

2018.05.22

輪講 (福岡)

Perception of bar graphs - A biased impression?

Claudia Godau, Tom Vogelgesang & Robert Gaschler, (2016) *Computers in Human Behavior*, 59, 67-73

Introduction

- コンピュータを利用するとき、しばしばグラフに接触する
 - Ex. スプレッドシートはグラフを生成するために、簡単に利用できる
- グラフを読み取る能力と生成する能力は相関している (Davis, 2011)
 - 情報の視覚化は、知識の統合にとって、価値のあるツール (Burley, 2010)
 - グラフは、変数間の量的な関連に関する結果の素早い理解を可能にする
- 一方、グラフデザインは、提示された結果の解釈に影響を及ぼす (Fischer, Dewulf, & Hill, 2005; Huestegge & Philipp, 2011)
 - 知覚がグラフからデータの概要を引き出す際に、バイアスの影響がある可能性
- 本研究の目的は、棒グラフが偏った印象を与えるかどうかを検討すること

- グラフの種類のような特性についての研究は広範囲に渡る
 - 棒グラフは、円グラフと比較して、より早く正確に情報を読み取れる (Simkin & Hastie, 1987)
 - 垂直な棒グラフは、水平な棒グラフより、よりユーザーに身近である (Fischer et al., 2005)
 - 個々の価値については棒グラフが好ましい一方、グラフの傾向については線グラフが好ましい (cf. Zacks & Tversky, 1999)
 - ◇ しかし先行研究は、棒グラフが個々の価値についてのみに役立つということを示唆してはいない
 - ◇ 棒グラフは個々の価値を伝えるだけでなく、平均や条件の変動性など全体的な情報も伝える
- データをグラフで表す強みは、データの全体的な性質に関する簡易的な印象を、読み手が得られること
- 線グラフと棒グラフ、チャートを比較した先行研究 (Peebles, 2008)
 - 人々が、データがより高いか低いかを判断する場合、棒グラフでは過小評価されることを示した
- 本研究では2つの実験を通して、バイアスが、棒グラフに見られるかどうかを検討した
 - 実験1: 棒グラフにおける平均の過小評価について、異なる方法を用いて検討
 - 実験2: 平均の推測のバイアスに対する外れ値の影響を検討

Literature review

1. 棒グラフの平均の過小評価

2018.05.22

輪講 (福岡)

- 棒グラフは最も一般的な形式であり、メディアでもよく扱われる (Zacks, Levy, Tversky, & Schiano, 2002)
 - しかし、棒グラフは全ての意見において、最も良い提示ができるわけではない
 - ◇ バーの高さを判断する際の偏った印象は、度々報告されている (Jarvenpaa & Dickson, 1988; Kosslyn, 2006; Peebles, 2008; Zacks, Levy, Tversky, & Schiano, 1998)
- 一方、棒グラフの知覚を実験的に検討した結果は矛盾している
 - いくつかの条件では、人はバーの高さを過大評価する (Jarvenpaa & Dickson, 1988; Kosslyn, 2006; Zacks et al., 1998)
 - 別の条件では、過小評価することも示されている (Peebles, 2008; Zacks et al., 1998)
- 50本のバーの高さと、テストバーの高さを合わせる課題 (Zacks et al., 1998)
 - 参照するバーとテストバーの間の差が大きいとき、テストバーの高さはより過大評価された
 - ◇ 人には高いバーを過小評価する傾向がある一方、低いバーを過大評価する傾向があることを示した
 - ◇ 原因の一つとして、バーの高さに対する偏った印象の可能性がある (Jarvenpaa & Dickson, 1988; Kosslyn, 2006)
 - 実験1では、この可能性をフォローするために、バーの高さを操作した
- グラフから平均を推測する際のバイアスを検討するために、様々な方法を使用
 - 高さ (高いバー vs 低いバー)
 - 形式 (棒グラフ vs 点グラフ)
 - 評価 (グラフ的判断 vs 数値的判断)

2. 平均の推測のバイアスに対する外れ値の影響

- 実験2は、バーの高低の刺激を用いた実験1において検討された棒グラフの平均の過小評価を再検討し、拡張するために実施された
- 外れ値は2つのレベルでパフォーマンスに影響する可能性がある
 - 1つ目に、外れ値を理解することは、制御プロセスの必要性の認識を上げる (e.g., Brewer, 2012; Schützwohl, 1998)
 - ◇ 外れ値のあるグラフとないグラフでは、平均を推測する際のバイアスの量が異なるかもしれない
 - 2つ目に、高低の予測は、外れ値の刺激に関連して行われる
 - ◇ 外れ値の高さによって注意を引くことが考えられ、参加者の平均の推測は外れ値に向けて偏る
 - 高い外れ値は上向きの平均の推測を引き起こし、低い外れ値は下向きの平均の推測を引き起こすだろう
 - ◇ 一方で、平均の推測が外れ値から離れることも考えられる

2018.05.22

輪講 (福岡)

Methods

実験 1 : 棒グラフの平均の過小評価

- 参加者 : 53 名 ($M_{age} = 27, SD_{age} = 7.83$)
- 提示グラフ : 8 つのデータで構成した棒グラフ、もしくは点グラフ (Figure 1)
 - データはエクセルでランダムに作成
- 4 ブロック × ランダムに 80 試行実施
 - 全体で 320 個のグラフを提示
- 1~3 ブロックでは、赤い水平線と共にグラフを提示
 - 赤い水平線 (基準線) : データの平均を判断する基準として使用
 - 課題は、正しい平均が、基準線より高いか低いかを判断し、素早くボタンを押すこと
 - ◇ ブロック 1 と 2 では、基準線は正確に平均値を示した
 - もし参加者にグラフから推測した平均のバイアスがあるのなら、結果は五分五分から逸脱するだろう
- ブロック 1 : バーの間の変動性は一定に保っていたが、高いバーと低いバーを使用
 - 低いバーより高いバーで、より平均を過小評価するバイアスを示すかどうか検討
 - 高いバー (8~12 の範囲) のグラフ 40 個と、低いバー (1~5 の範囲) のグラフ 40 個を提示
- ブロック 2 : 同一のデータで、棒グラフと点グラフを使用
 - バイアスが、棒グラフだけに起こるのか、点グラフでも見られるのか検討
 - 高いバーの棒グラフ 40 個と、点グラフ 40 個を提示 (8~12 の範囲)
- ブロック 3 : 基準線と平均の距離を変更
 - 線は平均の周囲 $\pm 1.5SD$ で引かれた
 - 高いバー (8~12 の範囲) の棒グラフが 80 個提示された
- ブロック 4 : 結果が線に依存することを避けるため、平均を数値で提示
 - 課題は、正しい平均が、提示された数値より高いか低いかを判断し、素早くボタンを押すこと
 - 縦軸にスケールを記入したグラフを使用
 - グラフが消えた後、小数点第 2 位までの数値 (平均) を提示
 - 高いバー (8~12 の範囲) の棒グラフが 80 個提示された
- ブロック 1~3 では、グラフは 3 秒提示され、グラフが消えた後、ボタンを押すことを求める文が提示され、参加者の反応用に 2 秒間与えられた
- ブロック 4 では、グラフは 8 秒提示され、グラフが消えた後、平均が数値で提示され、反応用に 7 秒間与えられた
 - ボタンを押した後、2 秒後に次の試行の教示が始まった
 - フィードバックは与えなかった

実験 2 : 平均の推測のバイアスに対する外れ値の影響

2018.05.22

輪講 (福岡)

- 参加者 : 大学生 80 名 ($M_{age} = 24, SD_{age} = 5.35$)
- 水平線の書かれた棒グラフを提示され、正しい平均が線より高いか低いかを判断する課題
 - 参加者が反応する時間として、2 秒間与えられた
- 提示グラフ : 11 本のバーで構成された棒グラフ
 - 10 本のバーは、1SD に等しくなるようランダムに決定
 - 11 本目のバーの高さが $\pm 3SD / \pm 1SD$ 異なった
 - ◇ +3SD 条件 (30 試行) : 11 本目のバーが、10 本の平均から +3SD 離れるように設定
 - ◇ -3SD 条件 (30 試行) : 11 本目のバーが、10 本の平均から -3SD 離れるように設定
 - ◇ +1SD 条件 / -1SD 条件 (各 20 試行) : 統制条件として用いられた
 - ◇ 11 本目のバーはランダムな位置に挿入された
 - 棒グラフの底辺と黄色の水平線 (基準線) の間の距離は、全ての試行で一定

Result

実験 1 : 棒グラフの平均の過小評価

ー 反応時間が 100ms 以下の場合には分析から除外

ー 1,2,4 ブロックにおける反応の割合を Figure 2 に示す

- ブロック 1 : バーの高低におけるバイアス
 - 「より低い」と反応した割合とチャンスレベル (50%) を比較する t 検定を実施
 - ◇ 参加者は、高低どちらのバーにおいても平均を有意に過小評価した
 - 高いバー : $t(52) = 6.38, p < .001$
 - 低いバー : $t(52) = 6.12, p < .001$
 - ◇ 低いバー条件と高いバー条件で「より低い」と反応した割合に差はなかった ($t(52) = .75, p > .46$)
 - 大部分の参加者に影響が現れたかどうかを検討するために、参加者の 2 変数バイアス得点を使用
 - ◇ 二項分布によれば、参加者が 40 試行中 26 試行 (62.5%) 以上「より低い」と反応した場合、チャンスレベル (50%) と有意な差が見られた
 - 高いバー条件における参加者の 64.15%
 - 低いバー条件における参加者の 62.26% (Figure 2 の 青い点)
- ブロック 2 : 棒グラフと点グラフにおけるバイアス
 - 棒グラフ条件 : 参加者は平均を有意に過小評価した ($t(52) = 9.13, p < .001$)
 - 点グラフ条件 : 参加者は平均を有意に過大評価した ($t(52) = -4.58, p < .001$)
 - ◇ 2 つの条件間の差は有意 ($t(51) = 9.59, p = .001$)
 - 過小評価効果はグラフの種類に依存する
 - ◇ ただし、後日、点グラフのみを提示し実験したとき、点グラフにおける過大評価が観察されなかったことは、注意すべき
- ブロック 3 : 平均からの距離におけるバイアス

- 基準線は、正しい平均の周囲 1.5SD 以内で変化した
- 平均から離れた距離における「より低い」という反応の割合分布 (Figure 3)
 - ◇ もし平均の推測にバイアスがないなら、基準線が正しい平均と重なっているとき、「より低い」と反応した割合は 50%程度のはず
 - ◇ しかし、「より低い」という反応の分布が 50%程度となるのは、0SD から左へシフトした
 - シグモイド関数 ($R^2 = .976$) によると、正しい平均より .278SD 低い
- 平均の過小評価を相殺するには、基準線は、正しい平均から約.28SD 低く提示される必要がある
- ブロック 4 : 数値に対する判断をしたときのバイアス
 - 基準は数値で提示されたが、参加者は平均を過小評価した
 - ◇ 「より低い」と反応した割合は、50%より有意に高かった ($t(51) = 2.92, p = .01$)
- 反応時間 (Figure 4)
 - 2つの反応 (より高い/より低い) 間に有意な差はなかった ($F < 1$)

実験 2 : 平均の推測のバイアスに対する外れ値の影響

—参加者が 2 秒間の反応画面内に反応しなかった試行 (6.65%) を除外

- 外れ値の大小に関わらず、全ての条件で「より低い」反応は 60%以上 (Figure 5)
 - 全て、50%から有意な差があった ($t(79) = 5.32 \sim 10.4, ps < .001$)
- 平均の推測は、外れ値から反れるより、外れ値に向かうように偏った
 - 参加者はしばしば、他のバーの平均と外れ値の差が小さいとき (-1SD) より、外れ値のバーが短いとき (-3SD) のとき、「より低い」と答えた
 - 強い正の外れ値は、弱い外れ値より、「より高い」という反応を及ぼした
- 「より低い」と反応した割合を従属変数とした 2 要因分散分析を行った
 - 独立変数 : 外れ値の強さ (3SD vs 1SD) × 外れ値の距離 (- vs +)
 - 外れ値の距離の主効果 ($F(1, 79) = 4.62, p < .035$) と交互作用 ($F(1, 79) = 12.065, p < .001$) が有意
 - ◇ 外れ値の強さの主効果は有意ではなかった ($F < 1$)
- 反応時間は、外れ値が制御プロセスを駆動させることを示唆しなかった (Figure 5)
 - 参加者は-3SD の外れ値で最も早く反応した
 - 正の外れ値は、遅い反応を引き出した
 - ◇ 予想外なことは、「より低い」反応に優位な弱い刺激 (-1SD) のとき、反応時間が遅くなったこと
- 反応時間を従属変数とした 2 要因分散分析を行った
 - 外れ値の距離と外れ値の強さの交互作用が有意だった ($F(1, 79) = 7.3, p = .008$)
 - それぞれの主効果は有意ではなかった ($Fs < 1.47$)

Discussion

実験 1 : 棒グラフの平均の過小評価

- 実験 1 の結果は、人々が棒グラフで示されたデータの平均を過小評価することを示唆
- 近年、グラフはコンピュータによって簡単に生成できる
 - グラフを用いることは、学校で教えられる (Ainley, 1994)
 - 棒グラフは、情報を伝えるためのツールとして、インターネットの様々なところにある (Roth, Kolojechick, Mattis, & Goldstein, 1994)
 - ◇ 日常的にコンピュータを利用することから、棒グラフを見たとき平均を過小評価する問題の関心は増加するかもしれない
- なお、過小評価の問題は、電子画面によって起きるわけではない
 - バイアスが、人とコンピュータの相互作用の状況に限定される可能性を除外するため、実験 1 ブロック 2 の課題について、質問紙を用いて、38 名の学生に調査
 - ◇ その結果、棒グラフは平均の過小評価を導いた
- 点グラフに見られた平均の過大評価について
 - おそらく、棒グラフと点グラフを一緒に提示したことによる影響
 - ◇ 点グラフのみを提示した追実験では、平均の過大評価は見られなかった
 - 原因の一つとして、棒グラフの過小評価に対する逆の影響があった可能性
 - ◇ 例えば、全体的に反応のバランスを取るため、参加者は棒グラフに対して「より低い」と答えたとき、点グラフで「より高い」と答えたのかもしれない

実験 2 : 平均の推測のバイアスに対する外れ値の影響

- 外れ値を含んでも、棒グラフにおける平均の過小評価のバイアスは再確認された
 - 極端に高いバー (+3SD の外れ値) を含んだ棒グラフでさえ、「より低い」という反応は現れた
 - ◇ 平均の推測は外れ値に対して偏ったにも関わらず

Conclusion

- グラフは全体の平均や変動性のような情報を早く伝えることができる
 - しかし、本研究の結果は、頻繁に使用されるグラフの種類から平均を推測することに対するバイアスを示唆した
 - ◇ 棒グラフにおける平均は、実際よりも、低く見積もられる
 - 棒グラフは傾向の偏った印象を与える
 - 本研究の結果は、棒グラフにおいて、平均と比較して、個々の価値を過小評価した結果と一致した (Peebles, 2008)
 - While Zacks et al. (1998) は、バーを高くすることで、過小評価の程度が増えることを示したが、本研究でその影響は観察されなかった

- 棒グラフにおける平均の過小評価の影響は大きいように思われる
 - 「平均はより低い／高い」反応の転換点は、正しい平均より.28SD程度低かった
 - 平均を数値で示したり、平均を上回る外れ値を提示したりしたときでさえ、バイアスは現れた
- バーの強さは過小評価に影響しなかったため、他の原因の可能性を考える必要がある
 - Peebles (2008) は、棒グラフと点グラフにおける分布の空間的注意の差による影響かもしれないことを示唆する
 - ◇ 人は、点グラフでは、点と x 軸間の空間より点に関心を向ける一方で、棒グラフでは、バーの頂点よりむしろバーの中間に注目する可能性
 - 今後の課題としては、グラフ理解の間の眼球運動を操作することで、この可能性を検証すること
- グラフに対する注意の配分は、学習によって影響を与えることができる
 - 注意は、グラフリテラシーに依存する (Okan, Galesic, & Garcia-Retamero, 2015)
 - ◇ グラフリテラシーの低い人々は、紛らわしい空間的概念のマッピングにしばしば時間を割いた
 - 一方で、グラフリテラシーの高い人々は、正確な解釈のために必要な情報を含む部分を見ることに、より時間を割いた
 - ノービスは顕著な知覚的特徴に焦点を置く (e.g. color; Ali & Peebles, 2013; Lowe, 1999; 2003)
 - ◇ 彼らの注意を引くダイナミックな視覚化をすることで、グラフ理解を促進することができる (Ploetzner, Lippitsch, Galmbacher, Heuer, & Scherrer, 2009)
- 実践的観点から考えるなら、どんな種類のグラフをいつ使うかを知る必要がある
 - 棒グラフは、条件全体の平均のような情報に関連するとき、良い選択ではない一方で、他においては役に立つかもしれない
 - ◇ 例えば、棒グラフは、参加者の解釈のエラーが少ないので、線グラフと比較して、3つの変数間の関連を示すためには好ましい (Peebles & Ali, 2009; Peebles, 2011)
 - 単純で頻繁に使用されるグラフ形式における知覚的バイアスに焦点を置くことは、より複雑な形式の知覚的特性を学ぶための基礎として、役に立つだろう

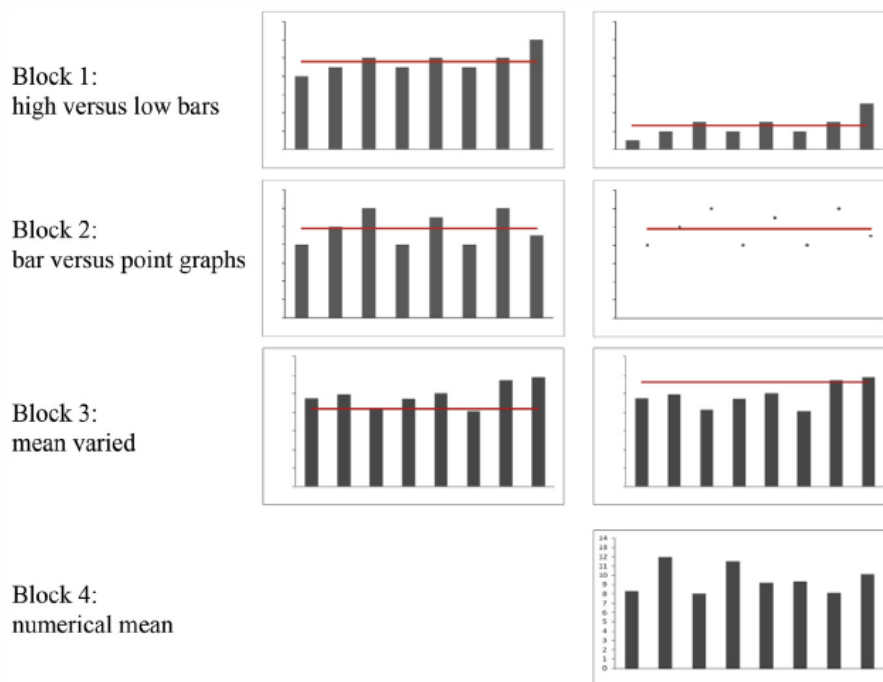


Fig. 1. Conditions and examples of the presented graphs for all blocks of the first experiment – underestimation of the mean in bar graphs.

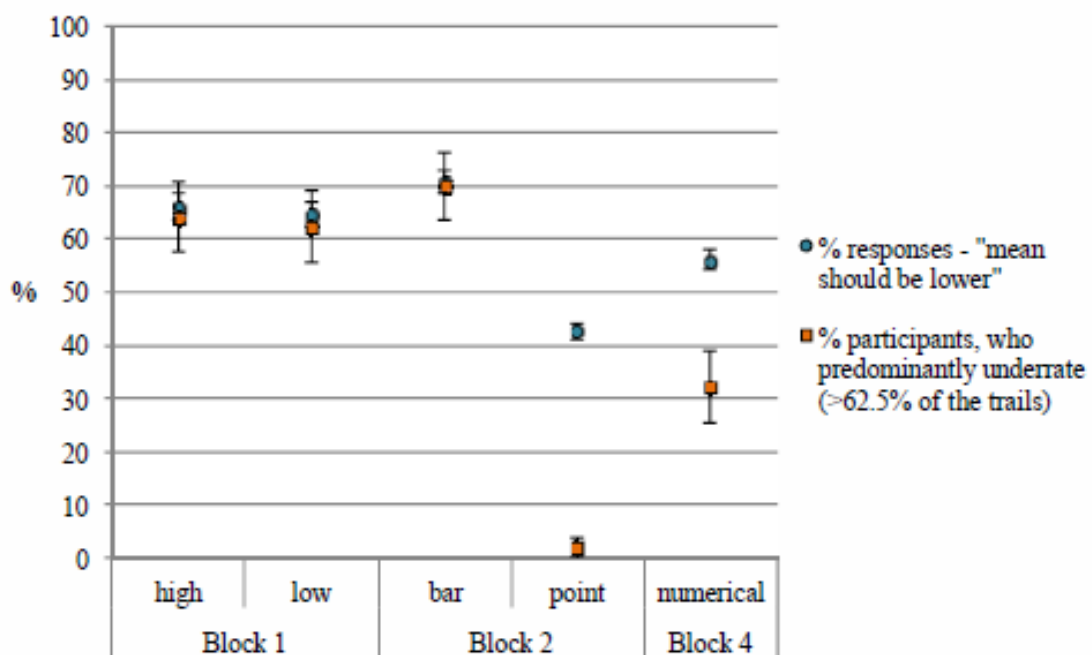


Fig. 2. The percentage of participants, who predominantly underrate the mean (>62.5% of all trails), is presented in orange squares, and the percentage of "mean should be lower" responses is presented in blue points for the Blocks 1, 2, 4 of the first experiment – underestimation of the mean in bar graphs. The error bars show the standard error of the mean. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

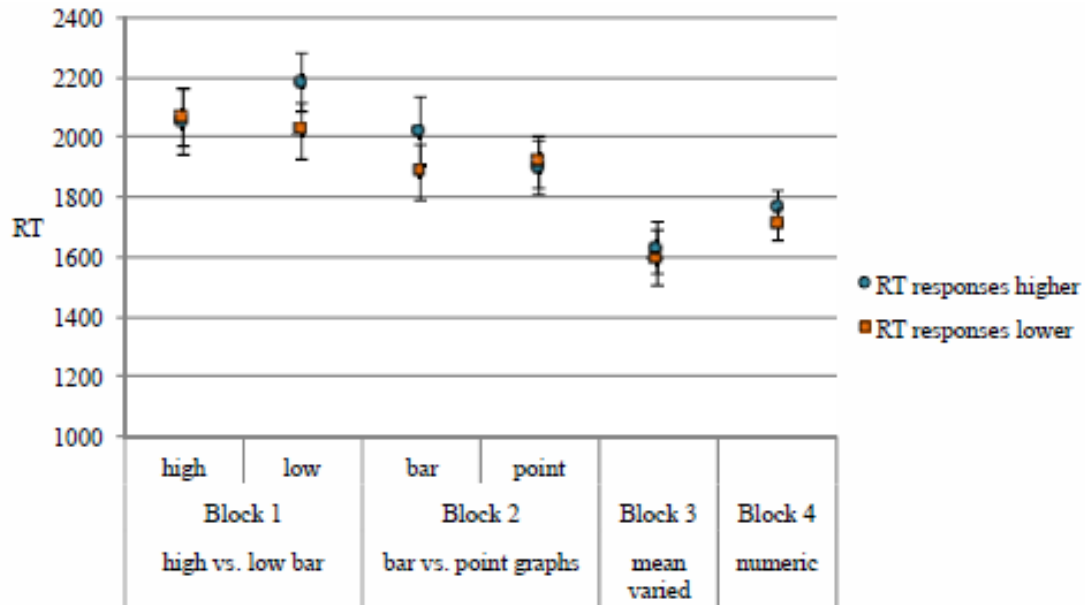


Fig. 4. Reaction times in msec. split by “mean should be lower/higher” responses for Block 1, 2, 3 and 4 of the first experiment – underestimation of the mean in bar graphs.

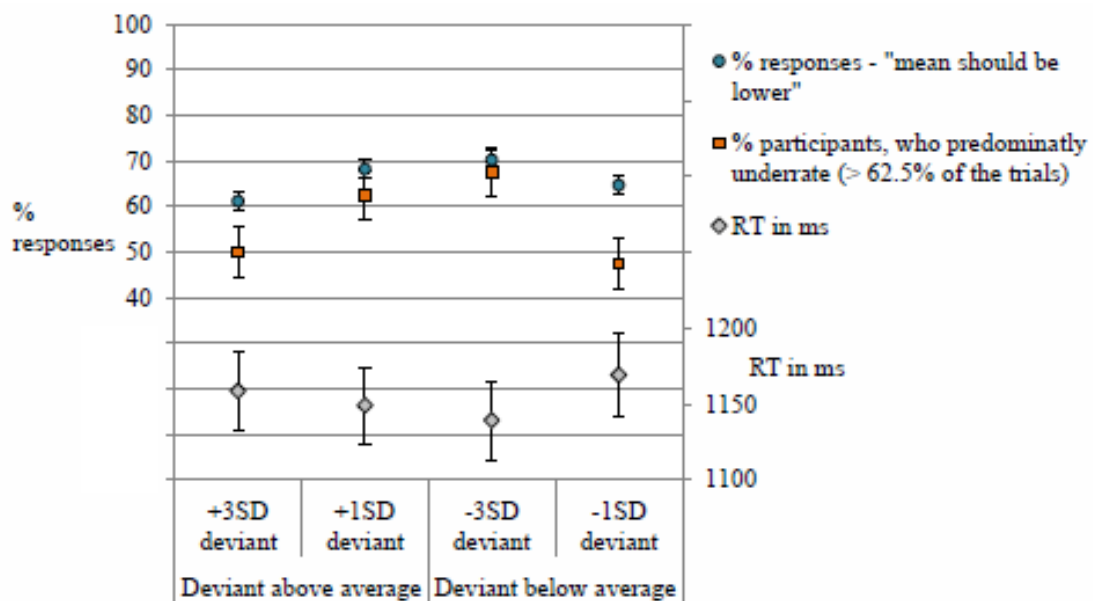


Fig. 5. The percentage of participants, who predominantly underrate the mean (>62.5% of all trails), is presented in orange squares, and the percentage of “mean should be lower” responses is presented in blue points (left y-axis) of the second experiment – effect of outliers on the bias in average estimation. Gray diamonds show RTs in ms (right y-axis). The error bars show the standard error of the mean. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 3. The percentage of "mean should be lower" responses is presented per deviation of the reference line from the mean in standard deviations (Block 3 data of the first experiment – underestimation of the mean in bar graphs).

