

アンカリング効果は意味プライミングか: Mussweiler & Strack (2000) の概念的追試 Is the anchoring effect a form of semantic priming?: A conceptual replication of Mussweiler and Strack (2000)

西田 勇樹¹, 山崎大暉², 服部 雅史³
Yuki Nishida¹, Daiki Yamasaki², Masasi Hattori³

¹立命館大学BKC社系研究機構, ²京都大学大学院文学研究科, ³立命館大学総合心理学部

¹BKC Research Organization of Social Sciences, Ritsumeikan University

²Graduate School of Letters, Kyoto University

³College of Comprehensive Psychology, Ritsumeikan University

yuki-n@fc.ritsumei.ac.jp

概要

アンカリング効果は初めに提示された数値（アンカー）が後の数値推定に影響を与える現象である。Mussweiler & Strack (2000) は、ドイツの年間平均気温に関する低気温のアンカー（5度）を提示した後の語彙判断課題で、冬に関する単語への反応時間が短くなることを示した。この結果は、アンカーが関連知識のアクセシビリティを高めることでアンカリング効果を生じさせるとする選択的アクセシビリティモデルを支持する。本研究では、この理論の妥当性を複数の概念的追試とメタ分析によって再検討した。その結果、アンカリング効果は確認されたが、アンカーの高さに関連する単語への反応時間の短縮は一貫して認められなかった。この結果は、Mussweiler & Strack (2000) の再現性および、理論の一般性に疑問を投げかけるものである。

キーワード: アンカリング効果 (anchoring effect), 選択的アクセシビリティモデル (selective accessibility model), 概念的追試 (conceptual replication)

1. はじめに

アンカリング効果は、初めに提示された数値（アンカー）が後の数値推定に影響を与える現象である。たとえば、参加者に「国連加盟国に占めるアフリカ諸国の割合は10%（または65%）より多いか少ないか」の比較判断をさせた後に、絶対判断としてその数値を回答させると、低いアンカーが提示された条件で絶対判断の回答の値がより低くなる (Tversky & Kahneman, 1974)。

アンカリング効果を説明する数ある理論のうち、選択的アクセシビリティモデル (Strack & Mussweiler, 1997) は、この領域で頻繁に取り上げられる。この理論は、アンカーに合致する知識が選択的に生成され、その知識のアクセシビリティが高まり、絶対判断で利用され、その結果としてアンカリング効果が生じると説明する。また Mussweiler & Strack (1999) は、このプロセスが意味プライミングに類似すると述べている。

Mussweiler & Strack (2000) (以後、MS2000) の Study 1

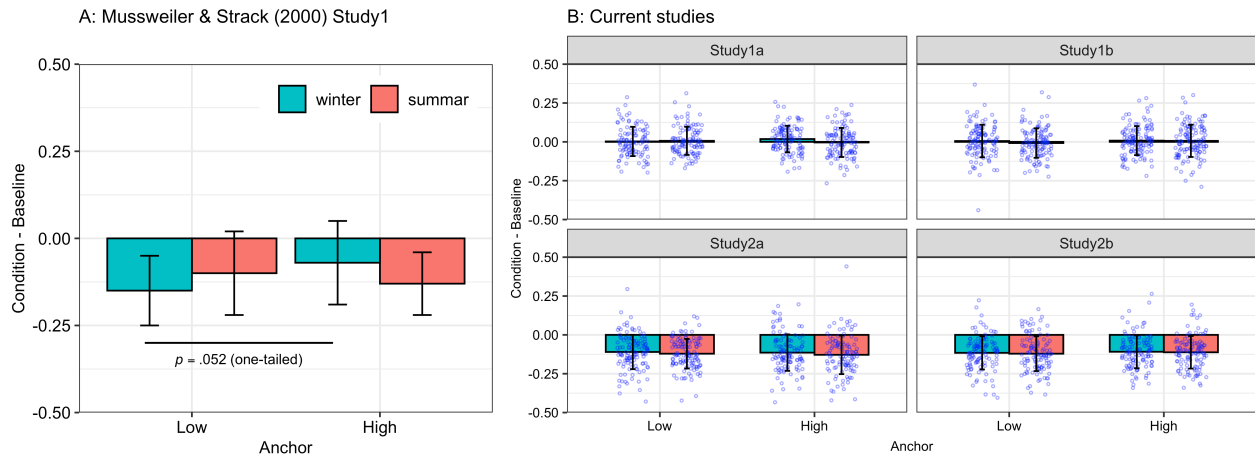
は、ドイツ語圏の参加者に比較判断として半数の参加者にドイツの年間平均気温は5度（または20度）より高いか低いかわね、絶対判断としてドイツの年間平均気温を数値で回答させた。続く語彙判断課題 (lexical decision task; 以後、LDT) では、冬に関連する単語の冬語、夏語、中立語、意味をもたない非単語の全60語を1語ずつ提示した。参加者はそれらに対し単語かどうかできるだけ素早く判断した。MS2000は、比較判断で5度を提示した低アンカー条件の冬語に対する反応時間が高アンカー条件に比べ、短いことを報告した (Figure 1A)。MS2000はLDTを用いた類似の実験を Study 2 でも行い、同様の傾向を報告した。MS2000は、アンカー値に関連する単語の反応が促進されたことで、意味プライミングに類似したプロセスの関与を示唆した。

MS2000の結果は選択的アクセシビリティモデルの中心的証拠の一つであり、この理論は本邦における心理学の専門書籍や論文 (e.g., 広田他, 2018; 本田, 2023) でも引用され、これらの影響力は近年も続いていると考えられる。しかし、英語圏で行われた MS2000 の直接追試は再現に失敗し (Harris et al., 2019), その方法論と理論に対する疑念が生じている。一方で、MS2000の結果が日本語刺激を用いた文脈で再現可能かは検討されていない。理論の有効性が特定の言語や刺激によって左右される可能性がある以上、異なる文脈での検証は理論的一般性の評価に不可欠である。そこで本研究では、選択的アクセシビリティモデルを検証するために、MS2000と類似の方法による4件（実験1a, 1b, 2a, 2b）の概念的追試を実施した。加えて一連の実験の効果を評価する目的でメタ分析を実施した。

2. 実験1

方法 実験 1a (<https://osf.io/hebxu;>) および 1b

Figure 1
Mussweiler & Strack (2000) Study 1 (A) と本研究 (B) の結果



注) 軸と凡例はすべて共通である。縦軸は参加者ごとに冬語/夏語条件の反応時間からベースラインの反応時間(中立語と非単語の平均)を差し引いた平均値を示す。バーが下に伸びるほど、反応時間がベースラインより短いことを意味する。左 (A) は MS2000 の Table 2 をもとに作成し、 p 値は同じ研究の t 値から算出された (MS2000 では “ $p < .05$, one-tailed” と報告)。右 (B) の実験 1a, b では中立語の単語親密度が高く、反応時間が短くなった結果、縦軸の値がゼロ付近または正の値を示したと考えられる。ドットは各参加者のデータ、エラーバーは標準偏差を示す。

(<https://osf.io/wgm94>) を OSF に事前登録した。実験 1a に 292 名、実験 1b に 296 名がオンラインで参加した。参加者は、低アンカー条件と高アンカー条件に分かれ、札幌市の年間平均気温は 4 度 (または 20 度) より高いか低いかを回答する比較判断を行った。絶対判断では、札幌市の年間平均気温の数値を推定し回答した。その後の LDT は夏語・冬語・中立語・非単語の 60 試行で構成され、参加者は、提示された語が意味を持つ単語かどうかできるだけ素早く判断した。なお、実験 1b でのみ比較判断で 10 秒間考えるよう指示が与えられ、その 10 秒後に回答可能となった。この手続きは MS2000 がないが、十分に考える時間を与えることで、アンカーに合致する知識の活性化を促す意図で導入された。

結果 実験 1a の最終的な分析対象者数は、226 名 (低アンカーと高アンカーでそれぞれ、 $n=110+116$) であり、実験 1b は 248 名 ($n=122+126$) であった。一貫して高アンカー条件は低アンカー条件よりも絶対判断の推定値が高かった (実験 1a: $t(214) = 10.3, p < .001, d = 1.37$; 実験 1b: $t(234) = 10.6, p < .001, d = 1.35$)。すなわち、アンカリング効果が確認された。

LDT の分析では、正答反応の試行における反応時間について、アンカーの要因 (低アンカー/ 高アンカー; 参加者間) と単語の要因 (冬語/ 夏語; 参加者内) について分散分析を実施した。MS2000 を踏襲し、反応時間を自然対数変換した上で、参加者ごとに冬語または夏語の反応時間平均から中立語と非単語を合わせた反応時間平均を減算した値の平均値を従属変数とした。その結果 (Figure 1B)、アンカーと単語の交互作用は有意

ではなかった、 $a: F(1, 224) = 3.3, p = .071$; $b: F(1, 246) = 0.6, p = .443$ 。一貫して冬語に対する高アンカー条件と低アンカー条件の差は認められず、 $a: t(224) = -0.9, p = .391, d = -0.11$; $b: t(245) = 1.1, p = .270, d = 0.14$ 。夏語に対する高アンカー条件と低アンカー条件の差も認められなかった、 $a: t(224) = 1.3, p = .199, d = 0.17$; $b: t(241) = 1.1, p = .339, d = 0.03$ 。実験 1 は MS2000 の結果を支持しなかった。

3. 実験 2

実験 1 の手続きは、絶対判断の後に LDT が続くため、絶対判断が LDT の反応時間に影響した可能性もある。そこで、MS2000 の Study 2 と同様の手続きを実験 2 で採用し、比較判断、LDT、絶対判断の順で実施した。

方法 実験 2a (<https://osf.io/57cty>) および 2b (<https://osf.io/neamb>) を OSF に事前登録した。実験 2a に 296 名、実験 2b に 303 名がオンラインで参加した。実験 1 と同様、参加者は、札幌市の年間平均気温に関する比較判断を行った。アンカー値は実験 1 と同じであった。実験 2 では、比較判断の直後に LDT を実施した。LDT 終了後、札幌市の年間平均気温に関する絶対判断に回答した。実験 2a でのみ実験 1b と同じく比較判断で 10 秒間考えるよう指示が与えられた。実験 2 の LDT の中立語は、中立語の単語親密度を夏語と冬語に近づけるよう実験 1 とは異なる単語に変更された。

結果 実験 2a の分析対象者数は 238 名 ($n = 125 + 113$) で、実験 2b は 234 名 ($n = 118 + 116$) であった。実験 2 でも一貫して高アンカー条件は低アンカー条件よりも

絶対判断の推定値が高かった, $a: t(236) = 7.6, p < .001, d = 0.99$; $b: t(232) = 9.1, p < .001, d = 1.19$ 。すなわち, アンカリング効果が確認された。

LDT における分散分析の結果 (Figure 1B), アンカーと単語の交互作用は観察されなかった, $a: F(1, 236) = 0.03, p = .861$; $b: F(1, 232) = 0.02, p = .902$ 。冬語に対する高アンカー条件と低アンカー条件の差は認められなかった, $a: t(209) = -0.5, p = .629, d = -0.06$; $b: t(231) = 0.6, p = .526, d = 0.08$ 。また夏語に対する高アンカー条件と低アンカー条件の差も認められなかった, $a: t(230) = -0.3, p = .780, d = -0.04$; $b: t(232) = 0.5, p = .621, d = 0.07$ 。実験 2 は MS2000 の結果を支持しなかった。

4. メタ分析

分析対象 メタ分析の対象は, MS2000 の 2 件の実験, MS2000 の追試研究である Harris et al., (2019) の 4 件, 上述の本研究 4 件, および, 大学の心理学実験実習の授業で収集した 12 件の実験, 合計 22 件とした。実習授業のデータは, 12 の独立したクラスで, 本研究と類似の方法で対面で実施され, 参加の同意を得て収集された。MS2000 の Study 2 と Harris et al. (2019) の Study 2, 4 は, 本研究と類似の手続きであったが, 用いられたアンカリングと LDT の題材が自動車の価値に関するものであった。これを考慮し, 低アンカーが反応時間を短縮すると予測される単語を低アンカー関連語 (本研究では冬語に該当する) と呼ぶ。メタ分析では, 2 つの比較に関する効果量 (Hedges' g) に焦点をおいた。第一に高アンカー条件と低アンカー条件の推定値の比較, 第二に低アンカー関連語における高アンカー条件と低アンカー条件間の比較であった。それぞれ, アンカリング効果と MS2000 の LDT の結果に対応する。

結果 アンカリング効果は一貫して認められ, 大きな効果量を確認した, $t(21) = 7.8, p < .001, g = 1.22, 95\% \text{ CI} [0.92, 1.53]$ 。次に各実験の低アンカー関連語の効果量を Figure 2 に示した。低アンカー関連語の効果は認められず, プールされた効果量は小さかった, $t(21) = 1.3, p = .222, g = 0.06, 95\% \text{ CI} [-0.04, 0.16]$ 。実験間および著者間 ($P_{\text{Level } 2} < 0.1 \%$; $P_{\text{Level } 3} < 0.1 \%$) で異質性は確認されなかった。これは, アンカーから低アンカー関連語の反応時間に与える効果が一貫して小さいことを示唆する。

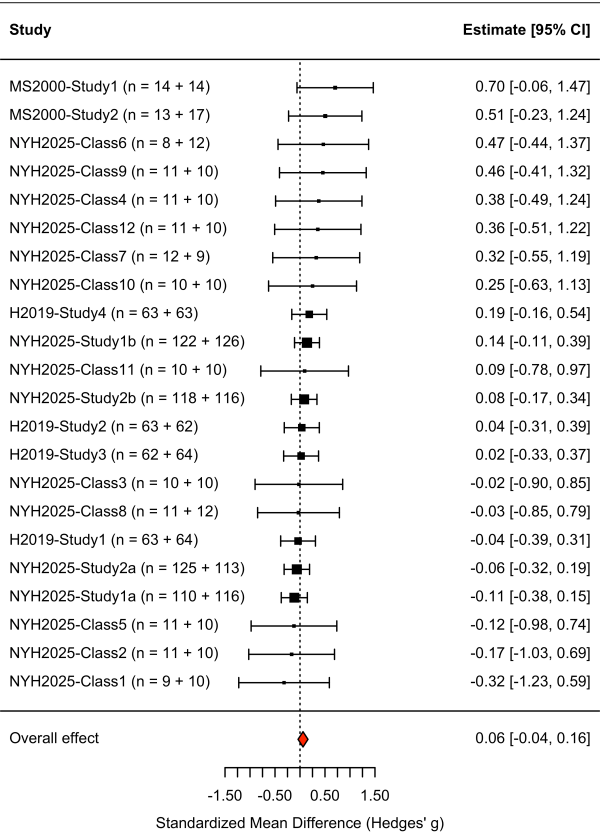
5. 考察

本研究結果は, アンカリング効果の頑健性を確認す

るとともに, その効果を意味プライミングに類似したプロセスで説明する見解に疑問を投げかけるものとなった。確かに, 本研究と MS2000 の間には, 日本語刺激を用いた LDT, オンライン実験の環境など, 方法上の違いが存在するため, こうした要因で LDT への効果が消失した可能性は否定できない。しかし, MS2000 の結果が特定の手法や刺激に依存しており, わずかな操作の違いで効果が消失するのであれば, その再現性の乏しさそのものが選択的アクセシビリティモデルの一般性に疑念を生じさせる。すなわち, この理論の適用範囲が限定的である可能性がある。

またメタ分析の結果は MS2000 を全面的に否定するものではない。低アンカー関連語に対するアンカー条件間のプールされた効果量 (Figure 2 の “Overall effect”) は小さく統計的に有意ではないものの, 効果の方向性は MS2000 の結果と一致しており, 信頼区間の下限も

Figure 2
LDT のメタ分析結果を示したフォレストプロット



注) 効果量の値が正であれば, 低アンカー関連語において低アンカー条件の方が高アンカー条件より反応時間が短縮したことを意味する。図の左から順に, 実験名, 効果量のプロット, 効果量の値と 95% 信頼区間を示し, 効果量の大きい順に上から並べている。プロットの位置は効果量を, プロットの大きさはメタ分析における推定精度 (標準誤差の逆数) を表し, エラーバーは 95% 信頼区間を示す。実験名は著者で分類され, “MS2000” は Mussweiler & Strack (2000), “H2019” は Harris et al. (2019), “NYH2025” は本研究を指す。

ゼロに近い。個別の効果量を見ると、MS2000 や実習授業で実施されたサンプルサイズの小さい実験では、比較的大きな効果が得られた。特に、MS2000 の2件の効果量が最大で、実習授業実験の6件の効果量はMS2000と同じ方向性で、小程度以上の効果量 ($g > 0.2$) を示した。しかし、サンプルサイズの大きい本研究や Harris et al. (2019) では、効果量はより小さく抑えられた。これはまさに、小規模実験の効果量がしばしば過大推定され (Button et al., 2013), 追試の効果量がオリジナル研究よりも小さくなる (Open Science Collaboration, 2015) といった、心理学で広く指摘されている問題と一致している。したがって、過大であった可能性がある。

これらの解釈に不確実性が残る背景には、アンカリング効果の諸理論に内在する構造的な問題があると考えられる。特に、理論間で明確に対立する予測が導きにくい (Schley & Weingarten, 2023), アンカリング効果を同時に複数の理論で説明できてしまい、個別理論の検証が困難になっている、たとえば、不十分な調整説 (Tversky & Kahneman, 1974) は、アンカリング効果を説明する別の理論として提案されている。この理論では、参加者がアンカーを出発点として推定値を調整し、その調整は不完全であると仮定する。不十分な調整説と選択的アクセシビリティモデルは異なるメカニズムを想定しているが、「知識のアクセシビリティが推定の出発点を方向づけ、その後不十分な調整が加えられる」のように、両者を統合的に解釈できてしまう。このような理論構造では仮説の反証や理論間の競合的検討が難しい。もっとも、理論間の境界の曖昧さや排他性の欠如は心理学全般に共通する課題でもある (Meehl, 1978)。重要なのは、このような状況を踏まえて、いかにして理論を精緻化し検証可能な形で構築していくかである。

今後のアンカリング研究では、理論を定性的に記述するだけではなく、数理モデルに基づいた定量的な予測の提示が一つの方向性となるかもしれない。たとえば、特定の仮定に基づいた絶対判断における推定値の分布のシミュレーションや、既存理論を形式的にモデル化し、それらの予測を比較する手法が考えられる。実際、アンカリング効果のメカニズムを記述する数理モデルはすでに検討されており、アンカーが数値判断に及ぼす影響を定量化する試みもある (Turner & Schley, 2016)。もっとも、現時点ではアンカリング効果に関する諸理論の数理モデルの枠組みは十分に確立されているわけではなく、重要研究の追試やメタ分析を含めた理論的整備が必要である。とはいえ、モデル比較を通じ

て理論間の差異を定量的に評価することは、アンカリング効果の理論の検証と精緻化に貢献するであろう。

文献

- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365–376. <https://doi.org/10.1038/nrn3475>
- Harris, A. J. L., Blower, F. B. N., Rodgers, S. A., Lagator, S., Page, E., Burton, A., Urlichich, D., & Speekenbrink, M. (2019). Failures to replicate a key result of the selective accessibility theory of anchoring. *Journal of Experimental Psychology: General*, 148(9), e30–e50. <https://doi.org/10.1037/xge0000644>
- 広田すみれ・増田真也・坂上貴之 (2018). 心理学が描くリスクの世界——行動的意思決定入門 第3版 慶應義塾大学出版会
- 本田秀仁 (2023). エラーか合理性理解の出発か?——認知バイアスの意味を再考する—— 基礎心理学研究, 42(1), 128–133. <https://doi.org/10.14947/psychono.42.22>
- Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46(4), 806–834. <https://doi.org/10.1037/0022-006X.46.4.806>
- Mussweiler, T., & Strack, F. (1999). Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology*, 35(2), 136–164. <https://doi.org/10.1006/jesp.1998.1364>
- Mussweiler, T., & Strack, F. (2000). The use of category and exemplar knowledge in the solution of anchoring tasks. *Journal of Personality and Social Psychology*, 78(6), 1038–1052. <https://doi.org/10.1006/jesp.1998.1364>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aa4716. <https://doi.org/10.1126/science.aac4716>
- Schley, D., & Weingarten, E. (2023, October 6). 50 Years of Anchoring: A Meta-Analysis and Meta-Study of Anchoring Effects. *PsyArXiv*. <https://doi.org/10.31234/osf.io/teq7k>
- Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology*, 73(3), 437–446. <https://psycnet.apa.org/doi/10.1037/0022-3514.73.3.437>
- Turner, B. M., & Schley, D. R. (2016). The anchor integration model: A descriptive model of anchoring effects. *Cognitive Psychology*, 90, 1–47. <https://doi.org/10.1016/j.cogpsych.2016.07.003>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1130. <https://doi.org/10.1126/science.185.4157.1124>