Pre-analysis plan:
Anchoring and Subjective Belief Distributions

Håkan Holm, Margaret Samahita, Roel van Veldhuizen, Erik Wengström

May 10, 2023

1) Purpose and Motivation

The purpose of this project is to study how anchors affect estimations of subjective belief distributions (SBDs). Anchoring is a well-known judgment bias in decisions. Although the impact of anchors has been studied extensively on estimations of single-number summary statistics, its impact on higher moments of SBDs is to a large extent unexplored. This makes it valuable to study since SBDs play an important role in economic theory.

Our overall aim is to investigate how anchoring affects SBDs. More specifically, we will study: 1) if elicitations of averages through SBDs are less affected by anchors than direct elicitations of means and 2) how anchors affect the 2nd (and, of secondary importance, 3rd) moment of SBDs. In addition, we will investigate how individual characteristics can explain heterogeneities in the properties of SBDs and how these impact the anchoring of SBDs.

In the presentation of our pre-analysis plan, we have followed the suggestion of moderation in Duflo et al. (2020) when it comes to detail. While we have tried to give sufficient detail to the planned study for it to generate testable hypotheses, some details of the parts of secondary importance have not been specified and may be left as an exploratory analysis.

2) Design

We conduct an online experimental survey on Prolific. In the online survey, we randomly allocate participants to five treatment conditions where we elicit beliefs in various ways to be described in more detail below. After the belief elicitations, all subjects answer survey questions about their individual characteristics.

2.1) Online Survey/Experiment

We start by giving all subjects information about central concepts in the survey such as the average and the frequency distribution of a set of values. We also let subjects answer a few control questions that are designed to check that the subjects understand the concepts.

The next step is to inform subjects that we have collected information about a given price distribution. It will be a distribution of historical prices for a one-night hotel room in the city of Rome. This distribution was selected since we want the subject to have some idea about the price distribution but not too much information about it.

Subjects are then randomly allocated to one of the following five treatments, where subjects' beliefs are elicited with monetary incentives.

- *Control elicitation of Distribution (CD).* We elicit the SBDs of the hotel room prices using the procedure suggested by Crosetto and de Haan (2022). There is no anchoring in this treatment.

- *Low anchor elicitation of Mean (LM).* The subjects are first asked if they believe that the average price is lower or higher than the value of a low anchor. After that subjects are asked to guess the average price of the hotel room.

- *Low anchor elicitation of Distribution (LD).* The subjects are first asked if they believe that the average price is lower or higher than the value of a low anchor. After that, we elicit the SBDs of the hotel room prices by the same technique as in *CD*.

- *High anchor elicitation of Mean (HM).* The subjects are first asked if they believe that the average price is lower or higher than the value of a high anchor. After that subjects are asked to guess the average price of the hotel room.

- *High anchor elicitation of Distribution (HD).* The subjects are first asked if they believe that the average price is lower or higher than the value of a high anchor. After that, we elicit the SBDs of the hotel room prices by the same technique as in *CD*.

We then run a 2nd elicitation so that subjects who received the *LM* and *HM* will receive *CD* and subjects who received *CD*, *LD*, and *HD* will be asked to guess the average hotel prices. (Hence, they receive the same treatment as in *LM* and *HM* but without any anchor, which will be denoted as *M*.) In connection with the 2nd elicitation, we will ask how certain the subjects are about their estimations following the elicitation of cognitive uncertainty (CU) by Enke and Graeber (2022). The 2nd round elicitations are of secondary importance and will only be used in the exploratory analysis.

After the treatment, all subjects answer questions about demographics, cognitive reflection, investment behavior, and financial literacy.

Demographic questions:
1) How old are you? ___________ (years)
2) What is your gender: ______(man/woman/non-binary/other(please specify))
3) What is the highest education level you have reached?
1. Elementary School
2. High school graduate
3. Some College
4. Associate Degree
5. Bachelor's Degree
6. Master's Degree
7. Doctorate Degree

Questions designed to measure cognitive reflection inspired by Frederick (2005)
4) A house contains a living room and a kitchen that are perfectly square. The living room has four times the area of the kitchen. If the walls of the kitchen are four meters long, how long are the walls of the living room?
5) Yesterday a store owner reduced the price of a pair of $100 shoes by 10 percent. This morning he reduced the price further by 10 percent. How much does the pair of shoes cost now?
6) If it takes 4 machines 4 minutes to make 4 forks, how many minutes would it take 80 machines to produce 80 forks?
7) In a lake there is a patch of lily pads. Every day it doubles in size. If it takes 100 days for the lily pads to cover the entire lake. How long (in days) does it take for the lily pads to cover half the lake?
8) A meal and a drink cost $11 in total. The meal costs $10 more than the drink. How much does the meal cost?

Questions concerning risk-taking, investment behavior, and financial literacy (risk question is taken from the German Socio-Economic Panel, and the financial literacy questions are taken from Lusardi and Mitchell, 2014)
8) How willing are you to take risks, in general?
[Respondents rate their willingness on a scale from 0 to 10.]
9) Do you own stocks in any form: a) in equity funds? ______(yes/no) b) individual stocks in specific companies? _____ (yes/no)
10) Suppose you had $100 in a savings account and the interest rate was 2 percent per year. After 5 years, how much do you think you would have in the account if you left the money to grow:
Alternatives: [more than $102; exactly $102; less than $102; do not know]
11) Imagine that the interest rate on your savings account was 1 percent per year and inflation was 2 percent per year. After 1 year, would you be able to buy:
Alternatives: [more than, exactly the same as, or less than today with the money in this account; do not know.]
12) Do you think that the following statement is true or false? "Buying a single company stock usually provides a safer return than a stock mutual fund."
Alternatives : [true; false; do not know]

The overall design is summarized in the table below:

Table of Design

Tasks / Number of Subjects	*150*	*150*	*150*	*150*	*150*
General information	X	X	X	X	X
Information about means/distributions and control questions	X	X	X	X	X
1st Elicitation	CD	LM	LD	HM	HD
2nd Elicitation	M	CD	M	CD	M
Cognitive uncertainty question	X	X	X	X	X
Questionnaire on demographics, Cognitive reflection, Risk-taking, Financial literacy	X	X	X	X	X

X- subjects will be doing the task. 150- number of subjects in each treatment (motivated by power calculations specified in section 4.2 below).

3) Analysis

To explain our analysis we start by defining some variables. All variables refer to the first elicitation round if not otherwise explicitly stated:

Type of anchor: *L-Low, H-High, C-Control (no anchor)*
Elicitation of: *M-Mean; D-SBD*
Treatment $X \in \{CD, LM, LD, HM, HD\}$
N = number of subjects in Treatment *X*
TrueMean = objective mean of the elicited variable's true distribution.
$Match_i(X)$ = The score indicating how close *i*'s SBD is to the true distribution in treatment *X*. The maximum and minimum score is 100 and 0, respectively.
$Mean_i(X)$ = Elicited mean for individual *i* (directly in M-treatments and indirectly in D-treatments) in Treatment *X*
$\overline{Mean}(X) = \frac{\sum_i^N Mean_i(X)}{N}$ = group mean
$\sigma_i(X)$ = standard deviation of individual SBD in Treatment *X*
$CV_i(X) = \frac{\sigma_i(X)}{Mean_i(X)}$ = Coefficient of variation of *i*'s SBD
$\overline{CV}(X) = \frac{\sum_i^N CV_i(X)}{N}$ = group mean of $CV_i(X)$

$Skew_i(X)$ = the Fisher's moment coefficient of skewness of *i*'s SBD
$\overline{Skew}(X) = \frac{\sum_i^N Skew_i(X)}{N}$ = mean skewness of group receiving treatment *X*.
$CU_i(X) \in \{0, 5, 10, \dots, 100\}$ = Cognitive uncertainty in Treatment *X* of individual *i*.
$\overline{CU}(X) \frac{\sum_i^N CU_i(X)}{N}$ = group mean of cognitive uncertainty
Age – age of respondent
Gender – gender of the respondent
Education – highest educational level of respondent
CRT – number of correct answers on questions 4-7.
Risk – willingness to take risks
IF – investment in funds (dummy)
IS – Investment in individual stocks (dummy)
FL – score on financial literacy questions 10-12

Our main hypotheses are as follows:

H1A: $\overline{Mean}(LM) < \overline{Mean}(HM)$, and H1B: $\overline{Mean}(LD) < \overline{Mean}(HD)$
Hypothesis 1 is about the anchoring effect per se. H1A will seek to confirm that the standard anchoring effect on means also exists in our data. In H1B, based on the documented robustness of the anchoring effect (e.g., Furnham and Boo, 2011), we conjecture that anchors affect elicitations of SBD in a directionally similar way. To test these claims statistically we use a one-sided two-sample t-test in each case. We then run an OLS regression with $Mean_i(X)$ as the dependent variable with a dummy for the high-anchor treatment (HD or HM) and controlling for Age, Gender, and Education.[1] For this regression and all other regressions we will use robust standard errors unless otherwise specified.

H2: $\overline{Mean}(HM) - \overline{Mean}(LM) > \overline{Mean}(HD)$- $\overline{Mean}(LD)$
Hypothesis 2 postulates that the anchoring effect is smaller when eliciting SBDs. This is based on previous research suggesting that the anchoring effect may be driven by heuristics such as the anchoring-and-adjustment process, which require cognitive effort (see e.g., Epley and Gilovich,2006; Simmons et al, 2010). If the elicitation of the SBD requires more cognitive effort than the mean, as seems likely, this may leave fewer cognitive resources for the anchoring-and-adjustment heuristic and hence less scope for an anchoring effect. In addition, compared to the mean, the SBD is less compatible with the anchor, which is thought to lower the scope for the anchoring effect (e.g., Strack and Mussweiler, 1997; Li et al., 2021). We test Hypothesis 2 using a one-sided difference-in-difference test, that is, an OLS regression of $Mean_i(X)$ on a dummy for the two high-anchor treatments (HD, HM), a dummy for the two D-treatments (HD, LD), and their interaction (the main variable of interest). We also run a second regression controlling for Age, Gender, and Education.

H3A: $\overline{CV}(CD) = \overline{CV}(LD)$, and H3B $\overline{CV}(CD) = \overline{CV}(HD)$
Hypothesis 3 is about the coefficient of variation of the elicited distribution. A popular explanation for the anchoring effect in means is the anchoring-and-adjustment heuristic (Tversky and Kahneman, 1974). This heuristic implies that the final estimate will be a linear combination between the anchor (*a*) and the prior belief (*x*) (that is, the estimate the participant would give without an anchor). There exist several potential ways of extending this

[1] We will control for age using the following categorical dummies: 18-24, 25-34, 35-44, 45-54, 55-64, 65 and above. For gender, we will include a dummy for men and a dummy capturing other/non-binary individuals, provided these are present in our data. For education, we will include dummy variables for each of the 7 categories, using elementary school as the reference category.

heuristic to the case of SBDs. Assume that the subjects have a prior SBD and that they form a distribution of beliefs over the potential anchors (of which they observe one realization). We can then model the final estimates as a linear combination of two random variables (*a* and *x*). Alternatively, the final distribution may be viewed as a weighted mixture distribution of the two distributions.

If we consider the first approach and study the distribution of a linear combination of two (independent) random variables, the variance of the final estimate will depend on the variances of the anchor distribution and the prior SBD, as well as the weighting parameter. If the variance of the anchor is equal to or smaller than the variance of the initial belief, the variance of the final estimate will be lower than the variance of the prior belief distribution. Correspondingly, the coefficient of variation of the final belief distribution will be lower than the prior SBD, at least for the situation in which the mean of the anchor distribution is greater than the mean of the prior SBD. However, it is possible to imagine that the effect goes in the other direction. For example, if the variance of the anchor is much larger than the variance of the initial beliefs, and the adjustment towards the anchor is strong.

If we instead consider the second approach and model the final distribution as a weighted mixture distribution of the distributions of prior beliefs and anchors, the effect is again ambiguous. The coefficient of variation of the final estimate will depend on the variances and means of the anchor and the initial belief, as well as the degree of adjustment towards the anchor.

As a result, the overall prediction is ambiguous: the coefficient of variation may both decrease or increase if the converse holds. We will test this hypothesis using two two-sided two-sample t-tests and then run two OLS regressions for subjects receiving the relevant treatment (either HD or LD, with CD as the control group) with $CV_i(X)$ as the dependent variable, and with a dummy for the relevant treatments (HD or LD) and controlling for Age, Gender, and Education.

Hypotheses of secondary importance and secondary analysis:

H4A: $\overline{Skew}(LD) = \overline{Skew}(CD)$. H4B: $\overline{Skew}(CD) = \overline{Skew}(HD)$.
It could be interesting to look at higher moments as well. Lacking an a-priori hypothesis, we will use two two-sample two-sided t-tests to test whether the skewness in the respective anchor treatments differs from the skewness in the control treatment. We will also run an OLS regression for subjects with the skewness coefficient as the dependent variable and with the dummy variables D_{HD} and D_{CD} for treatments HD and LD while controlling for Age, Gender, and Education.

Analysis of factors affecting $CV_i(X)$
We will study how various factors correlate with the individual coefficient of variation. We expect that $CV_i(X)$ is positively correlated with $CU_i(X)$ for all *X.* The intuition for this is that an individual who is certain about a variable should have a more concentrated SBD than an individual who is uncertain and therefore finds it difficult to exclude many values as plausible. We expect *Risk* (taking) (and IF, IS) to be negatively correlated with $CV_i(X)$ since a high variation in the SBD can be seen as “insurance” against getting everything wrong and thereby not receiving any payment at all in the elicitation. We also conjecture that cognitive reflection (*CRT*) and financial literacy (*FL*) may negatively affect $CV_i(X)$ (see e.g., Bergman et al., 2010). We study this by an OLS where $CV_i(X)$ is the dependent variable and $CU_i(X)$, *Risk*, *CRT*, *IF*, *IS*, and *FL* as independent variables. We control for treatment dummies in addition to the standard demographic controls (Age, Gender, Education).

Heterogeneous treatment effects for H1-H3

We will also explore heterogeneous treatment effects for H1-H3 using Age, Gender, Education, Risk, CRT, FL, CU, IF and IS. For H1, we will regress *Mean*$_i$(*X*) on a dummy for the high anchor treatment (HM or HD), the relevant heterogeneous effect variable, and the interaction of the two. For H2, we regress *Mean*$_i$(*X*) on a dummy for the two high-anchor treatments (HD, HM), a dummy for the two D-treatments (HD, LD), and their interaction, as well as the heterogeneous treatment effect variable and its interaction with all three previously listed variables. For H3, we use two OLS regressions on *CV*$_i$(*X*) relevant treatment (either HD or LD, with CD as the control group) with *CV*$_i$(*X*) as the dependent variable, and the dummy for the anchor treatment (either HD or LD), the heterogeneous effect variable and their interaction as independent variables.

Analysis of factors affecting $Match_i(X)$

We will study how various factors affect how good subjects are in estimating the true distribution. We expect that Education, cognitive reflection (CRT), and financial literacy (FL) are all positively correlated with $Match_i(X)$. In addition, we expect that $CU_i(X)$ and treatments including anchors will negatively affect $Match_i(X)$.

We study this by an OLS using subjects in CD, LD, and HD where $Match_i(X)$ is the dependent variable and *Education, CRT, FL,* $CU_i(X)$, and a treatment dummy for anchor treatments (LD, HD) are independent variables. We also include Age and Gender as controls.

In a second specification, we explore heterogeneous interaction effects between the treatment dummies with the independent variables mentioned above. For instance, it is plausible that anchors impact high-CRT subjects less than low-CRT subjects.

4) Data collection and sample size

4.1) Data collection

The experimental and survey data are collected on the online platform Prolific. We aim for a sample as close as possible to a representative one of the population in the USA that Prolific can provide in the age group over 18. We aim to collect data from 750 subjects with 150 in each of the five treatment groups (based on the 1st elicitation), based on the power calculations presented in section 4.2.

We will conduct our analysis using the following samples:
i) Full sample of all participants who finished the full survey.
ii) The full sample from (i) minus participants who completed the study in 5 minutes or less.
iii) The sample from (ii) minus outlier responses: those who state a mean price of 50 USD or less (in the elicitation of the mean) or those who state a mass of 5% or more for the lowest or highest bin (that is, the 0-49 USD and 750-799 USD intervals in the elicitation of SBD).

4.2) Power Calculation

Our power calculation is based on a study by Lee and Morewedge (2022) that used a relatively similar design as we do in terms of domain and anchors. They asked subjects in different treatment groups to estimate their willingness to pay in USD for 4-star hotel rooms, not in Rome (as we will do) but in Miami providing them with either no anchor, a low anchor, or a high anchor. The mean prices and standard deviations (in parentheses) were 197 (100), 147 (81), and 330 (176) for the no anchor, low anchor, and high anchor treatments, respectively. The low anchor in this study was $44 and the high anchor was $610. Our low and high anchors are $134 and $546, which are based on the low and high boundaries of the true underlying distribution of hotel prices. Our anchors are slightly less dispersed, which may lead to less extreme results in both anchor treatments. For the purposes of our power calculation, we therefore assume mean prices (standard deviations) of 180 (100) and 300 (160) in the low-anchor and high-anchor treatments, where the standard deviations are adjusted in proportion to the change in means. We further assume that the anchoring effect will be 50% smaller in the

SBD treatments. In our pilot (N=143 across all treatments) we found somewhat larger effects, but we stick with these initial estimates to keep our power estimates conservative.

We compute power for hypothesis 1 using “power twomeans” in Stata for a one-sided two-sample t-test, one for (LM,HM) and one for (LD,HD), for a power of 0.80. For hypothesis 2, we simulate normally distributed data and look for the minimum sample size required to have a significant difference-in-difference coefficient in 80% of simulated samples. We perform these power calculations both for an uncorrected significance threshold of $p<0.05$ and for the Benjamini-Hochberg correction for multiple testing.[2] Since we do not have a directional hypothesis for H3, we do not include this hypothesis in our analysis, but do take it into account in our adjustment for multiple testing.

The Table below presents the results, which show that a sample size of 150 participants per treatment would be sufficient to have a power of 0.80 to observe a significant difference-in-difference result even after applying the multiple testing correction (and a higher power for the other tests). As a result, we aim for a sample size of 150 participants per treatment.

	Standard	Benjamini-Hochberg correction
H1a: $\overline{Mean}(LM) < \overline{Mean}(HM)$	17	27
H1b: $\overline{Mean}(LD) < \overline{Mean}(HD)$	62	85
H2: difference-in-difference	130	150

References

Bergman, O., Ellingsen, T., Johannesson, M., & Svensson, C. (2010). Anchoring and cognitive ability. *Economics Letters*, *107*(1), 66-68.

Chapman, G.B., Johnson, E.J., 1999. Anchoring, activation, and the construction of values. Organizational Behavior and Human Decision Processes 79, 1–39.

Crosetto, P., & de Haan, T. (2022). Comparing input interfaces to elicit belief distributions, HAL Open science (halshs-03816349).

Duflo, E., A. Banerjee, A.Finkelstein, L. F. Katz, B. A. Olken, and A. Sautmann. (2020) In Praise of Moderation: Suggestions for the Scope and Use of Pre-analysis Plans for RCTs in Economics. NBER Working Paper 26993

Enke,B., & Graeber, T. (2022). Cognitive uncertainty, wp, 2022.

Epley, N., & Gilovich, T. 2006. The anchoring and adjustment heuristic: Why adjustments are insufficient. Psychological Science, 17: 311-318.

Frederick, S. W., & Mochon, D. (2012). A scale distortion theory of anchoring. Journal of Experimental Psychology: General, 141, 124–133.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The journal of socio-economics*, *40*(1), 35-42.

Jacowitz K. E. and K Kahneman D. (1995) Measures of anchoring in estimation tasks, Personality and Social Psychology Bulletin 21, 1161–1166.

Lee, C. Y., & Morewedge, C. K. (2022). Noise Increases Anchoring Effects. *Psychological Science*, 33(1), 60-75.

Li, L., Maniadis, Z., & Sedikides, C. (2021). Anchoring in Economics: A Meta-Analysis of Studies on Willingness-To-Pay and Willingness-To-Accept. Journal of Behavioral and

[2] This correction ranks all *n* hypothesis from the lowest p-value to highest and then multiplies the significance threshold for the lowest p-value by (1/*n*), the second-lowest p-value by (2/*n*), etc. In our case, this implies multiplying the thresholds for H1a, H1b and H2 by 1/5, 2/5 and 3/5 respectively, given a total of five hypotheses including H3a and H3b.

Experimental Economics, 90, 101629.
Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. Journal of Economic Literature, 52(1), 5–44.
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, 136–164.
Simmons, J. P., LeBoeuf, R. A., & Nelson, L. D. (2010). The effect of accuracy motivation on anchoring and adjustment: Do people adjust from provided anchors? Journal of Personality and Social Psychology, 99, 917–932.
Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. Journal of Personality and Social Psychology, 73(3), 437–446.
Wegener, D.T., Petty, R.E., Blankenship, K.L., Detweiler-Bedell, B., (2010). Elaboration and numerical anchoring: implications of attitude theories for consumer judgment and decision making. Journal of Consumer Psychology 20, 5–16.