大正大学認知心理学レポート
2020, Vol. 4, 33-44.

人間科学専門演習Ⅰ・Ⅱ
人間科学応用演習Ⅰ・Ⅱ

複数グループからなるグラフの平均を読み取る際には面積の情報が利用される

高尾肇・中島瑠菜・惣野昇太

大正大学心理社会学部人間科学科

指導教員：井関龍太

要旨：複数のグループからなるグラフから平均値を読み取ることを求めるとき，人は平均値そのものを計算するのではなく，各グループを表す棒やドットの面積を手がかりとして判断している可能性がある。本研究では，棒グラフや点グラフの太さを操作して平均値の判断を求めた。実験の結果，棒や点の面積の大きいグループが平均値の大きいグループとして選ばれやすくなることがわかった。

問　題

多くのデータを一目で見て取れるようにまとめる方法としてグラフを提示することは効果的である。グラフは私たちの身近な場面で使われている。多くのグラフは実際に数値を並べているわけではない。それにもかかわらずグラフから情報を読み取ることができるのは，私たちがグラフの視覚的特徴を数量の代替（proxy）として利用できるからである。グラフには棒グラフ，円グラフなどの種類がありデータのどの部分を見せたいかによって使い分けられる。これに応じて，私たちもグラフの種類によって代替として使用する視覚的特徴を使い分けている。主にはグラフの空間位置，長さ，面積などを代替として使用しているが，これらの代替として使用する情報によってはグラフからデータを読み取る精度が異なる（Cleveland & McGill, 1984; Heer & Bostock, 2010）。

Yuan, Haroz, & Franconeri（2019）は，二つのグループからなるデータを棒グラフで表すときにグループ同士で棒の本数が異なると，本数が多いグループの平均値が高いと考えてしまう現象を報告した。彼らは，代替として使用する情報はグラフの種類によって変化するだけではなく，同じ種類のグラフでも単一のグループを表示する場合と複数のグループを表示する場合で異なると考えた。彼らは，グループを表すグラフから平均値を推測する際に位置情報が代替情報として利用される点グラフ（dot plot），位置情報と面積の情報が利用される棒グラフ，面積のみが利用される不整列グラフ（棒グラフをｘ軸から浮かし，棒ごとに位

置がばらばらにずれているグラフ）を比較した。実験の結果，複数グループの比較では位置ではなく，面積を代替として使用していることが示唆された。グループを構成するデータの数（棒の本数）が異なる場合，棒グラフと不整列グラフでは平均値の判断がより困難になった。参加者は複数の棒から構成される領域の広さ（面積）を代替としているため，棒の本数が異なると，本数が多いグループの平均値が高いと考えることによるものと解釈された。また，彼らはグラフの総ピクセル数によって判断の精度に変化が起きるのかを調べる追加実験も行った。この実験ではグラフの総ピクセル数を少なくすることで（棒を細くするなど）複数グラフの面積を全体的に小さくした結果，データを読みとる精度が向上したという結論が得られた。しかし，この実験ではピクセル数を増加させた場合や，棒の本数の異なるグループ間で棒の太さなどが異なる場合が検討されていなかった。そこで，本研究では Yuan et al.（2019）の実験をもとに，グループごとの棒や点の太さを操作することで平均値の判断に変化が起きるのかを調べる。このことから複数グループの平均値を判断する場合に面積が代替として使用されるという解釈の妥当性を確かめる。

本研究では，Yuan et al.（2019）にならって，2 つのグループからなるグラフから平均値を読み取って比較する場合，人は面積を代替として使用するという仮説を検証する。実験では左右に並んだ 2 つのグループを比較して，平均値が高いと判断したグループがどちらかを回答することを求める。面積を操作するために High，Low，Even 条件を設けた。High 条件では平均値が高いグループの棒や点の太さを増やすことでグラフの面積を大きくした。Low 条件では平均値が低いグループの棒や点の太さを増やすことでグラフの面積を大きくした。Even 条件では面積が両グループで等しくなるようにした。仮説が正しければ，グラフの面積が大きい方のグループの平均値が高いという回答が増えると考えられる。そのため，平均値が高いグループの面積が大きい High 条件では，面積が両グループで等しい Even 条件よりも正答率が高まり，平均値が低いグループの面積が大きい Low 条件では，Even 条件よりも正答率が低くなると予測される。

方　法

実験参加者

大学生 27 名（男性 15 名，女性 12 名）が実験に参加した。参加者の平均年齢は 20.63 歳（$SD = 1.25$）であった。視力に問題のある者はいなかった。

刺激と装置

実験はコンピュータ（HP, Parilion 500-340jp/CT）にインストールされた実験ソフトウェア PsychoPy（version 3.2.4）を用いて行われた。刺激は 24 インチのモニター（BENQ, XL2420Z）に提示され，参加者の反応はキーボードを用いて計測した。参加者のモニターとの距離は約 60 cm であった。

刺激は 2 つのグループからなるデータをグラフで表したものを用いた。図 1 に例を示した。刺激となるグラフグループには点グラフ（以下，Dot），棒グラフ（以下，Bar），不整列グラフ（棒グラフを x 軸から浮かして，棒ごとに位置がばらばらにずれているグラフ；以下，Mis）の 3 種類があった。グラフの太さについては，左右のグラフのうち平均値が低いグループの方が太い Low 条件，平均値が高いグループの方が太い High 条件，両方のグループの太さが同じ Even 条件を設けた。グラフの太さ（棒グラフと不整列棒グラフでは棒の水平方向の大きさ，点グラフでは点の直径に当たる）は太いものが 14 pix，細いものが 7 pix であり，左右のグラフの間隔は 28 pix であった。Even 条件では 14 pix と 7 pix のどちらに統一するかは試行ごとにランダムに決定した。なお，いずれのグラフにおいても左側のグラフは赤色，右側のグラフは青色であった。

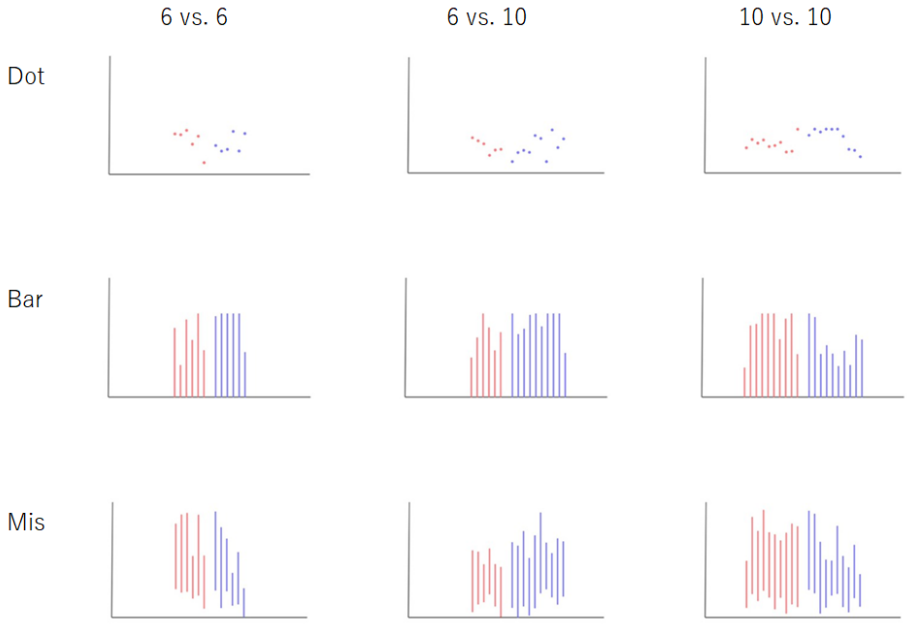

図 1　刺激の例

1 グループが 6 個のデータからなる 6 vs. 6 条件と 1 グループが 10 個のデータからなる 10 vs. 10 条件，一方のグループが 6 個，もう片方のグループが 10 個のデータからなる 6 vs. 10 条件の 3 種類のセットサイズを用意した。6 vs. 10 条件については，データが 6 個のグループと 10 個のグループのいずれが左右のどちらに配置されるのかは試行ごとにランダムに決定した。

手続き

本試行を行う前に参加者に実験概要と回答の方法について教示した。参加者には，グラフは車のレースゲームにかかった時間だと説明した。青色で塗られている方が青チーム，赤色

で塗られている方が赤チームで，平均してどちらのチームの所要時間が長いかを判断してキーボードを使って回答するように指示した。

各試行では，モニターの中央に注視点が 500 ms 表示された後，グラフが画面の中央に表示され，キーが押されるまでの時間の計測を行った。

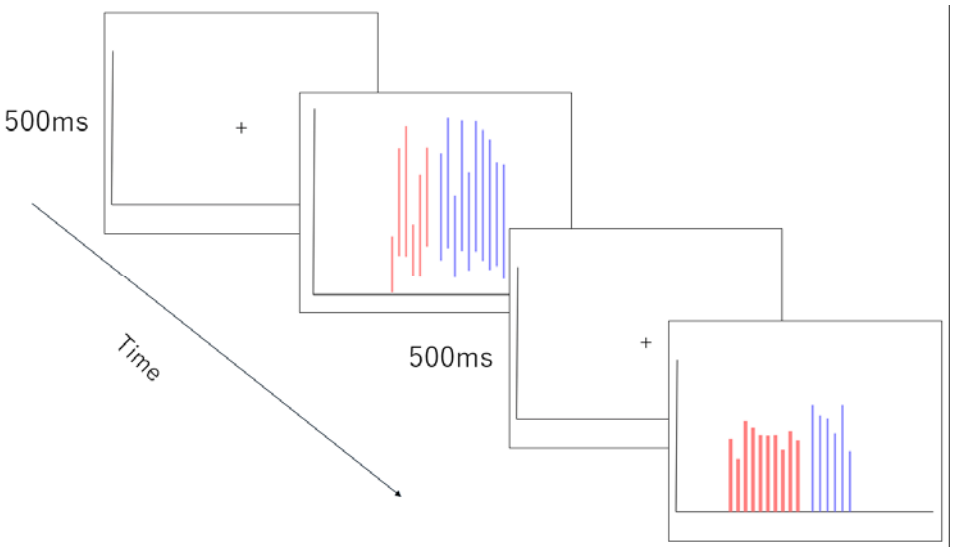

図 2　試行系列

参加者はモニターに表示された二つのグループのどちらの平均値が高いのかを判断し，左側に表示されたグループの平均値が高いと判断した場合は z キーを，右側に表示されたグループの平均値が高いと判断した場合はバックスラッシュキーを押して回答を行った。キーボードを押したときに正答であればそのまま次の試行に進み，誤答であれば誤答と表示された。z キーとバックスラッシュキー以外のキーを押した場合も誤答扱いとなった。グラフはキーを押されるまで画面の中央に表示された。その後，500 ms のインターバルの後，次の試行に進んだ。

実験では棒グラフ，不整列棒グラフ，点グラフの三つのグラフのブロックがランダム順に提示された。また，グラフの種類ごとに 6 vs. 6，10 vs. 10，6 vs. 10 の三つの比較の種類があった。これらの条件もブロック化してランダム順に提示された。High 条件，Low 条件，Even 条件は各ブロック内でランダム順に提示された。2 つのグループの平均値の違いは 60 ピクセルから始まり最低値を 1 ピクセルに設置した 3 アップ 1 ダウンの階段法で決められた。1 ブロック 30 試行とし，一人あたり 27 ブロックの合計 810 試行を行った（3 グラフタイプ×3 比較タイプ×3 つの太さの種類×30 試行＝810 試行）。1 ブロックごとに 5 秒の休憩表示が現れ，いずれかのキーを押すことで試行を再開した。

結　　果

正答率

比較タイプとグラフの太さ別に回答の正答率を算出した。図 3 は条件別の正答率を示したものである。

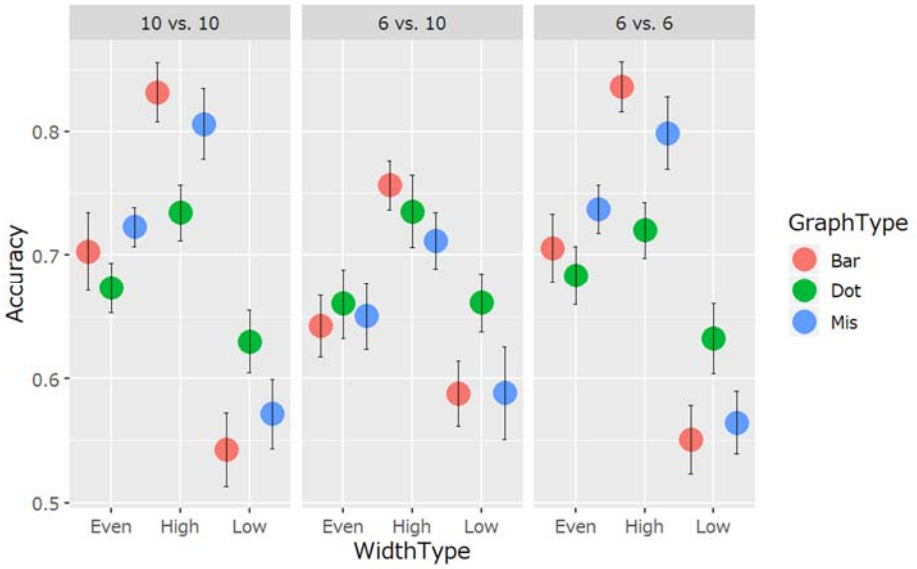

図 3　各条件における正答率の平均値（エラーバーは 95%信頼区間）

正答率におけるグラフタイプと比較タイプとグラフの太さの 3 要因参加者内計画の分散分析を行った。有意水準は 5%であった。グラフタイプの主効果は見られなかった（$F(2, 52) = 0.14$, $p = 0.87$）。比較タイプの主効果が見られた（$F(2, 52) = 11.34$, $p < 0.001$）。比較タイプについて Shaffer 法による多重比較を行った。多重比較では全体の有意水準が 5%になるように調整した。その結果，6 vs. 10 条件よりも 10 vs. 10 条件の方が正答率が高く（$t(26) = 3.76$），6 vs. 10 条件より 6 vs. 6 条件の方が正答率が高く（$t(26) = 3.67$），6 vs. 6 条件と 10 vs. 10 条件では違いが見られなかった（$t(26) = 0.36$）。グラフの太さ条件の主効果が見られた（$F(2, 52) = 84.29$, $p < 0.001$）。グラフの太さ条件について Shaffer 法による多重比較を行った。グラフの太さでは，Even よりも High の方が正答率が高く（$t(26) = 10.54$），Low よりも High の方が正答率が高く（$t(26) = 10.41$），Low よりも Even 条件の方が正答率が高かった（$t(26) = 6.53$）。

グラフタイプと比較タイプの交互作用が有意であった（$F(4, 104) = 4.33$, $p = 0.003$）。交互作用のパターンをくわしく検討するため，図 4 にグラフタイプにおいての比較タイプごとの正答率の平均を示した。

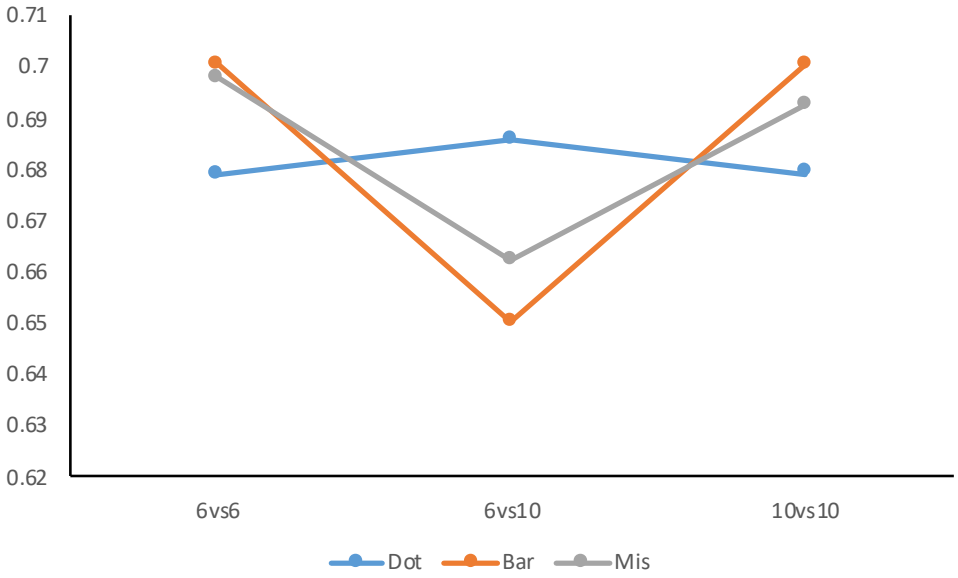

図 4　比較タイプごとのグラフタイプの正答率

グラフタイプ別に比較タイプの単純主効果を検定した結果，Dot 条件における比較タイプの単純主効果は有意ではなかった（$F(2, 52) = 0.29$, $p = 0.75$）。Bar 条件における比較タイプの単純主効果は有意であった（$F(2, 52) = 6.15$, $p = 0.004$）。Bar 条件にける比較タイプの単純主効果ついて Shaffer 法による多重比較を行った。その結果，6 vs. 10 条件よりも 6 vs. 6 条件の方が正答率が高く（$t(26) = 3.84$），6 vs. 10 条件よりも 10 vs. 10 条件の方が正答率が高く（$t(26) = 2.60$），6 vs. 6 条件と 10 vs. 10 条件では差が見られなかった（$t(26) = 0.45$）。Mis 条件における比較タイプの単純主効果は有意であった（$F(2, 52) = 10.89$, $p < 0.001$）。Mis 条件における比較タイプの単純主効果について Shaffer 法による多重比較を行った。その結果，6 vs. 10 条件よりも 10 vs. 10 条件の方が正答率が高く（$t(26) = 3.67$），6 vs. 10 条件よりも 6 vs. 6 条件の方が正答率が高く（$t(26) = 3.47$），6 vs. 6 条件と 10 vs. 10 条件では差が見られなかった（$t(26) = 0.0019$）。

グラフタイプとグラフの太さの交互作用が有意であった（$F(4, 104) = 13.14$, $p < 0.001$）。交互作用のパターンをくわしく検討するため，図 5 にグラフタイプにおいての文字の太さ条件ごとの正答率の平均を示した。

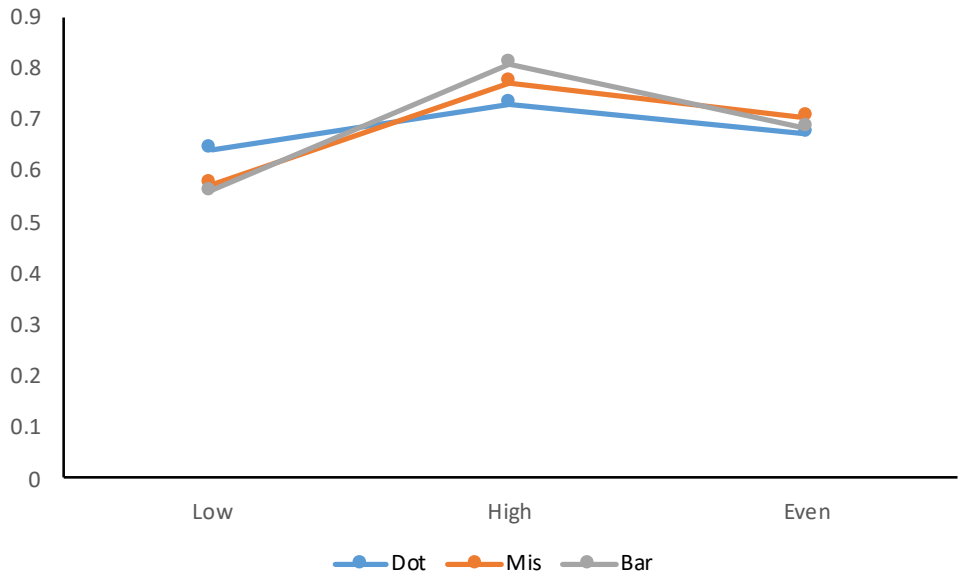

図 5　文字の太さごとのグラフタイプの正答率の平均

グラフタイプ別にグラフの太さの単純主効果を検定した結果，Dot 条件におけるグラフの太さの単純主効果は有意であった（$F(2, 52) = 11.36$, $p < 0.001$）。Dot 条件におけるグラフの太さの単純主効果ついて Shaffer 法による多重比較を行った。その結果，Low 条件よりも High 条件の方が正答率が高く（$t(26) = 3.99$），Even 条件よりも High 条件の方が正答率が高く（$t(26) = 3.78$），Low 条件と Even 条件では差が見られなかった（$t(26) = 1.69$）。Mis 条件におけるグラフの太さの単純主効果は有意であった（$F(2, 52) = 48.08$, $p < 0.001$）。Mis 条件にけるグラフの太さの単純主効果ついて Shaffer 法による多重比較を行った。その結果，Low 条件よりも High 条件の方が正答率が高かった（$t(26) = 8.04$）。Low 条件よりも Even 条件の方が正答率が高かった（$t(26) = 5.94$）。Even 条件よりも High 条件の方が正答率が高かった（$t(26) = 5.11$）。Bar 条件におけるグラフの太さの単純主効果は有意であった（$F(2, 52) = 90.17$, $p < 0.001$）。Bar 条件にけるグラフの太さの単純主効果ついて Shaffer 法による多重比較を行った。その結果，Low 条件よりも High 条件の方が正答率が高かった（$t(26) = 13.38$）。Low 条件よりも Even 条件の方が正答率が高かった（$t(26) = 5.62$）。Even 条件よりも High 条件の方が正答率が高かった（$t(26) = 8.87$）。

比較タイプとグラフの太さの交互作用が有意であった（$F(4, 104) = 8.65$, $p < 0.001$）。交互作用のパターンをくわしく検討するため，図 6 に比較タイプにおいての文字の太さ条件ごとの正答率の平均を示した。

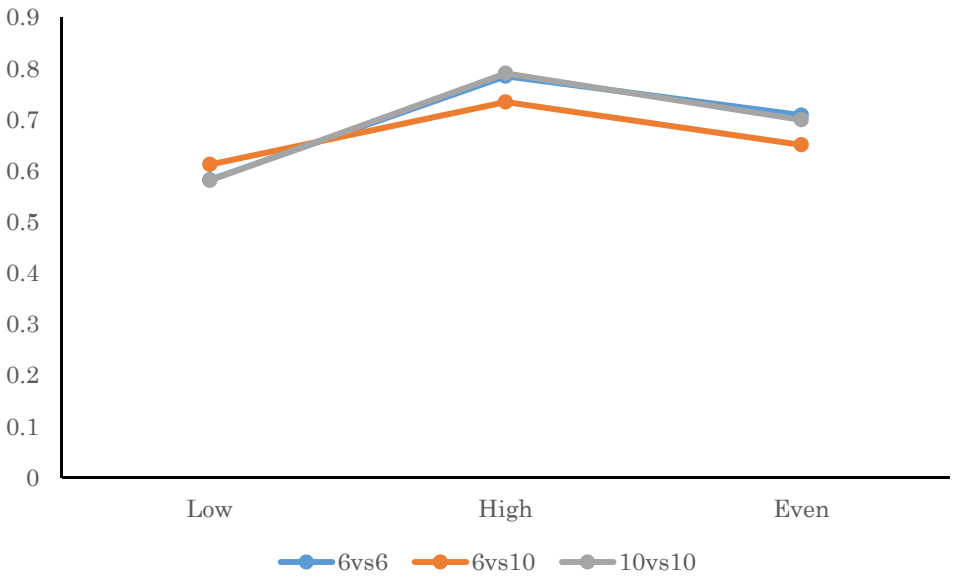

図 6　文字の太さごとの比較タイプの正答率の平均

比較タイプ別にグラフの太さの単純主効果を検定した結果，6 vs. 6 におけるグラフの太さの単純主効果は有意であった（$F(2, 52) = 69.51$, $p < 0.001$）。6 vs. 6 条件にけるグラフの太さの単純主効果ついて Shaffer 法による多重比較を行った。その結果，Low 条件よりも High 条件の方が正答率が高く（$t(26) = 10.99$），Low 条件よりも Even 条件の方が正答率が高かった（$t(26) = 7.53$）。Even 条件よりも High 条件の方が正答率が高かった（$t(26) = 4.54$）。比較タイプ別にグラフの太さの単純主効果を検定した結果，6 vs. 10 におけるグラフの太さの単純主効果は有意であった（$F(2, 52) = 22.19$, $p < 0.001$）。6 vs. 10 条件にけるグラフの太さの単純主効果ついて Shaffer 法による多重比較を行った。その結果，Low 条件よりも High 条件の方が正答率が高く（$t(26) = 5.69$），Even 条件よりも High 条件の方が正答率が高かった（$t(26) = 6.80$）。Low 条件と Even 条件では差は見られなかった（$t(26) = 1.83$）。比較タイプ別にグラフの太さの単純主効果を検定した結果，10 vs. 10 におけるグラフの太さの単純主効果は有意であった（$F(2, 52) = 79.05$, $p < 0.001$）。6 vs. 10 条件にけるグラフの太さの単純主効果ついて Shaffer 法による多重比較を行った。その結果，Low 条件よりも High 条件の方が正答率が高く（$t(26) = 10.47$），Low 条件よりも Even 条件の方が正答率が高かった（$t(26) = 6.65$）。Even 条件よりも High 条件の方が正答率が高かった（$t(26) = 8.33$）。

グラフタイプと比較タイプとグラフの太さの交互作用は有意でなかった（$F(8, 208) = 0.90$, $p = 0.51$）。

反応時間

モニターにグラフが表示されてから反応キーが押されるまでの反応時間の平均値を算出した。その際に誤答と，反応時間の平均値から 3 標準偏差以上離れている反応時間は外れ値

として除外し，3 標準偏差以内の正答の反応時間のみをデータに使用した。図 7 に条件別の平均反応時間を示した。反応時間におけるグラフタイプと比較タイプとグラフの太さの 3 要因参加者内計画の分散分析を行った。

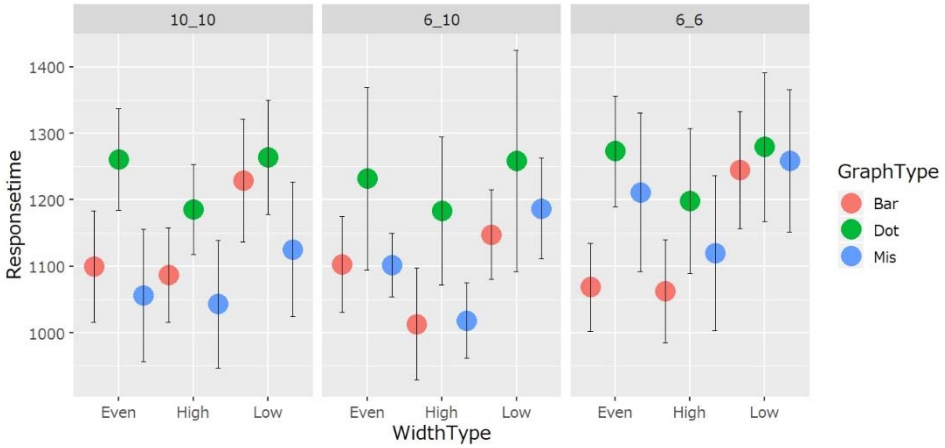

図 7　各条件における反応時間の平均値（エラーバーは 95%信頼区間）

グラフタイプの主効果は有意であった（$F(2, 52) = 3.24, p = 0.047$）。グラフタイプの主効果について Shaffer 法による多重比較を行った。Dot と Bar（$t(26) = 2.35$），Dot と Miss（$t(26) = 1.90$），Miss と Bar（$t(26) = 0.16$）のいずれの条件においても反応時間の差は有意ではなかった。比較タイプの主効果は見られなかった（$F(2, 52) = 0.43, p = 0.65$）。グラフの太さ条件の主効果が有意であった（$F(2, 52) = 31.14, p < 0.001$）。グラフの太さの主効果について Shaffer 法による多重比較を行った。その結果, High よりも Low の方が反応時間が長く（$t(26) = 7.47$），Even よりも Low の方が反応時間が長く（$t(26) = 4.45$），High よりも Even の方が反応時間が長かった（$t(26) = 3.67$）。

グラフタイプと比較タイプの交互作用（$F(4, 104) = 0.43, p = 0.79$）は有意ではなかった。グラフタイプとグラフの太さの交互作用が有意であった（$F(4, 104) = 2.53, p = 0.04$）。グラフタイプ別にグラフの太さの単純主効果を検定した結果，Dot におけるグラフの太さの単純主効果は有意であった（$F(2, 52) = 4.59, p = 0.01$）。Shaffer 法による多重比較を行った結果，High よりも Low の方が反応時間が長く（$t(26) = 2.67$），High よりも Even の方があ h ン脳時間が長く（$t(26) = 2.17$），Low とでは Even では差が見られなかった（$t(26) = 0.53$）。Mis におけるグラフの太さの単純主効果は有意であった（$F(2, 52) = 27.02, p < 0.001$）。Shaffer 法による多重比較を行った結果，High よりも Low の方が反応時間が長く（$t(26) = 7.68$），Even よりも Low の方が反応時間が長く（$t(26) = 4.90$），High とでは Even では差が見られなかった（$t(26) = 1.69$）。Bar におけるグラフの太さの単純主効果は有意であった（$F(2, 52) = 13.29, p < 0.001$）。Shaffer 法による多重比較を行った結果，High よりも Low の方が反応時間が長く（$t(26) =$

5.34），Even よりも Low の方が反応時間が長く（$t(26) = 2.54$），High よりも Even の方が反応時間が長かった（$t(26) = 2.53$）。比較タイプとグラフの太さの交互作用（$F(4, 104) = 0.35, p = 0.85$）は有意でなかった。また，グラフタイプと比較タイプとグラフの太さの交互作用は有意でなかった（$F(8, 208) = 0.92, p = 0.50$）。

JND

ブロックごとの JND の平均値を算出した。各ブロックの 16 試行から 30 試行間の 15 試行の平均値を算出することで JND を求めた。JND が低いほど弁別の精度は高いといえる。図 8 は比較タイプ別にグラフタイプの JND を表したものである。

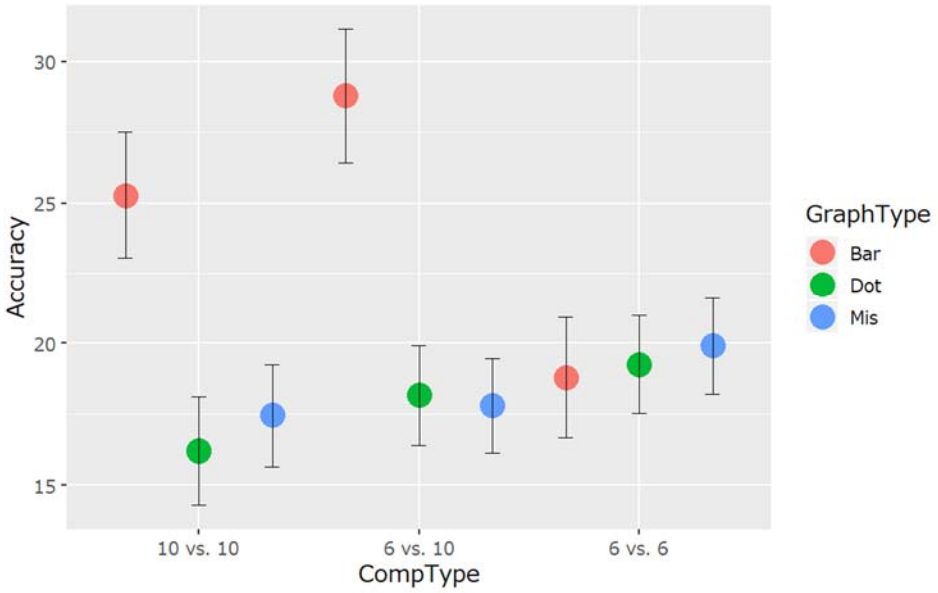

図 8　比較タイプ別のグラフタイプの JND（エラーバーは 95%信頼区間）

JND についてグラフタイプと比較タイプの 2 要因参加者内計画の分散分析を行った。その結果，グラフタイプの主効果は有意であった（$F(2, 52) = 18.52, p = 0.001$）。グラフタイプの主効果について Shaffer 法による多重比較を行った。Miss よりも Bar の方が JND が高く（$t(26) = 5.08$），Dot より Bar の方が JND が高く（$t(26) = 4.53$），Dot と Miss では差は見られなかった（$t(26) = 0.59$）。比較タイプの主効果は有意ではなかった（$F(2, 52) = 2.22, p = 0.12$）。グラフタイプと比較タイプの交互作用は有意であった（$F(4, 104) = 8.44, p < 0.001$）。Dot における比較タイプの単純主効果は有意でなく（$F(2, 52) = 1.76, p = 0.18$），Mis における比較タイプの単純主効果は有意でなかった（$F(2, 52) = 1.19, p = 0.31$）。Bar における比較タイプの単純主効果は有意であった（$F(2, 52) = 11.76, p < 0.001$）。Shaffer 法による多重比較を行った結果，6 vs. 6 条件よりも 6 vs. 10 条件の方が JND が高く（$t(26) = 4.83$），6 vs. 6 条件よりも 10 vs. 10 条件の方が JND が高く（$t(26) = 2.98$），6 vs. 10 条件と 10 vs. 10 条件では差が見

られなかった（$t(26) = 1.73$）。

考　察

本研究では，2 つのグループからなるグラフの太さを操作することによって平均値の判断に変化が生じるかを調べた。正答率を調べたところ，平均値が低いグループのグラフが太い Low 条件では，他の条件よりも正答率が低く反応時間は遅かった。平均値が高いグループのグラフが太い High 条件では，他の条件よりも正答率が高く反応時間は速かった。2 つのグループの棒や点の太さが同じであった Even 条件は，Low 条件よりも正答率は高く反応時間は速く，High 条件よりも正答率は低く反応時間は遅かった。これは，Low 条件では平均値が低いにもかかわらずグラフが太いことで，グラフの面積に誘導されて誤って選択したのに対して，High 条件では平均値が高いという判断が強化されて選択したことによると考えられる。この結果は High 条件が Even 条件よりも正答率が増加し Low 条件は Even 条件よりも正答率が減少するという予測と一致した。これらのことから，2 つのグループからなるグラフから平均値を読み取るとき，人は面積を代替として利用するという仮説は支持された。

Yuan et al.（2019）は，二つのグループからなるデータを棒グラフで表すときに棒の本数が異なる 6 vs. 10 条件において，棒の本数が多いグループの方が面積が大きくなるため平均値が高いと判断されると考察した。本研究では，6 vs. 6，10 vs. 10 条件でも同様に棒グラフの平均値を面積から判断していることがわかった。つまり，同数の棒グラフの比較のデータであっても，平均値を読み取る際に面積を手がかりにしているということである。

2 つのグループからなるグラフから平均値を読み取るときに面積の情報が用いられるという主張は，日常的な場面でのグラフ読み取りにおいて棒グラフから読み取られるデータの正確性に疑問を投げかける。本研究ではグラフの面積を操作することで正答率の差が生じた。しかし，一般には，異なるグループを表すためには太さよりも色の違いが用いられることが多い。グラフに用いる色を操作することによっても読み取りに違いは生じるのだろうか。ビジネス等でよく使われる Microsoft Excel でグラフを作成すると，デフォルトで寒色系の色と暖色系の色でグラフが作られる。もし寒色と暖色において読み取りの正確さに違いが生じるのであれば，多くの会議や講演などでグラフのデータ読み取りにおいてミスリーディングが生じ，誤った意思決定を導く可能性があるのかもしれない。

引用文献

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, **79**, 531-554.（Yuan et al., 2019 の引用による）

Heer, J., & Bostock, M. (2010). Crowdsourcing graphical perception: Using Mechanical Turk to assess visualization design. *Proceedings of the 28th Annual Chi Conference on Human Factors in Computing Systems* (pp. 203-212).（Yuan et al., 2019 の引用による）

Yuan, L., Haroz, S., & Franconeri, S. L. (2019). Perceptual proxies for extracting averages in data visualizations. *Psychonomic Bulletin & Review*, **26**, 669-676.