret Thatcher: A New Illusion. *Perception*, 9(4), 483-484.  
<https://doi.org/10.1068/p090483>
- Torkar, R., Furia, C. A., Feldt, R., Neto, F. G. O., Gren, L., Lenberg, P., & Ernst, N. A. (2022). A Method  
to Assess and Argue for Practical Significance in Software Engineering. *IEEE Transaction on  
Software Engineering*, 48(6), 2053-2065. <https://doi.org/10.1109/TSE.2020.3048991>
- Yamada, Y. (2018). How to crack pre-registration: Toward transparent and open science. *Frontiers in  
Psychology*, 9, 1831. <https://doi.org/10.3389/fpsyg.2018.01831>
- Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, 45.  
<https://doi.org/10.1017/S0140525X20001685>
- Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress.  
*Perspectives on Psychological Science*, 13(4), 411-417. <https://doi.org/10.1177/1745691617751884>
- 植田 航平・益田 佳卓・佐々木 恭志郎・山田 祐樹 (印刷中). これからの再現性問題の話をしよう  
電子通信情報学会誌.
- 宇佐美 慧. (2011). 階層的なデータ収集デザインにおける 2 群の平均値差の検定・推定のためのサ  
ンプルサイズ決定法と数表の作成—検定力および効果量の信頼区間の観点から—. 教育心  
理学研究, 59(4), 385-401. <https://doi.org/10.5926/jjep.59.385>
- Utz, S., & Carbon, C. C. (2015). Is the Flashed Face Distortion Effect expertise-based?-a systematic  
experimental investigation. *Journal of Vision*, 15(12), 147-147. <https://doi.org/10.1167/15.12.147>

- Weisberg, M. (2013). *Simulation and Similarity: Using models to understand the world*. Oxford University Press. 松王政浩 (訳) (2017) 科学とモデル: シミュレーションの哲学入門 名古屋大学出版会.
- Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, 25(9), 1115-1125.  
<https://doi.org/10.1177/01461672992512005>
- Wen, T., & Kung, C. C. (2014). Using functional magnetic resonance imaging to explore the flashed face distortion effect. *Journal of Vision*, 14(12), 29-29. <https://doi.org/10.1167/14.12.29>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., ... & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of open source software*, 4(43), 1686.  
<https://doi.org/10.21105/joss.01686>
- Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, 41. <https://doi.org/10.1017/S0140525X17001972>

## サプリメンタリ

ここでは実験2に関連して、実験プログラムのミスで条件設定が適切に行えていなかった50名のデータとその結果について報告する。本来この実験は、顔刺激1枚あたりの呈示時間(150 ms・300 ms, 2水準)及び顔刺激の呈示回数(20回・60回, 2水準)をそれぞれ独立変数とする2要因4水準参加者内計画であった。しかし、実験プログラムのミスで「150 ms・20回」条件と「300 ms・60回」条件のみが実施された。これらの条件は、事前の予測において最もFFDEが弱い条件と最も強い条件に相当する。

本実験の手続きや実験の制御に使用したライブラリなどは実験2と同一であった。ただし、本来は2要因4水準で行う実験であったが、それを2条件で行っているため1条件あたりの繰り返し回数は実験2の倍である30回であった。なお、ここで報告する分析結果はいずれも探索的であることには留意されたい。

## 参加者

参加者の募集方法などは実験2と同一であった。実験には50名がオンラインで参加した。事前登録した除外基準(<https://osf.io/x25v6>)に基づき、全回答のうち90%が同じ選択肢を選んでいった参加者2名を除外した。その後、1フレームレートあたりの平均時間が32 msを超えている試行を除外した。これにより、2名の参加者が除外された。分析には46名(女性7名・男性37名・無回答2名、平均年齢45.0歳、標準偏差12.0歳)のデータが用いられた。

## 結果

除外後のデータについて、参加者ごとかつ条件ごとに不気味さ・歪みの大きさの評定値の平均値を算出した。効果量とその信頼区間の算出は、*effectsize* 0.6.0.1 パッケージ (Ben-Shachar et al., 2020) を用いた。それ以外の分析環境は実験 2 と同一であった。

### 不気味さについて

それぞれの条件における不気味さの評定値を Figure 20 に示した。評定値の平均値 ( $M$ ) と標準偏差 ( $SD$ ) は、150 ms・20 回条件では  $M = 4.07$ ,  $SD = 0.95$  であった。300 ms・60 回条件においては  $M = 5.43$ ,  $SD = 0.95$  であった。これらの条件間で評定値に差があるかを検討するため、探索的に対応のある  $t$  検定を行った。ただし、この際 1 フレームレートあたりの平均時間による除外で 150 ms・20 回条件のみデータがある参加者が 1 名いた。対応のある分析を行うため、その参加者の全データを除外した。除外後の 150ms・20 回条件の平均評定値は  $M = 4.09$ ,  $SD = 0.95$  であった。対応のある  $t$  検定の結果、300 ms・60 回条件の平均評定値は 150 ms・20 回条件の平均評定値よりも有意に大きかった ( $t(44) = 8.27, p < .001$ , Hedges'  $g = 1.21$ , 95% CI [0.84, 1.61])。また、評定値の大きさなどは、実験 2 における当該条件と概ね同様であった。

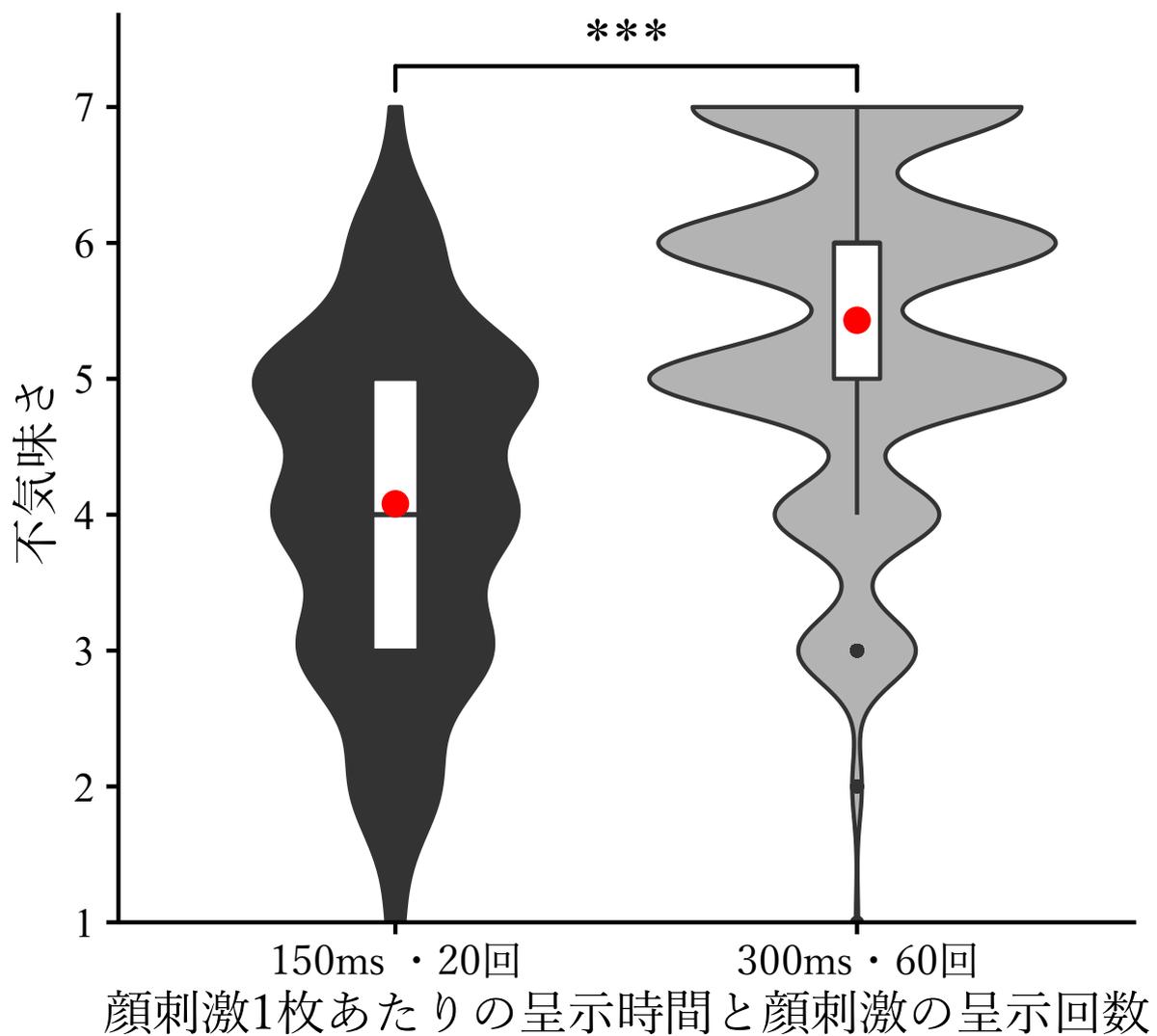


Figure 20. 条件ごとの不快味さ評定値。\*\*\*は  $p < .001$  を表す。

## 歪みの大きさについて

それぞれの条件における歪みの大きさの評定値を Figure 21 に示した。評定値の平均値 ( $M$ ) と標準偏差 ( $SD$ ) は 150 ms・20 回条件では  $M = 4.25$ ,  $SD = 0.90$  であった。300 ms・60 回条件においては  $M = 5.63$ ,  $SD = 0.80$  であった。これらの条件間で評定値に差があるかを検討するため、探索的に対応のある  $t$  検定を行った。ただし、この際フレームレートによる除外で 150 ms・20 回条件のみデータがある参加者が 1 名いた。対応のある分析を行うため、その参加者の全データを除外した。除外後の 150 ms・20 回条件の平均評定値の少数第 3 位を四捨五入した値は、除外前と同一であった。対応のある  $t$  検定の結果、300 ms・60 回条件の平均評定値は 150 ms・20 回条件の平均評定値よりも有意に大きかった ( $t(44) = 8.97, p < .001$ , Hedges'  $g = 1.31$ , 95% CI [0.92, 1.73])。また、評定値の大きさなどは、実験 2 における当該条件と概ね同様であった。

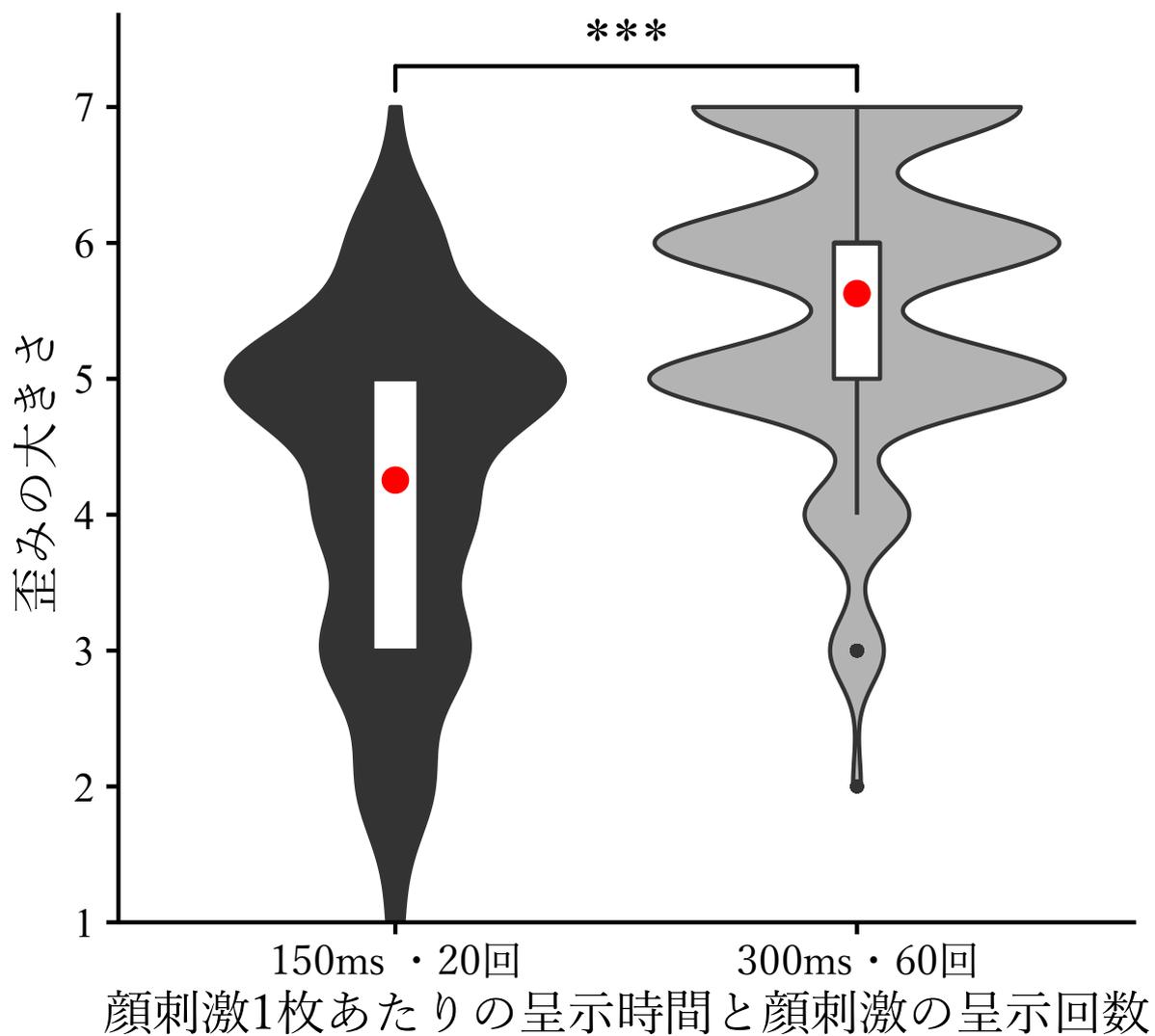


Figure 21. 条件ごとの歪みの大きさ評定値。\*\*\*は  $p < .001$  を表す。

## 考察

探索的な分析ではあるが、不気味さ・歪みの大きさの平均評定値はいずれも 150 ms・20 回条件のほうが 300 ms・60 回条件よりも有意に小さかった。これは事前の仮説を概ね支持する結果であった。この分析は探索的であり、かつ 2 要因が交絡しているため、明確な結論を下すことは控えるが、顔刺激 1 枚あたりの呈示時間ないし顔刺激の呈示回数が短い・少ないほど FFDE が弱くなる傾向が読み取れる。また、150 ms・20 回条件の平均評定値は、不気味さが 4.07、歪みの大きさが 4.25 と小さいことから、FFDE の生起において、これらの要因 (のいずれか、または両方) が重要であると考えられる。

## 研究倫理

本研究のプロトコルは、九州大学研究倫理審査委員会の承認を受け実施された (承認番号: 2022-018)。全ての参加者は実験参加前に教示を読み、同意した上で参加した。実験への参加はいつでも取りやめることができ、その場合であっても参加者には一切の不利益はなかった。実験を中断した参加者のデータは保存されなかった。参加者の個人情報個人が特定されない形で収集され保護される。本研究はヘルシンキ宣言に基づいて実施された。

## 利益相反

本研究に関して表明すべき利益相反はない。

## オープンプラクティス

本研究における全ての事前登録の内容や、実験プログラム・刺激・分析プログラム等は以下の URL からアクセスできる ([osf.io/jzvt5](https://osf.io/jzvt5))。

## 謝辞

本修士論文は九州大学大学院人間環境学府行動システム専攻修士課程在学中に行った研究をまとめたものです。本研究の実施及び論文の執筆にあたり、ご指導ご鞭撻をいただきました指導教官の山田祐樹先生に心より感謝致します。また、研究室の同期の植田航平さんや中京大学・立命館大学の鈴木萌々香さん、ReproducibiliTea Tokyo, ReproducibiliTea Fukuoka の皆さまには、本研究において研究の方針や進め方などさまざまな点について議論し、多くの示唆を与えていただきました。研究室の先輩である劉歆緒さん、張倫博さん、高嶋魁人さんには修士論文の執筆にあたり、多くのご助言を頂きました。そして、九州大学文学部技術専門職員の黒木大朗さんには、実験プログラムの作成をしていただきました。衷心より御礼申し上げます。

最後に、行動システム専攻の同期の友人・後輩、実験に参加していただいた実験参加者の皆様、そしていつも私を支えてくれた両親と多くの議論をしてくれた弟に心より感謝致します。誠にありがとうございました。