

Influencing Message Reception via Anchoring

Andrew Wright

Brett Brandom

Michelle Cheung

2024-12-15

Abstract

This project examines how the anchoring-and-adjustment heuristic influences business decision-making, specifically in interpreting a company's quarterly sales performance. Participants were presented with internal sales forecasts (anchors) followed by actual results. Three groups received different anchors: a control group, a low anchor (\$10M forecast vs. \$15M actual), and a high anchor (\$20M forecast vs. \$15M actual). We measured their perceptions of whether Q4 sales were successful after showing them either an upward or downward trending chart of the past 4 quarters of sales. Results showed that the high anchors biased respondents toward less favorable assessments, while the low anchors led to more positive evaluations. These effects persisted regardless of sales trends but were stronger with negative news. This suggests that anchors shape how people interpret ambiguous outcomes, with implications for business and behavioral economics.

Background

The anchoring heuristic is the disproportionate influence on decision makers to make judgements that are biased towards an initially presented value. While anchoring had been studied previously, the anchoring-and-adjustment heuristic was introduced by Tversky and Kahneman (1974)¹, and this is the effect we will be considering here. Since the introduction by Tversky and Kahneman, the anchoring effect has been observed in multiple domains. Table 1 in Furham and Boo² lists several research articles demonstrating the effect. The literature indicates that in decision making, higher ambiguity, lower familiarity, relevance or personal involvement with the problem, a more trustworthy source or more plausible bid/estimate all lead to stronger anchoring effects. Researchers measured the concept by splitting subjects into two groups, exposing them to a low or high anchor value according to group, asking a question with a specific numerical answer, and comparing the responses of the two groups.

Research Question

This study seeks to explore the influence of anchoring on decision making in a business context. Specifically, we aim to answer the following research question: How does exposure to an initial sales forecast (anchor) influence individuals' judgments of a company's sales performance when the actual results differ from the forecast, and does the direction of the sales trend (upward vs downward) moderate this effect?

Hypothesis

We hypothesized that mentioning an expected outcome could bias the audience members' interpretation of the results so that their conclusions are more likely to align with the initial anchor. Specifically, we

¹<https://www.science.org/doi/10.1126/science.185.4157.1124>

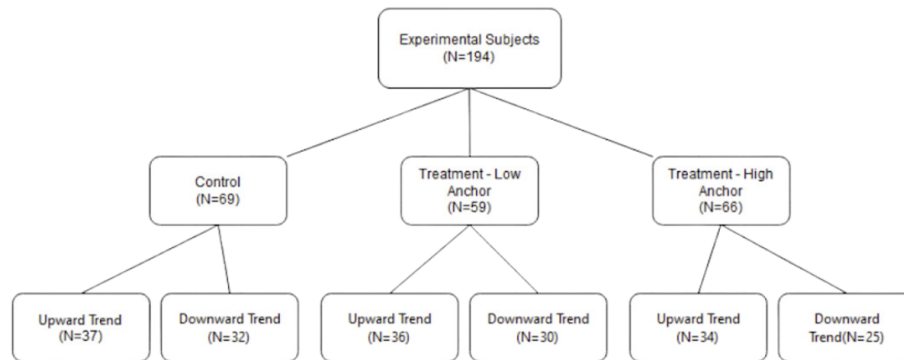
²<https://www.sciencedirect.com/science/article/pii/S10533535710001411>

hypothesize that exposure to low sales forecast (\$10 million) will bias participants toward a more favorable evaluation of the company’s Q4 sales performance, while exposure to a high sales forecast (\$20 million) will bias participants toward a less favorable evaluation.

Experiment Design

Experiment Overview

To test this hypothesis, we conducted a between-subjects experiment. In the experiment, participants were presented with a hypothetical business scenario. Below is a description of each treatment and outcome question, as well as a waterfall diagram of the number of subjects who received each question.



Control: “Suppose you work for a fictional retail company. The next page will present you with a scenario and ask you to interpret it.”

Treatment 1 - Low Anchor: “Suppose you work for a fictional retail company. Internal forecasts predicted Q4 sales of the company you work at would be \$10 million, and actual sales exceeded the forecast with a total of \$15 million. The next page will present you with a scenario and ask you to interpret it.”

Treatment 2 - High Anchor: “Suppose you work for a fictional retail company. Internal forecasts predicted Q4 sales of the company you work at would be \$20 million, and actual sales failed to meet expectations with a total of \$15 million. The next page will present you with a scenario and ask you to interpret it.”

Outcome 1 - Upward Trend [Appendix Fig 1.2]: “The chart above shows historical sales data for your company for the past four quarters in millions of dollars. Based on the information you’ve been given about your company, would you say that its Q4 sales were successful or not?”

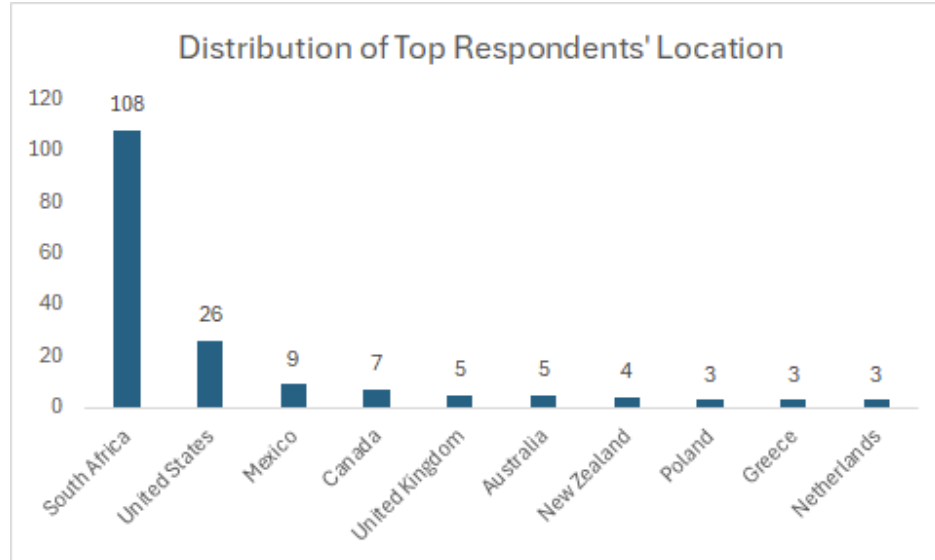
Outcome 2 - Downward Trend [Appendix Fig 1.1]: “The chart above shows historical sales data for your company for the past four quarters in millions of dollars. Based on the information you’ve been given about your company, would you say that its Q4 sales were successful or not?”

Communication Tooling

Subjects were found on Prolific, a service used to pay participants for taking surveys or completing other tasks. We offered participants \$1 for completing the survey, which was above Prolific guidelines for the amount to offer based on the length of the survey. Participants were located globally with no restrictions on who could participate other than the language of the survey being in English. The survey itself was hosted on Qualtrics, with a link provided to the survey for each Prolific user.

Enrollment & Recruiting Process

We requested a total of 200 respondents, who were asked to enter their unique Prolific id at the beginning of the survey. A total of 194 out of these 200 entered their id, with the other 6 not entering a valid id and therefore not being eligible to complete the survey. Out of those who did enter a valid id, all completed the survey in full and received payment, meaning that there was no attrition in the study. The overall distribution of location among respondents was just over half from South Africa, ~20% from North America, and the rest split between Europe, Oceania, Asia, and other countries in Africa. We suspect that there may be a large proportion of Prolific users from South Africa due to the favorable exchange rate from USD to the South African Rand. Approximately 60% of respondents were female versus 40% male.



Randomization Process

We randomized subjects by assigning them uniformly first to either of the three treatments, and then to either of the two outcomes. Randomization was done within the Qualtrics platform, utilizing random selection at both the treatment stage and at the outcome stage. This resulted in 36% of subjects assigned to the control, 30% to the low anchor treatment, and 34% to the high anchor treatment. 45% were assigned to the downward trend outcome, and 55% were assigned to the upward trend outcome. We used a chi-square test to confirm that subjects were randomly assigned to each block with equal probability ($p=0.7$).

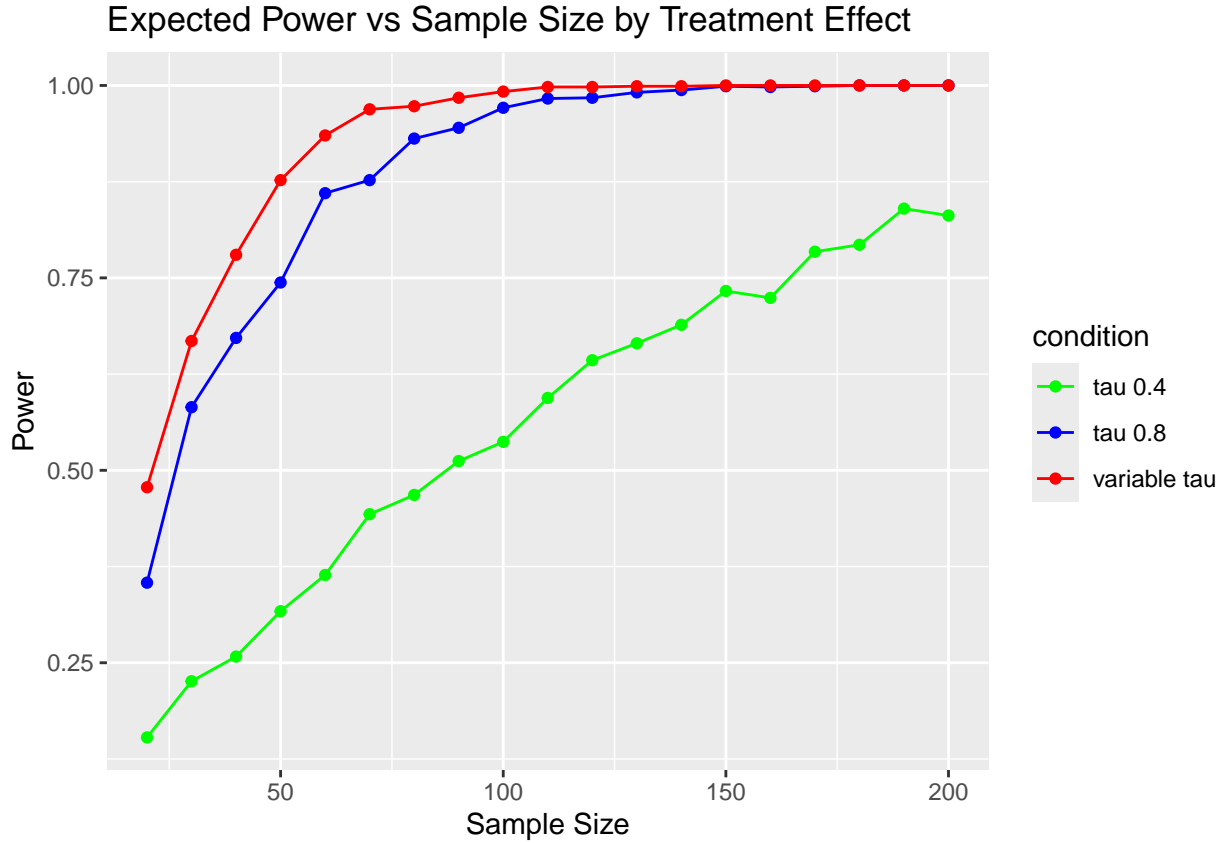
Observation and Outcome Measures Results

Our experiment is a between-subjects design. The only difference between subjects in either of the two treatment groups and the control is their exposure to their assigned treatment scenario instead of the control scenario. This means that differences in outcomes between these groups can be attributed to a causal effect of the treatment. The specific outcome being measured is a “Yes” or “No” response to one of the two outcome questions mentioned above for each subject. Potential outcomes therefore would be which of these two responses an individual would have given if they had received the corresponding treatment or control scenario. Below is a ROXO diagram outlining the three paths from randomization to outcome measurement. Note that there was no pretreatment measurement performed in this study.

R	X	O
Control	Control scenario	Binary outcome measured
Treatment - Low	Low anchor scenario	Binary outcome measured
Treatment - High	High anchor scenario	Binary outcome measured

Power Calculation

Given pre-experiment assumptions about effect size and experiment size, the plot below shows that power increases with sample size in all three scenarios. Interestingly, the variable treatment effect ($\tau \sim N(0.8, 0.4)$) reaches sufficient power (0.8 or higher) slightly quicker than the constant moderate treatment effect ($\tau = 0.8$), though both are quite similar. In contrast, the scenario with the smallest effect size ($\tau = 0.4$) requires a much larger sample size to achieve the same power, emphasizing the challenge of detecting weaker effects. We expect that as long as τ is close to the 0.8 level here, even with a heterogeneous treatment effect, that our sample size will be sufficient to detect a statistically significant effect.



Results

Linear Regression

To measure the effectiveness of our anchoring, we used an ordinary least squares regression. We encoded our ternary treatment variable as a factor with baseline value of control (absorbed into the constant of our model), and high and low anchor treatments appearing explicitly in the regression. Similarly, we encoded our binary chart type variable so that downward is the baseline. And we encoded our outcome variable to be 1 if

the subject said “Yes” to the question posed to them, and 0 if they said “No.” We note that we considered including age, gender, and country of residence as additional covariates, but we found that they did not increase the precision of our coefficient estimates and so we excluded them.

As a sanity check, we first regressed our outcome against the chart type variable:

$$\text{response} = \theta \cdot \text{chart-type-up}$$

We find that on average, when the chart is going up, subjects are more likely to respond positively, as we expect.

We next regressed our outcome against anchor and chart type:

$$\text{response} = \theta \cdot \text{chart-type-up} + \delta_1 \cdot \text{anchor-high} + \delta_2 \cdot \text{anchor-low}$$

Recall that a low anchor means subjects were told sales exceeded forecasts, and a high anchor means sales did not meet forecasts. We find $\hat{\delta}_1 = -0.106$ and $\hat{\delta}_2 = 0.144$, indicating that priming the audience with an expected outcome biased their favorability of the quarterly results, in line with our expectation from the anchoring hypothesis.

To measure heterogeneous treatments effects dependent on the chart type, we fit an additional model with an interaction term between anchor and chart type:

$$\begin{aligned} \text{response} = & \theta \cdot \text{chart-type-up} + \delta_1 \cdot \text{anchor-high} + \delta_2 \cdot \text{anchor-low} \\ & + \gamma_1 \cdot \text{chart-type-up} \times \text{anchor-high} + \gamma_2 \cdot \text{chart-type-up} \times \text{anchor-low} \end{aligned}$$

The signs of the anchors make sense, and the coefficients for anchor low, chart type upward, and their interaction are all significant. The negative sign of the interaction coefficient suggests some kind of a saturation effect, suggesting that anchoring might help best when making bad news seem less bad, but that good news is not made even better. However the coefficient for anchor high and its interaction with chart type are both not statistically significant. We wouldn’t expect anchoring to be significant in one direction but not the other.

```

##
## Regression Results for the Hypothetical Business Scenario Experiment
## =====
##                               Dependent variable:
##                               -----
##                               Perceived Success of Q4 Sales
##                               (1)                (2)                (3)
## -----
## AnchorLow                      0.144**          0.346***
##                               (0.057)          (0.116)
##
## AnchorHigh                    -0.106*          -0.027
##                               (0.059)          (0.102)
##
## ChartUp                       0.636***          0.640***          0.813***
##                               (0.055)          (0.052)          (0.070)
##
## AnchorLow:ChartUp                                -0.374***
##                                                    (0.119)
##
## AnchorHigh:ChartUp                                -0.149
##                                                    (0.122)
##
## Constant                      0.299***          0.280***          0.187***
##                               (0.049)          (0.053)          (0.070)
## -----
## Observations                   194                194                194
## R2                             0.439                0.484                0.511
## Adjusted R2                    0.436                0.476                0.498
## Residual Std. Error    0.359 (df = 192)    0.346 (df = 190)    0.339 (df = 188)
## F Statistic      150.309*** (df = 1; 192)  59.371*** (df = 3; 190)  39.226*** (df = 5; 188)
## =====
## Note:                                     *p<0.1; **p<0.05; ***p<0.01
##                               Regression results from a 3x2 factorial experiment.

```

Logistic Regression

Because we are dealing with binary response variables, linear regression may not be an appropriate model, and thus our previous models may be misspecified. We used logistic regression with the same model formulae to see if this yielded clearer results. Writign p for the probability of a yes response, we fit the following three models:

$$\log\left(\frac{p}{1-p}\right) = \theta \cdot \text{chart-type-up},$$

$$\log\left(\frac{p}{1-p}\right) = \theta \cdot \text{chart-type-up} + \delta_1 \cdot \text{anchor-high} + \delta_2 \cdot \text{anchor-low},$$

$$\begin{aligned} \log\left(\frac{p}{1-p}\right) = & \theta \cdot \text{chart-type-up} + \delta_1 \cdot \text{anchor-high} + \delta_2 \cdot \text{anchor-low} \\ & + \gamma_1 \cdot \text{chart-type-up} \times \text{anchor-high} + \gamma_2 \cdot \text{chart-type-up} \times \text{anchor-low}. \end{aligned}$$

Here, we find all interaction terms are highly insignificant, suggesting our third model is inappropriate and there are no heterogeneous treatments effects (when using a log odds ratio instead of ATE to measure causal effect). Interpreting our second model without interactions, we find an odds ratio of 3.3 for the effect of the low anchor and 0.34 for the high anchor.

```
##
## Logistic regression Results for the Hypothetical Business Scenario Experiment
## =====
##                               Dependent variable:
##                               -----
##                               Perceived Success of Q4 Sales
##                               (1)           (2)           (3)
## -----
## AnchorLow                        1.195**      1.600***
##                               (0.511)      (0.582)
##
## AnchorHigh                      -1.059*      -0.192
##                               (0.582)      (0.709)
##
## ChartUp                        3.512***      4.046***      20.032
##                               (0.456)      (0.547)      (1,072.315)
##
## AnchorLow:ChartUp                                -16.611
##                                                    (1,072.316)
##
## AnchorHigh:ChartUp                               -16.834
##                                                    (1,072.316)
##
## Constant                       -0.853***      -1.119***      -1.466***
##                               (0.234)      (0.381)      (0.453)
##
## -----
## Observations                    194           194           194
## Log Likelihood                  -78.914        -70.615        -67.575
## Akaike Inf. Crit.              161.829        149.229        147.151
## =====
## Note:                               *p<0.1; **p<0.05; ***p<0.01
##                               Regression results from a 3x2 factorial experiment.
```

Conclusion

Based on the results of our study, we found that when presenting quantitative results, audience members' perception of the results can be swayed by priming them with an anchor value before presenting the results. In this specific case, we found that historical sales results are interpreted more (less) favorably when audience members are first told that sales exceeded (did not exceed) forecasts. These results can be used by presenters to garner more favorable responses by comparing their numbers to some initial forecast first.

Limitations and Future Improvements

One major limitation in our study is its generalizability. While we have demonstrated an anchoring effect in this specific situation, and previous literature has demonstrated it more generally, it is not apparent what will result in the "best" anchoring for a specific condition, or under what circumstances such an effect will be more prevalent compared to others.

An implicit assumption in our experiment is that an assessment of the performance of a company's quarterly sales ought to be independent of the accuracy of any sales forecasts. We tried to choose an anchor scenario that would make sense in the context of a presentation, rather than something as divorced from reality as, say, spinning a roulette wheel as has been done in early studies. With that said it could be argued that a reasonable person would take sales exceeding forecasts to mean good quarterly results. Variations of the experiment with seemingly more disconnected anchors could further elucidate how well anchoring works in a presentation scenario. Future improvements to this study could make this distinction more clear in the setup. Finally, a larger sample size may have allowed us to better understand if there are real interaction effects between the treatments and different outcome questions in this study.

Appendix



Figure 1: Fig 1.1 - Downward trend outcome survey question



The chart above shows historical sales data for your company for the past four quarters in millions of dollars. Based on the information you've been given about your company, would you say that its Q4 sales were successful or not?

☐ Yes

☐ No



Figure 2: Fig 1.2 - Upward trend outcome survey question

Country of Respondant	Number of Respondants
South Africa	108
United States	26
Mexico	9
Canada	7
United Kingdom	5
Australia	5
New Zealand	4
Poland	3
Greece	3
Netherlands	3
Spain	3
Germany	2
Kenya	2
Chile	2
India	2
South Korea	2
Croatia	1
Malaysia	1
Austria	1
Slovakia	1
France	1
Hungary	1
Sweden	1
Portugal	1
Japan	1
Italy	1
Philippines	1
Slovenia	1

Figure 3: Fig 2.1 - Location of respondents

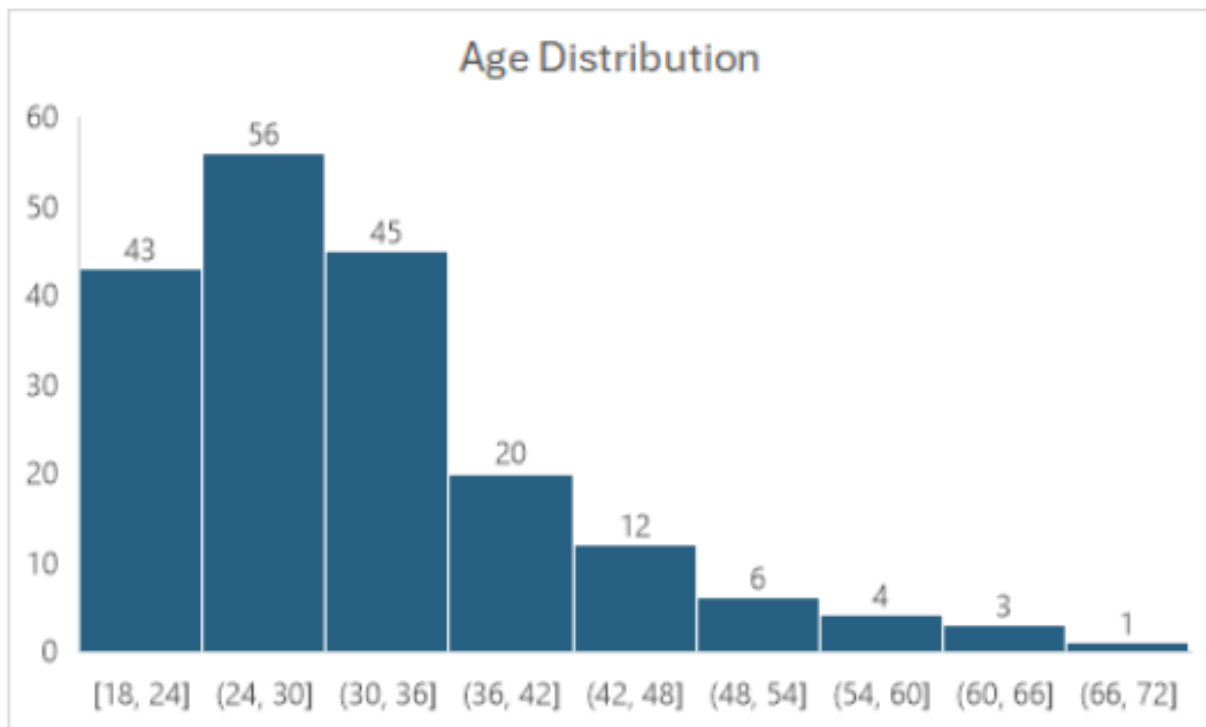


Figure 4: Fig 2.2 - Age of respondents