// NO.23-072 | 12/2023

DISCUSSION PAPER

// CHRISTOPH KARL BECKER, PETER DUERSCH,
THOMAS EIFE, AND ALEXANDER GLAS

Households' Probabilistic Inflation Expectations in High-Inflation Regimes

Leibniz Association

ZEW

Households' probabilistic inflation expectations in high-inflation regimes*

Christoph Karl Becker[†] Peter Duersch[‡] Thomas Eife[§] Alexander Glas[¶]

December 21, 2023

Abstract

Central bank surveys frequently elicit households' probabilistic beliefs about future inflation. However, most household surveys use a response scale that is tailored towards low-inflation regimes. Using data from a randomized controlled trial included in the Bundesbank Online Panel Households, we show (i) that keeping the original scale in high-inflation regimes distorts estimates of histogram moments and forces households to provide probabilistic expectations that are inconsistent with the point forecasts and (ii) how shifting the scale improves the consistency of predictions by allowing respondents to state more detailed beliefs about higher inflation ranges. We also explore potential disadvantages of adjusting the response scale.

Keywords: Probabilistic expectations, inflation, survey data

JEL classification: D84, E31, E58

---

*We are grateful to the Bundesbank for including our questions in the Bundesbank Online Panel Households. Our research was improved through very helpful comments and suggestions by Christian Conrad, Jonas Dovern, Johannes Frank, Lora Pavlova, Morten Ravn, Norbert Schwarz, Georg Strasser, Michael Weber and participants of the ifo Dresden Workshop on Macroeconomics and International Finance in April 2023 and the Economic Expectations Workshop at Heidelberg University in July 2023.

[†]Heidelberg University, christoph.becker@awi.uni-heidelberg.de.

[‡]University of Mannheim, duersch@xeeron.de.

[§]Heidelberg University, thomas.eife@awi.uni-heidelberg.de.

[¶]Corresponding author. ZEW – Leibniz Centre for European Economic Research, alexander.glas@zew.de.

1 Introduction

Survey data are a popular source of information about the macroeconomic expectations of experts, households and firms. In addition to point forecasts, many surveys collect probabilistic expectations by asking respondents to assign probabilities to pre-defined outcome intervals ('bins'). These probability distributions offer important insights into how survey participants assess the uncertainty, skewness and tail risk associated with their predictions (Manski, 2004).

In this paper, we analyze the quality of the probabilistic inflation expectations from the Bundesbank Online Panel Households (BOP-HH) in light of the recent surge of inflation rates in Germany and the euro area as a whole. In particular, we assess whether adjusting the bin definitions improves the consistency between the point forecasts and the probabilistic expectations by conducting a randomized experiment in June 2022 where some of the participants receive the original response scale, while others receive an alternative design where the center of the intervals is closer to—but still below—the actual German inflation rate at the time.

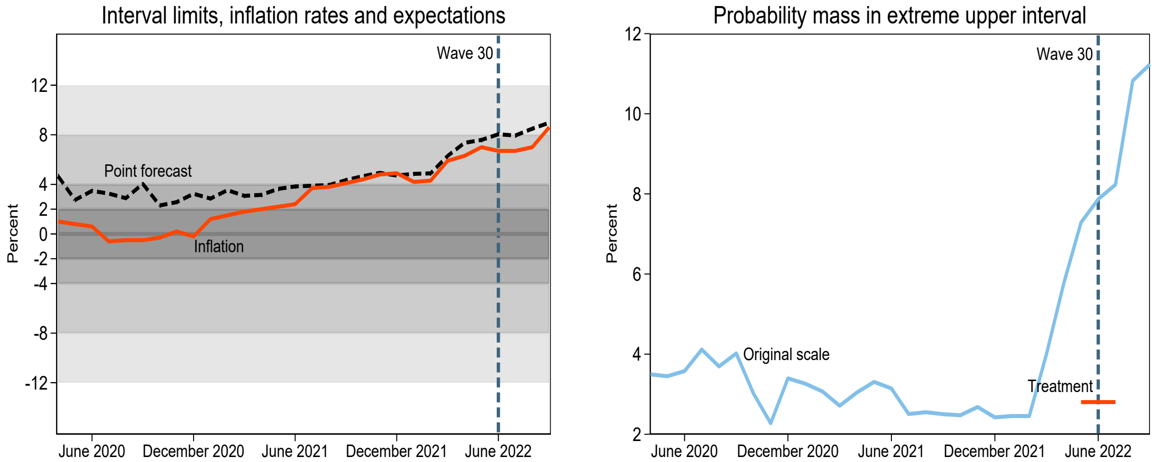
Our central finding is that the alternative design leads to considerably more consistent responses with the probabilistic expectations closely matching both actual inflation and point forecasts. This improved match between point forecasts and probabilistic expectations is driven by the fact that the original scale offers respondents a relatively small set of reasonable choices at times when inflation is very low or very high. For example, respondents who expect inflation rates of eight percent or higher only have two intervals at their disposal. This forces them either to provide inconsistent answers or to assign probabilities in extreme, marginal intervals, which are half-open and thus provide no means to signal the upper bound of their beliefs. Our finding is relevant for all surveys that employ scales similar to that used in BOP-HH such as the Federal Reserve Bank of New York's Survey of Consumer Expectations (SCE) and the European Central Bank's Consumer Expectations Survey.[1]

As illustrated by the shaded areas in the left plot in Figure 1, the question about probabilistic expectations consists of ten bins which are centered around an inflation rate of 0%. The interior bins cover the range from −12% to +12%. The two exterior bins are half-open. A major advantage of using this scale is that it facilitates comparisons both within surveys (across time) and between surveys (across different geographical locations). The red line shows the monthly German inflation rate based on the consumer price index. Before 2021, inflation rates were close to the center of the response scale. Inflation began to rise during the COVID-19 pandemic and further accelerated after the Russian invasion of Ukraine in February 2022 and the associated energy crisis. The inflation rate in June 2022 (Wave 30), when our experiment was conducted, was 6.7%. By September 2022, inflation further increased to 8.6%, which is above the lower bound of the rightmost interior bin.[2] The black line shows average inflation expectations in the BOP-HH. Clearly, households take notice of this development and adjust their point forecasts accordingly.

[1]See Armantier et al. (2017) for an overview of the SCE. Other examples include similar surveys conducted by the central banks of Canada, France, the Netherlands, Ukraine and the United Kingdom.

[2]The first estimates of German inflation in 2022 were even higher than the final releases in Figure 1. For example, inflation in June (September) was initially estimated to be 7.6% (10%). These figures more closely resemble the information available to households at the time they participated in the BOP-HH.

Figure 1: Probabilistic inflation expectations and interval definitions



Notes: The left plot shows monthly German consumer price inflation (red line). The dashed black line depicts the average inflation expectations of German households (individuals responses are trimmed by 1% from bottom and top in each month). Shaded gray areas correspond to the original bin definitions in the BOP-HH. The dashed blue line indicates the June 2022 wave of the BOP-HH to which we contributed an alternative bin design. The right plot shows the average probability mass in the rightmost bin based on participants presented with the original bin design (blue line). The red bar shows the corresponding average probability mass for the individuals presented with the alternative design.

The increase in households' point predictions is accompanied by an upward shift in their probabilistic inflation expectations. The blue line in the right plot shows the average probability mass assigned to the rightmost (half-open) bin. Before February 2022, the average probability fluctuated at relatively low levels between 2% and 4%. Consistent with the higher average point forecasts, we observe a steep increase in the average probability following the Russian invasion of Ukraine. The average probability in the rightmost bin was 7.9% in June 2022 and rose even further to more than 11% by September 2022. Since it is unknown what respondents consider a likely upper bound for inflation, the information provided by the open interval is limited. One has to make an assumption about the upper bound to derive a belief distribution from the answers. Thus, the evidence in Figure 1 puts into question the reliability of moments derived from the probabilistic expectations based on the original survey design since 2022.

We contributed an alternative bin design to Wave 30 of the BOP-HH where the center of the intervals is shifted from 0% to 4%, while keeping the relative bin width identical to the original design. As a result, the interior bins in the alternative treatment cover a range from −8% to +16%. The red bar in the right plot shows that for this treatment group, the average probability mass assigned to the rightmost bin is only 2.8%, which is much more in line with the figures observed in earlier survey waves. We show that these respondents use more bins and have higher histogram means that are more consistent with their point forecasts. In addition, households in the treatment group report lower standard deviations than those in the baseline group since they are able to use the tighter interior bins. We conclude that the distortion of moments of the obtained belief distribution can be reduced by adjusting the bin definitions at times when inflation is unusually high.

The observed differences in histogram moments may also be driven by other factors beyond allowing households to better express their expectations. For example, the higher histogram means may be the result of central tendency bias (see Becker et al., 2023), i.e., the tendency of respondents to assign more probability mass to bins near the center of the scale. Moreover, even under otherwise idential beliefs the estimated moments might differ mechanically due to the way the moments are calculated from the discrete survey responses. We explore such alternative explanations and show that they cannot fully explain the differences between the two groups.

Our research relates to the literature that explores how households form their macroeconomic expectations. Important covariates include households' socioeconomic characteristics such as gender, income and education (Bruine de Bruin et al., 2010; Das et al., 2020), their sources of information about monetary policy and the state of the economy (Coibion et al., 2022; Conrad et al., 2022) as well as individual and macroeconomic lifetime experiences (Malmendier and Nagel, 2011, 2016; D'Acunto et al., 2021). While these studies focus mostly on households' point forecasts, we consider the probabilistic expectations. Using the Michigan Survey of Consumers, Bruine de Bruin et al. (2011) show that consumers are generally willing and able to provide meaningful probability distributions that are consistent with the point predictions. Similarly, Zhao (2023) finds that the point forecasts of households in the SCE tend to be well-aligned with their probabilistic expectations. We contribute to the literature by analyzing whether the quality of the probabilistic expectations is related to the formulation of the corresponding question in the survey questionnaire in high-inflation regimes. As such, our analysis also relates to the literature that analyzes how specifics of the survey design influence the responses. Here, Schwarz (2010) gives a good overview in general while Weber et al. (2022) and Becker et al. (2023) discuss this point in the context of inflation expectations.

The rest of this paper is organized as follows. Section 2 explains the data and discusses the competing designs of the question used for the probabilistic inflation expectations. Section 3 presents the results. We discuss our findings in Section 4. Section 5 concludes.

2 Bundesbank Online Panel Households

We use data from the BOP-HH, a representative online survey of German households operated by the Bundesbank. The survey targets individuals aged 16 years or older (see Beckmann and Schmidt, 2020, for details on the elicitation process). Among other questions, participants are asked to state their inflation expectations and socioeconomic characteristics. The survey started in 2019 with three pilot surveys. Starting with Wave 4 (April 2020), the BOP-HH is issued on a monthly basis. We focus on the responses from Wave 30 (June 2022) to which we contributed alternative formulations for the question on the probabilistic inflation expectations.

In total, 4,460 households participated in Wave 30. We remove observations from the sample whenever the household did not report probabilistic inflation expectations or if information for any of the socioeconomic characteristics is missing. We also exclude one respondent who did not state whether her point forecast represents a deflation rate or an inflation rate. This leaves 4,092 observations in our sample.

2.1 Probabilistic inflation expectations

BOP-HH participants receive the following question on their probabilistic expectations:[3]

`CM004: In your opinion, how likely is it that the rate of inflation will change as follows over the next twelve months?`

- `The rate of deflation (opposite of inflation) will be 12% or higher.`
- `The rate of deflation ([...]) will be between 8% and less than 12%.`
- `The rate of deflation ([...]) will be between 4% and less than 8%.`
- `The rate of deflation ([...]) will be between 2% and less than 4%.`
- `The rate of deflation ([...]) will be between 0% and less than 2%.`
- `The rate of inflation will be between 0% and less than 2%.`
- `The rate of inflation will be between 2% and less than 4%.`
- `The rate of inflation will be between 4% and less than 8%.`
- `The rate of inflation will be between 8% and less than 12%.`
- `The rate of inflation will be 12% or higher.`

Respondents are asked to rate the probability of inflation falling into each bin on a scale from 0% to 100%, with 0% meaning that this outcome is completely unlikely and 100% meaning that they are absolutely certain it will happen. They also receive a notification that probabilities should add up to 100%. As mentioned above, the baseline design is centered around an inflation rate of 0%. Motivated by the recent surge in inflation rates, we contributed the following alternative bin design to the questionnaire of Wave 30:

`P3001A: In your opinion, how likely is it that the rate of inflation will change as follows over the next twelve months?`

- `The rate of deflation (opposite of inflation) will be 8% or higher.`
- `The rate of deflation ([...]) will be between 4% and less than 8%.`
- `The rate of deflation ([...]) will be between 0% and less than 4%.`
- `The rate of inflation will be between 0% and less than 2%.`
- `The rate of inflation will be between 2% and less than 4%.`
- `The rate of inflation will be between 4% and less than 6%.`
- `The rate of inflation will be between 6% and less than 8%.`
- `The rate of inflation will be between 8% and less than 12%.`

[3] All questions related to inflation include an info box that informs respondents that inflation is defined as the percentage change in the general price level as measured by the consumer prices index.

- `The rate of inflation will be between 12% and less than 16%.`
- `The rate of inflation will be 16% or higher.`

In the new formulation, the center of the scale is shifted upwards by four percentage points. As a result, the bins are centered around an inflation rate of 4%, which is closer to—but still below—the actual inflation rate in June 2022 (6.7%) relative to the baseline design. We leave the number of bins as well as their widths unchanged.[4]

The sample in Wave 30 was split into three randomly assigned groups. Approximately one third of the sample (1,355 observations) was presented with the baseline design used in all previous waves. Another third of the sample (1,376 observations) was presented with the alternative design which we refer to as the 'mean-shift' setting. The remaining 1,361 observations were presented with another bin design which we do not use in our analysis.[5] Thus, our analysis focuses on the 2,731 households in the baseline group and the mean-shift group. Their responses were collected between 15 June and 29 June. Accordingly, the most recent inflation figure available to them was the first estimate of the German inflation rate in May 2022 (7.9%; the final release is 7.0%) which was released by the German statistical office on 14 June. Only one response was elicited on 29 June when the first estimate of the June inflation rate was released (7.6%).

In the analysis below, we examine the impact of the alternative response scale on households' probabilisitic expectations. We are particularly interested in potential differences in the shape of the histograms reported by the baseline group and the mean-shift group. To assess such differences on an individual basis, we define the dummy variable *meanshift* that equals one if the individual is assigned to the mean-shift group, and zero else. Next, we calculate the number of bins with nonzero probability (*bins*) and the probability mass assigned to the rightmost bin (*phigh*). Note that the rightmost bin is defined differently across groups. We also consider the dummy variable *multipeak* which equals one if the histogram has multiple modes, and zero else. Table A.1 in the Appendix provides details on the construction of all variables.

After these steps, we compute the first four moments of each histogram. To do so, we follow Conrad et al. (2022) and assume that the probability in each bin is located at the midpoint.[6] To close the exterior bins, we assume that they have equal width to the

[4]Expert surveys such as the Survey of Professional Forecasters (SPF) operated by the ECB and the Federal Reserve Bank of Philadelphia cover a relatively narrow outcome range. As a result, their operators frequently adjust the scale in a way similar to our proposed alternativet desgin. This is usually done in reponse to macroeconomic shocks such as the Great Recession where a considerable pile-up of probabilities in the lowest bin for GDP growth was observed in the ECB-SPF. During the COVID-19 pandemic, the ECB-SPF introduced bins with unequal width. Glas and Hartmann (2022) show that this can have an impact on the mismatch between ex-ante uncertainty as measured by the histogram variance and ex-post uncertainty based on the variability of forecast errors.

[5]This design retains the centering around 0% but includes a more granular definition of the interior bins. See Becker et al. (2023) for the motivation behind this approach. Since the inflation rate during our period of study is far away from 0%, we do not use these observations in our analysis. However, all tables and figures for this alternative treatment are available upon request from the authors.

[6]Other alternatives include assuming uniformly distributed probabilities or fitting a continuous distribution as in Engelberg et al. (2009). However, Glas (2020) shows that this choice has little impact on estimates of the mean or the standard deviation. Moreover, Becker et al. (2022) show that fitting continuous distributions can lead to misleading results in the presence of varying interval widths.

adjacent bins, i.e., four percentage points.[7] Based on the 'mass-at-midpoint' approach, mean (μ), standard deviation (σ), skewness (γ) and kurtosis (κ) of the histogram reported by household $i = 1, \ldots, n$ are calculated as follows:

$$\mu_i = \sum_{k=1}^{K} m_k \times p_{i,k} \tag{1}$$

$$\sigma_i = \sqrt{\sum_{k=1}^{K} (m_k - \mu_i)^2 \times p_{i,k}} \tag{2}$$

$$\gamma_i = \frac{\sum_{k=1}^{K} (m_k - \mu_i)^3 \times p_{i,k}}{\sigma_i^3} \tag{3}$$

$$\kappa_i = \frac{\sum_{k=1}^{K} (m_k - \mu_i)^4 \times p_{i,k}}{\sigma_i^4} \tag{4}$$

In Eqn. (1)-(4), the index $k = 1, \ldots, K$ denotes the different bins, m_k is the midpoint of the k-th bin and $p_{i,k}$ is the probability assigned to this bin by household i. Using the BOP-HH, Dovern (2023) shows that σ_i is strongly correlated with households' qualitative assessment of uncertainty. Thus, we treat σ_i as a measure of inflation uncertainty.

Panel A of Table 1 presents summary statistics for all histogram characteristics by treatment status. For skewness and kurtosis, we consider only the responses of participants who use at least three bins. For each variable, the last column shows the p-values for a test of equal group means. We find that, on average, the individuals in the mean-shift group use significantly more bins, assign lower probability mass to the rightmost bin, report higher histogram means and lower standard deviations. We explore these differences in more detail in the next section.

2.2 Point forecasts

In addition to the probabilistic expectations, the BOP-HH elicits households' perceptions of current inflation ($\hat{\pi}_i^P$) and point forecasts of inflation for the coming year ($\hat{\pi}_i^E$). The corresponding questions are identical for all treatment groups. We analyze the consistency of point forecasts and density means via the absolute difference between $\hat{\pi}_i^E$ and μ_i. Since it has been shown that there exists a tight link between perceived and expected inflation (Jonung, 1981; D'Acunto et al., 2021), we also consider $\hat{\pi}_i^P$, although only one third of the participants in Wave 30 were asked for their perception of the current inflation rate over the previous twelve months. To reduce the impact of outliers, we trim the top and bottom 1% of inflation perceptions/expectations.

[7] Armantier et al. (2017) and Zhao (2023) use ±38% as the bounds for the exterior bins in their analyses of the SCE. This choice is based on historically observed inflation rates in the US. For Germany after the second world war, such extreme inflation rates have not been observed. Zhao (2023) mentions in his footnote 14 that he also considered ±16% for the bounds and that this choice did not affect his findings. Our choice for the bounds also makes it more difficult to detect potentially existing differences between histogram means across treatments and when comparing histogram means to point forecasts.

Table 1: Summary statistics for Wave 30 of the BOP-HH

	Baseline group			Mean-shift group			
	Obs.	Mean	SD	Obs.	Mean	SD	p-value
Panel A: Probabilistic inflation expectations							
bins	1,355	2.97	1.96	1,376	3.27	2.21	0.00***
phigh	1,355	7.87	19.91	1,376	2.80	12.77	0.00***
multipeak	1,355	0.08	0.26	1,376	0.08	0.27	0.92
μ_i	1,355	6.57	3.82	1,376	7.24	3.41	0.00***
σ_i	1,355	2.02	1.74	1,376	1.83	1.74	0.00***
γ_i	694	0.06	0.84	777	0.11	0.82	0.27
κ_i	694	3.61	2.45	777	3.53	2.24	0.53
Panel B: Point forecasts							
$\hat{\pi}_i^P$	443	6.63	2.53	434	6.67	2.55	0.82
$\hat{\pi}_i^E$	1,327	8.11	3.53	1,349	8.15	3.31	0.77
$\lvert\hat{\pi}_i^E - \mu_i\rvert$	1,327	2.17	3.40	1,349	1.60	2.54	0.00***
Panel C: Socioeconomic characteristics							
age	1,355	56.98	14.35	1,376	56.85	14.61	0.82
east1989	1,355	0.16	0.37	1,376	0.16	0.37	0.74
female	1,355	0.36	0.48	1,376	0.39	0.49	0.12
fullemploy	1,355	0.44	0.50	1,376	0.42	0.49	0.37
hhsize	1,355	2.20	1.04	1,376	2.20	1.07	0.94
income	1,355	3.98	2.01	1,376	3.94	1.95	0.62
yoe	1,355	11.55	1.67	1,376	11.51	1.69	0.58

Notes: This table shows summary statistics for the probabilistic inflation expectations (Panel A), point forecasts (Panel B) and socioeconomic characteristics (Panel C) of participants in Wave 30 of the BOP-HH. For skewness and kurtosis, we focus on responses where nonzero probability is assigned to at least three bins. We trim $\hat{\pi}_i^P$ and $\hat{\pi}_i^E$ by 1% from top and bottom. Household income is expressed in 1,000 euro. The last column shows p-values for tests of equal group means. Asterisks '*', '**', and '***' indicate significant differences in group means at the 10%, 5%, and 1% critical level, respectively.

Panel B of Table 1 shows summary statistics for the point forecasts. The average perceived inflation rate (calculated as the weighted average across the two groups) is 6.65%. Thus, the average participant underestimates inflation in May 2022. The weighted average of expected inflation is 8.13%. Clearly, households take notice of the surge in inflation as is also shown by the upward trend in average expectations shown in Figure 1.

In contrast to the probabilistic expectations, the means of perceived/expected inflation are not significantly different across groups. This is to be expected, as $\hat{\pi}_i^P$ and $\hat{\pi}_i^E$ are elicited before the probabilistic expectation and thus our treatment intervention (and respondents cannot return to previous questions). In both cases, the average point forecasts exceeds the average histogram means. However, due to the higher average histogram mean, the mismatch between point forecasts and histogram means is markedly lower for the mean-shift group and the difference between groups is significant.

2.3 Socioeconomic characteristics

Besides households' inflation expectations, we use information on their socioeconomic status. We consider age (*age*), gender (*female*), employment status (*fullemploy*), whether the individual lived in East Germany at the time of the German reunification (*east1989*), household size (*hhsize*), monthly household income (*income*) and years of education (*yoe*). These variables have been shown to be robust predictors of households' macroeconomic expectations and uncertainty thereof (Bruine de Bruin et al., 2010, 2011; Das et al., 2020). In all regressions below, we use the natural logarithm of income as a covariate.

Panel C of Table 1 presents summary statistics. The average respondent in Wave 30 is 57 years old and has almost 12 years of education. 38% of the individuals are female, 43% are full-time employed and 16% lived in East Germany at the time of reunification.

As with the point forecasts, socioeconomic characteristics are distributed similarly across groups and none of the mean differences are significant. Table 2 shows the estimates from regressing the *meanshift*-dummy on the socioeconomic variables. The baseline is the group of households that were presented with the original bin design. None of the coefficients are significantly different from zero, the (adjusted) R^2 is zero and the hypothesis that the slope coefficients are jointly equal zero is not rejected ($p = 0.83$). Taken together, these findings imply successful random assignment of the treatment.

3 Differences in Inflation Expectations by Treatment Status

This section presents our results for the differences in inflation expectations between the baseline group and the mean-shift group. Table 3 shows estimates from linear regressions of inflation expectations on treatment status and socioeconomic characteristics. Although the treatment is assigned randomly, we include the latter to increase the efficiency of the estimates by reducing the residual variance. Columns (1)-(3) show the results for the number of bins with nonzero probability, the probability mass in the rightmost bin and the indicator for multimodal histograms. Columns (4)-(7) present the estimates for the histogram moments. Columns (8)-(9) show the findings for the point forecasts and the mismatch between point forecasts and histogram means. All regressions are estimated with heteroskedasticity-consistent standard errors.

3.1 Differences in histogram characteristics

The histogram characteristics in Columns (1)-(7) are most likely to be affected by the alternative bin design. Indeed, we observe some noticeable differences across groups. In particular, we find that those in the mean-shift group use significantly more bins, assign a considerably lower probability mass to the rightmost bin, report higher histogram means and have lower inflation uncertainty than those in the baseline group. With the exception of the relatively small estimate for *bins*, these effects are also economically significant. For example, Column (2) shows that the average probability mass in the rightmost bin is more than five percentage points lower for the mean-shift group than for the baseline group. This corresponds to the vertical difference in the right plot of Figure 1. Similarly,

Table 2: Treatment assignment and socioeconomic characteristics

	meanshift
age	-0.06
	(0.08)
east1989	-0.68
	(2.64)
female	2.65
	(2.04)
fullemploy	-2.01
	(2.37)
hhsize	-0.15
	(1.09)
ln(*income*)	0.39
	(2.12)
yoe	-0.34
	(0.60)
Constant	54.58***
	(16.52)
Observations	2,731
$\bar{R}^2$	0.00
p-value (F-test)	0.83

Notes: This table presents the estimates from a linear regression of treatment status on socioeconomic characteristics. The baseline group consists of the individuals that were presented with the original bin design. Heteroskedasticity-consistent standard errors are reported in parentheses. The reported coefficients and standard errors are the estimated ones times 100. Asterisks '*', '**', and '***' indicate significance at the 10%, 5%, and 1% critical level, respectively.

Column (4) shows that the histogram means in the mean-shift group are, on average, 0.65 percentage point higher than those in the baseline group.

The estimated treatment effect for inflation uncertainty deserves special attention. On the one hand, the negative estimate may appear counterintuitive. If indeed respondents in the baseline group are restricted by the upper bound of 12%, one may expect that their distribution is artificially compressed. This appears to be supported by the observation that those in the mean-shift group use more of the available bins than those in the baseline group (although the effect is not very large). In that case, shifting the center of the scale to the right should increase the average standard deviation. On the other hand, recall that the bins right next to the rightmost bin are twice as large as those near the center of the scale. If respondents assign considerable probability to these wide interior bins, the standard deviation of their histogram may inflated relative to a situation where they are

Table 3: Differences in inflation expectations across treatments

	(1) *bins*	(2) *phigh*	(3) *multipeak*	(4) μ_i	(5) σ_i	(6) γ_i	(7) κ_i	(8) $\hat{\pi}_i^E$	(9) $\lvert\hat{\pi}_i^E - \mu_i\rvert$
meanshift	0.30***	-5.16***	-0.07	0.65***	-0.20***	0.05	-0.08	0.01	-0.59***
	(0.08)	(0.64)	(1.00)	(0.14)	(0.07)	(0.04)	(0.12)	(0.13)	(0.12)
age	-0.03***	-0.04	0.14***	0.01**	-0.01***	-0.00	0.02***	0.00	-0.00
	(0.00)	(0.03)	(0.04)	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
east1989	-0.09	2.24**	2.02	0.33*	0.03	-0.00	0.04	0.54***	0.12
	(0.11)	(1.05)	(1.48)	(0.19)	(0.10)	(0.06)	(0.17)	(0.21)	(0.16)
female	-0.16*	3.43***	4.58***	0.64***	0.14*	-0.08*	-0.07	0.95***	0.50***
	(0.09)	(0.73)	(1.14)	(0.16)	(0.08)	(0.05)	(0.13)	(0.15)	(0.14)
fullemploy	-0.07	1.40*	-0.68	0.30*	-0.10	-0.01	0.06	0.33*	0.13
	(0.10)	(0.83)	(1.13)	(0.17)	(0.08)	(0.05)	(0.13)	(0.17)	(0.15)
hhsize	0.01	0.87**	0.55	0.17*	-0.01	-0.07***	0.04	0.26***	0.16*
	(0.05)	(0.42)	(0.51)	(0.09)	(0.04)	(0.02)	(0.06)	(0.09)	(0.08)
ln(*income*)	-0.10	-2.55***	-2.69**	-0.62***	-0.10	0.07	-0.23	-0.92***	-0.42**
	(0.10)	(0.81)	(1.06)	(0.16)	(0.08)	(0.05)	(0.15)	(0.17)	(0.17)
yoe	0.04	-0.39*	-1.13***	-0.06	-0.02	0.04***	0.09**	-0.09**	-0.06*
	(0.02)	(0.23)	(0.32)	(0.04)	(0.02)	(0.01)	(0.04)	(0.04)	(0.04)
Constant	4.92***	31.43***	31.57***	10.81***	3.82***	-0.71*	3.22***	15.37***	5.91***
	(0.78)	(5.79)	(8.99)	(1.32)	(0.65)	(0.39)	(1.10)	(1.40)	(1.47)
Observations	2,731	2,731	2,731	2,731	2,731	1,471	1,471	2,676	2,676
$\bar{R}^2$	0.04	0.05	0.02	0.02	0.02	0.01	0.01	0.04	0.02

Notes: This table presents the estimates from linear regressions of histogram characteristics and point forecasts on treatment status and socioeconomic characteristics. The baseline group consists of the individuals that were presented with the original bin design. In columns (6) and (7), we consider only the responses of participants who use at least three bins. For columns (8) and (9), we trim $\hat{\pi}_i^E$ by 1% from top and bottom. Heteroskedasticity-consistent standard errors are reported in parentheses. The reported coefficients and standard errors in Column (3) are the estimated ones times 100. Asterisks '*', '**', and '***' indicate significance at the 10%, 5%, and 1% critical level, respectively.

allowed to allocate probabilities to the tighter interior bins. We suspect that this is the main explanation for the negative estimate in Column (5).

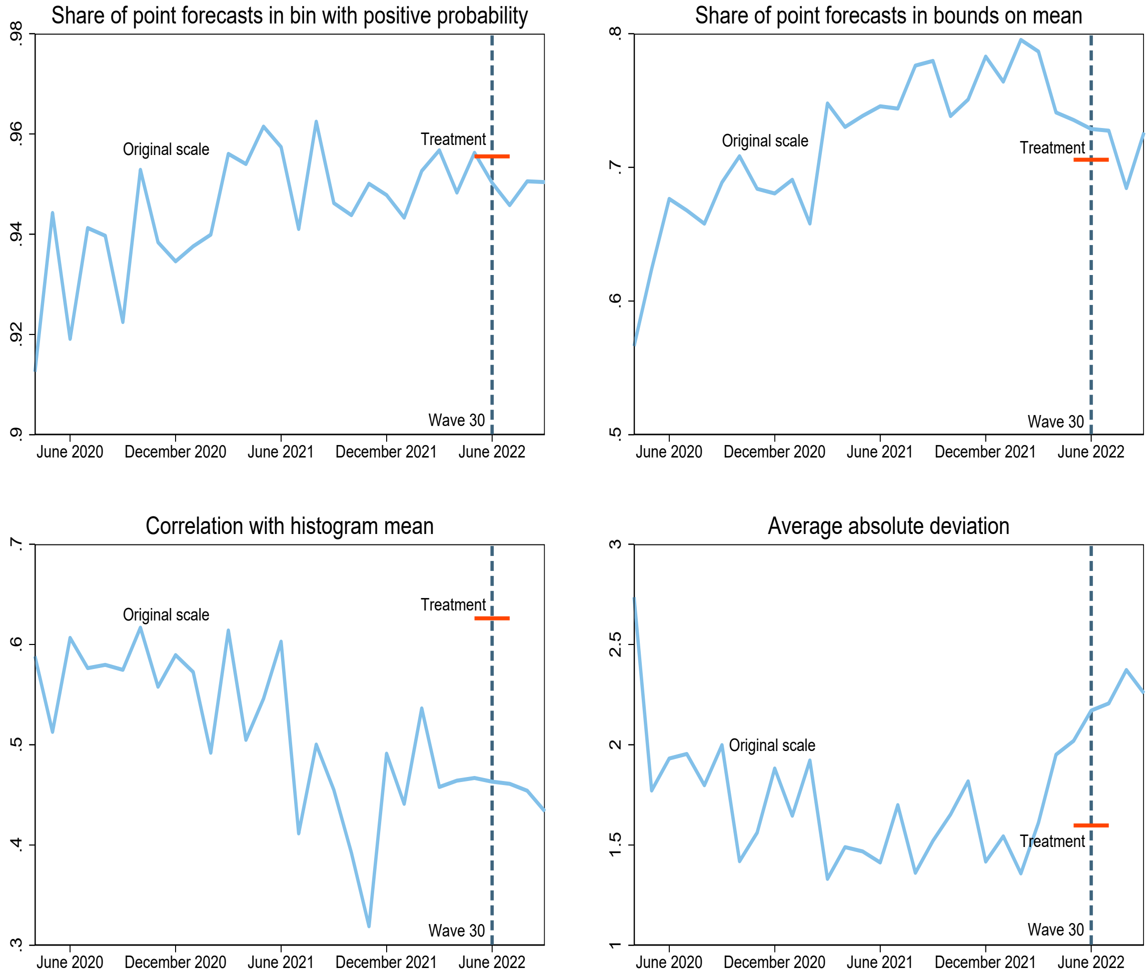
Given that all other factors such as the macroeconomic environment or the remaining questions in the survey questionnaire were identical for all respondents, the observed differences in the histogram characteristics are either due to genuinely higher expectations, framing effects or different discretization biases. We explore these issues in Section 4.

3.2 Consistency between point forecasts and probabilistic expectations

Bruine de Bruin et al. (2011) and Zhao (2023) find that the point forecasts of US households are well aligned with measures of central tendency such as the histogram mean. To assess whether this also is the case for German households, Figure 2 shows various measures of consistency between point forecasts and histogram means.

Around 95% of households report point forecasts that fall into a bin to which the respondents assigns nonzero probability. In line with the findings in Zhao (2023), approximately 70% of point forecasts lie within the individual bounds on the histogram mean, which are calculated by replacing the midpoint m_k in Eqn. (1) with the lower bound and the upper bound (see Engelberg et al., 2009, for details). These findings imply that point forecasts and probabilistic expectations of German households are relatively well-aligned

Figure 2: Consistency between point forecasts and probabilistic expectations



Notes: For each BOP-HH wave, the upper-left plot shows the share of point forecasts that fall into a bin to which the respondent assigns nonzero probability. The upper-right plot depicts the share of point forecasts that lie within the bounds on the histogram mean. The lower-left plot presents correlations between point forecasts and histogram means. The lower-right plot shows the average absolute deviation between point forecasts and histogram means. Point forecasts are trimmed by 1% from top and bottom. The red bars are the corresponding figures for the mean-shift group in Wave 30.

and supplement the evidence for the US. However, the correlation between point forecasts and histogram means exhibits a declining trend over time. Similarly, the average absolute deviation between $\hat{\pi}_i^E$ and μ_i has increased in recent waves. These results suggest that the alignment between point forecasts and histograms suffers at times when households are forced to assign considerable probablity to the exterior bins as was the case in recent waves (see Figure 1). At the same time, the last two subfigures show a much higher degree of consistency for the mean-shift group in Wave 30. For example, the correlation between $\hat{\pi}_i^E$ and μ_i is 0.63 for the mean-shift-group but only 0.46 for the baseline group. In light of these findings, we consider differences in the point forecasts and their alignment with the histogram means across treatment groups in the next step.

As discussed earlier, Column (8) of Table 3 shows again that the point forecasts of individuals in the baseline group and the mean-shift groups are not significantly different from each other. In fact, the estimated coefficient on the *meanshift*-dummy is essentially zero, which further confirms that the randomization of treatments was successful. Our combined findings of significantly higher histogram means for the mean-shift group and stable point forecast across groups imply that the consistency between point forecasts and probabilistic expectations must be higher for one of the two groups. Indeed, Column (9) shows that the average absolute deviation between point forecasts and histogram means is significantly smaller for the mean-shift group. The effect size of almost 0.6 percentage point is economically relevant and similar in magnitude to the observed difference in the histogram means across groups.

In sum, the findings in columns (2), (4), (8) and (9) suggest that participants in the mean-shift group are able to more adequately communicate their higher probabilistic beliefs about future inflation. This, in turn, leads to a higher degree of consistency between the point forecasts and the probabilistic expectations reported by those individuals.

3.3 Results for the control variables

Consistent with Armantier et al. (2021), we find that older respondents have significantly higher histogram means and report lower inflation uncertainty while using fewer bins. The *east1989*-dummy has a significantly positive effect on histogram means and point forecasts. This is in line with Goldfayn-Frank and Wohlfart (2020) who show that East Germans have higher inflation expectations than West Germans due to the inflationary shock experienced after reunification. Next, we find that women assign more probability mass to the rightmost bin and report higher histogram means and point forecasts. These findings align with similar evidence in Bruine de Bruin et al. (2011), Armantier et al. (2021) and Conrad et al. (2022). In addition, the probability of reporting a multi-peaked probability distribution is significantly higher for women and their point forecasts and histogram means tend to deviate more strongly. Full-time employed individuals and people living in larger households report higher point forecasts and histogram means. Higher income is associated with a lower probability mass in the rightmost bin (as in Armantier et al., 2021), lower point forecasts and a better match between point forecasts and histogram means (see Zhao, 2023). Lastly, higher education is associated with lower point forecasts and a better alignment between point forecasts and histogram means. The findings that high-income households and highly educated individuals have lower point forecasts are consistent with Bruine de Bruin et al. (2010).

3.4 Revisions of histogram moments

The rotating panel structure of the BOP-HH allows us to not only analyze cross-sectional differences in inflation expectations between treatment groups, but also differences in revisions of expectations over time. Of the 2,731 households in our sample for Wave 30, 737 also participated in Waves 29 and 31 (368 belong to the baseline group and 369 to the mean-shift group). For those individuals we can compute revisions in point forecasts and histogram characteristics. For example, the revision of the histogram mean between Wave 29 and Wave 30 is defined as $\Delta\mu_i = \mu_{i,June} - \mu_{i,May}$. Similarly, $\Delta\mu_i = \mu_{i,July} - \mu_{i,June}$

Table 4: Differences in revisions of inflation expectations across treatments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Differences in revisions from Wave 29 to Wave 30									
	$\Delta bins$	$\Delta phigh$	$\Delta multipeak$	$\Delta\mu_i$	$\Delta\sigma_i$	$\Delta\gamma_i$	$\Delta\kappa_i$	$\Delta\hat{\pi}_i^E$	$\Delta\lvert\hat{\pi}_i^E - \mu_i\rvert$
meanshift	0.02	-4.11***	-2.91	0.61**	-0.42***	0.09	0.08	0.26	-0.54**
	(0.12)	(1.41)	(2.00)	(0.29)	(0.11)	(0.11)	(0.23)	(0.23)	(0.27)
Observations	737	737	737	737	737	385	385	712	712
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$\bar{R}^2$	0.00	0.03	0.01	0.00	0.03	-0.01	0.00	0.01	0.00
Panel B: Differences in revisions from Wave 30 to Wave 31									
	$\Delta bins$	$\Delta phigh$	$\Delta multipeak$	$\Delta\mu_i$	$\Delta\sigma_i$	$\Delta\gamma_i$	$\Delta\kappa_i$	$\Delta\hat{\pi}_i^E$	$\Delta\lvert\hat{\pi}_i^E - \mu_i\rvert$
meanshift	-0.26***	7.20***	-1.04	-0.56**	0.25**	-0.08	0.10	-0.20	0.73***
	(0.10)	(1.40)	(1.81)	(0.28)	(0.10)	(0.12)	(0.31)	(0.22)	(0.26)
Observations	737	737	737	737	737	353	353	715	715
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$\bar{R}^2$	0.01	0.04	0.00	0.00	0.02	0.02	-0.01	0.01	0.01
Panel C: Net effect of revisions into and out of treatment									
	-0.24	3.17	-0.04	0.06	-0.17	0.01	0.17	0.07	0.18

Notes: This table presents the estimates from linear regressions of revisions of histogram characteristics and point forecasts on treatment status and socioeconomic characteristics for the subset of individuals that participated in Waves 29, 30 and 31. Panel A presents estimates for revisions between Wave 29 and Wave 30. Panel B focuses on revisions between Wave 30 and 31. The socioeconomic characteristics in Panel B are drawn from Wave 31 instead of Wave 30. Panel C shows the difference between the estimates in Panels B and those in Panel A. The baseline group consists of the individuals that were presented with the original bin design. In columns (6) and (7), we consider only the responses of participants who use at least three bins. For columns (8) and (9), we trim $\hat{\pi}_i^E$ by 1% from top and bottom. Heteroskedasticity-consistent standard errors are reported in parentheses. The reported coefficients and standard errors in Column (3) are the estimated ones times 100. Asterisks '*', '**', and '***' indicate significance at the 10%, 5%, and 1% critical level, respectively.

is the revision between Wave 30 and Wave 31. We analyze whether revisions of inflation expectations differ between the mean-shift group and the baseline group.[8]

Table 4 presents the results from linear regressions of revisions of point forecasts and histogram moments on treatment status and socioeconomic characteristics. The estimates for the latter are not shown to conserve space. Panel A presents estimates for revisions between Wave 29 and Wave 30. Panel B focuses on revisions between Wave 30 and 31, i.e., immediately after we conducted our experiment. Panel C shows the differences between the estimates in Panels B and A.

Panel A shows significant differences in revisions between groups for the probability mass assigned to the rightmost bin, the histogram mean, the standard deviation and $\Delta|\hat{\pi}_i^E - \mu_i|$. The magnitude of the estimated revisions into treatment is close in size to the treatment effects in Table 3. These findings support the notion that the observed differences in Wave 30 can indeed be ascribed to the alternative bin design.

Panel B presents the results for revisions between 30 and 31. Individuals who were assigned to the mean-shift group are again presented with the baseline bin definitions in Wave 31. We find that relative to the baseline group, these respondents more strongly adjust the number of bins, the probability mass in the rightmost bin, histogram means

[8]Table A.2 replicates Table 3 for the subset of respondents that participated in Waves 29 through 31. The estimates are very similar, although the magnitude of the effects tends to be slightly higher.

and standard deviations. Since the revisions of point forecasts do not differ significantly between groups, we also find a significant difference in the revised alignment of point forecasts and histogram means. These estimates have the opposite sign as those in Panel A and are relatively similar in size. However, Panel C shows that for $\Delta phigh$ and $\Delta|\hat{\pi}_i^E - \mu_i|$, the positive coefficients in Panel B exceed the negative coefficients in Panel A, which suggests that participants do not completely revert back to their pre-treatment expectations. Instead, they seem to at least partially retain their higher distribution from Wave 30.

4 Potential Transmission Channels

The results from the previous section reveal that changing the response scale in the current high-inflation regime affects the probabilistic expectations of BOP-HH participants. However, the reason behind the differences between the baseline group and the mean-shift is not immediately obvious. A first explanation is that the alternative design simply allows respondents to communicate more clearly their true beliefs at times when inflation is unusually high. However, other factors may also contribute to the observed deviations between treatment groups. This section explores these alternative channels.

4.1 Central tendency bias

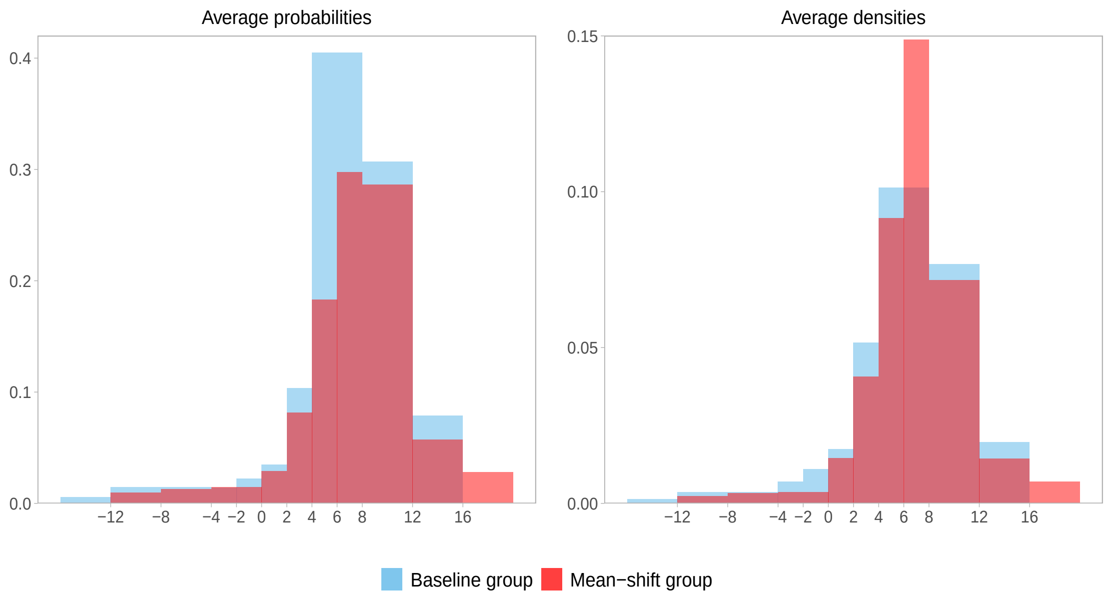
One alternative explanation is that the differences in responses are driven primarily by central tendency bias, i.e., a tendency of respondents to put more probability mass near the center of the scale. This central tendency bias could arise either from pure framing effects or as a result of rational updating by respondents. In particular, individuals may assume that inflation rates close to the center of the scale—0% for the baseline group, 4% for the mean-shift group—are deemed especially likely by the Bundesbank and adjust their expectations accordingly (see Becker et al., 2023, for a discussion). Note that, if updating were the main driver of central tendency bias, one may expect spillovers to probabilistic inflation expectations over subsequent survey waves (see Section 4.2) and/or to probabilistic expectations of other variables in the same wave (see Section 4.3).

Figure 3 shows the aggregate distributions of the individuals in both subsamples. The plot on the left depicts the average probability mass assigned to each bin across all respondents while the plot on the right shows the corresponding histogram by reporting densities instead of probabilities. Clearly, the aggregate distributions differ between the two groups. However, it appears that most of the differences occur for the interior bins while the probability mass assigned to inflation rates of 12% or more is nearly identical.

In order to disentangle framing effects from respondents being better able to express their true beliefs, we combine the ten original bins into seven wider bins which are defined identically for both groups. By comparing the average probability mass in each of these seven intervals, we are able to detect which regions of the distributions differ most strongly across groups. In particular, if the individuals in the mean-shift group assign more probability to the interval covering inflation rates between 4% (the center of the alternative scale) and below 8%, this would serve as evidence of central tendency bias. Table 5 shows the estimated treatment effects for each of the seven newly created bins.

The estimates in Column (5) show that the average probability mass assigned to inflation rates between 4% and below 8% is about eight percentage points higher for those

Figure 3: Average probabilistic inflation expectations by treatment status



Notes: The subfigures show the average responses for the individuals in the baseline group and the mean-shift group. The left plot depicts the average probability mass in each bin while the right plot shows the histograms by reporting densities instead of probabilities.

Table 5: Differences in bin probabilities across treatments

	(1) $(-\infty,-8)$	(2) $[-8,-4)$	(3) $[-4,0)$	(4) $[0,4)$	(5) $[4,8)$	(6) $[8,12)$	(7) $[12,+\infty)$
meanshift	-1.11***	-0.18	-2.18***	-2.67***	7.95***	-2.35*	0.53
	(0.37)	(0.31)	(0.37)	(0.78)	(1.30)	(1.28)	(0.78)
Observations	2,731	2,731	2,731	2,731	2,731	2,731	2,731
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$\bar{R}^2$	0.01	0.00	0.02	0.03	0.06	0.03	0.04

Notes: This table presents the estimates from linear regressions of the probability mass assigned to intervals of inflation outcomes on treatment status and socioeconomic characteristics. The baseline group consists of the individuals that were presented with the original bin design. In columns (6) and (7), we consider only the responses of participants who use at least three bins. For columns (8) and (9), we trim $\hat{\pi}_i^E$ by 1% from top and bottom. Heteroskedasticity-consistent standard errors are reported in parentheses. Asterisks '*', '**', and '***' indicate significance at the 10%, 5%, and 1% critical level, respectively.

in the mean-shift group. This is accomplished by assigning lower average porbability not just to the intervals to the left—as shown in Columns (1)-(4)—but also from the interval covering higher inflation rates between 8% and below 12%, as seen in Column (6). In contrast, Column (7) shows no evidence of significant differences in the probability mass assigned to inflation rates of 12% are more. These findings serve as evidence of central tendency bias, i.e., at least some of the observed difference between the two groups is a result of respondents assigning more probability mass to the new, higher center of the

Table 6: Differences in inflation expectations: interaction with *firsttimer*

	(1) *bins*	(2) *phigh*	(3) *multipeak*	(4) μ_i	(5) σ_i	(6) γ_i	(7) κ_i	(8) $\hat{\pi}_i^E$	(9) $\lvert\hat{\pi}_i^E - \mu_i\rvert$
meanshift	0.31***	-5.05***	0.65	0.68***	-0.17**	0.05	-0.13	-0.00	-0.63***
	(0.08)	(0.65)	(1.01)	(0.14)	(0.07)	(0.05)	(0.13)	(0.13)	(0.12)
firsttimer	0.78***	4.19*	10.98***	0.28	0.82***	-0.18*	-0.34	0.27	-0.06
	(0.23)	(2.17)	(3.63)	(0.33)	(0.22)	(0.09)	(0.22)	(0.35)	(0.29)
meanshift × *firsttimer*	-0.02	-0.83	-8.83*	-0.33	-0.38	-0.06	0.59	0.28	0.66
	(0.34)	(3.09)	(4.78)	(0.54)	(0.28)	(0.15)	(0.37)	(0.60)	(0.45)
Observations	2,731	2,731	2,731	2,731	2,731	1,471	1,471	2,676	2,676
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$\bar{R}^2$	0.04	0.05	0.03	0.02	0.02	0.01	0.01	0.04	0.02

Notes: This table presents the estimates from linear regressions of histogram characteristics and point forecasts on treatment status, a dummy variable for first-time participants, an interaction with treatment status and socioeconomic characteristics. The baseline group consists of the individuals that were presented with the original bin design. In columns (6) and (7), we consider only the responses of participants who use at least three bins. For columns (8) and (9), we trim $\hat{\pi}_i^E$ by 1% from top and bottom. Heteroskedasticity-consistent standard errors are reported in parentheses. The reported coefficients and standard errors in Column (3) are the estimated ones times 100. Asterisks '*', '**', and '***' indicate significance at the 10%, 5%, and 1% critical level, respectively.

response scale. Importantly, however, this result does not suggest that respondents in the alternative treatment are *more* susceptible to central tendency bias than those in the baseline group. Rather, it is likely that central tendency bias is a general issue regardless of the chosen center of the scale. This interpretation is supported by the estimates for the interval covering inflation between 0% (the center of the baseline definition) and below 4% in Column (4) where those in the baseline group assign the most additional probability relative to the mean-shift group.

4.2 Panel conditioning effects

In two recent papers, Weber et al. (2022) and Kim and Binder (2023) suggest that repeated participation induces individuals to learn about a topic they were previously unfamiliar with. This phenomenon is known as panel conditioning (or 'learning-through-survey') effect. In particular, Kim and Binder (2023) show that first-time participants in the SCE have higher point forecasts of inflation and display higher uncertainty than more experienced respondents.

We analyze whether the treatment effects of first-time BOP-HH respondents differ from those of other panelists. Of the 2,731 households in our sample, 196 (7%) participated in the BOP-HH for the first time. 106 of these individuals are assigned to the baseline group and the other 90 to the mean-shift group. The remaining 2,535 individuals participated at least once before. New entrants could simply assume that the alternative bin design represents the standard approach. On the other hand, it may be argued that participants with previous experience in the BOP-HH are somewhat 'anchored' around the original bin design. To explore these issues, we consider interactions between the *meanshift*-dummy and an indicator variable for first-time participants (*firsttimer*).

Table 6 shows that first-time participants use significantly more bins and report higher inflation uncertainty than more experienced respondents. This is consistent with the evidence in Kim and Binder (2023). In contrast, the coefficient on *firsttimer* is insignificant for both the point forecasts and the histogram means, although the estimates are rela-

Table 7: Differences in house price expectations across treatments

	(1) *bins*	(2) *phigh*	(3) *multipeak*	(4) μ_i	(5) σ_i	(6) γ_i	(7) κ_i	(8) $\hat{\pi}_i^E$	(9) $\lvert\hat{\pi}_i^E - \mu_i\rvert$
meanshift	0.01	-4.25**	3.48***	-0.58	0.02	0.05	-0.29*	-0.48	0.01
	(0.14)	(1.82)	(1.30)	(0.37)	(0.11)	(0.08)	(0.15)	(0.54)	(0.40)
Observations	864	864	864	864	864	462	462	844	844
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$\bar{R}^2$	0.05	0.04	0.00	0.02	0.04	-0.01	0.00	0.04	0.02

Notes: This table presents the estimates from linear regressions of histogram characteristics and point forecasts on treatment status and socioeconomic characteristics. The baseline group consists of the individuals that were presented with the original bin design. In columns (6) and (7), we consider only the responses of participants who use at least three bins. For columns (8) and (9), we trim $\hat{\pi}_i^E$ by 1% from top and bottom. Heteroskedasticity-consistent standard errors are reported in parentheses. The reported coefficients and standard errors in Column (3) are the estimated ones times 100. Asterisks '*', '**', and '***' indicate significance at the 10%, 5%, and 1% critical level, respectively.

tively large and have the expected positive sign. The lack of statistical significance may thus be the result of the small number of first-time respondents in our sample. We also find that the average probability mass in the rightmost bin is about four percentage points higher for first-time respondents and the probability of reporting a multi-peaked probability distribution is almost eleven percentage points higher. Importantly, however, the interactions between *meanshift* and *firsttimer* are all insignificant, except for a weakly significant coefficient in Column (3). This suggests that first-time participants do not react differently when presented with the alternative bin design compared to panelists with more survey experience.[9]

4.3 Spillovers to expectations of other variables

Another potential problem could be that changing the bin definitions for the probabilistic inflation expectations might lead those in the treatment group to also adjust their probabilistic expectations for other variables. For example, the BOP-HH asks households to provide probabilistic expectations for the growth rate of local house prices over the next year. For all respondents (including those in the mean-shift group), the bins for house price expectations correspond to the baseline definition of the bins for inflation. Therefore, there is no immediate reason to suspect that house price expectations differ between the baseline group and the mean-shift group. Importantly, however, the survey questionnaire asks for the house price expectations *after* the inflation expectations. Hence, households in the mean-shift group may interpret the different bin definitions for inflation and house prices as a signal that informs their expectations for the latter. To investigate the existence of spillover effects, Table 7 re-estimates Table 3 when we replace all dependent variables with their counterparts for house price expectations (using the same notation for the distinct variables). Note that only a third of the respondents in Wave 30 received the question about probabilistic house price expectations.

[9] Similarly, Table A.3 shows that the group-specific differences in the average probability mass assigned to certain inflation ranges do not differ between first-time respondents and other households.

Although most of the estimates are insignificant, a few things stand out. First, the average probability mass in the rightmost bin is about four percentage points lower for those in the mean-shift group. It may be the case that these individuals notice that different scales are used for inflation and house price expectations and therefore assume that the Bundesbank deems outcomes in the exterior bins particularly unlikely. Second, the probability of reporting a multi-peaked probability distribution is about three percentage points higher for those in the mean-shift group. This may suggest that using different scales for distinct variables is confusing for some respondents. Third, although we do not find evidence of significant differences in point forecasts or histogram moments (except for a weakly significant coefficient for the kurtosis), the estimates in Columns (4) and (8) are relatively large and suggest that, on average, those in the mean-shift group have lower house price expectations. These respondents may notice the lower center of the scale for the house price expectations relative to the inflation expectations and take that as a signal that the Bundesbank expects house prices to grow at a lower rate relative to inflation. The lack of significance could be the result of the relatively small sample size. Overall, our findings for the house price expectations should be interpreted as evidence of weak spillovers from changing the bin definitions for one variable to the probabilistic expectations of other variables.

4.4 Discretization errors

A final explanation for the estimated treatment effects is that at least some of the differences in responses can be ascribed to different discretizations of the scale across treatments which affects histogram moments even under the assumption of stable beliefs. To explore the magnitude of such 'technical errors', we consider a hypothetical setting where a household with fixed probabilistic expectations is confronted with the two bin designs. The expectations of this household are normally distributed with known mean μ_0 and variance σ_0^2, i.e., $\mathcal{N}(\mu_0, \sigma_0^2)$. While it is unrealistic to assume that all households have normally distributed expectations, this may be an appropriate assumption for highly educated respondents. Also, Table 1 shows that the average skewness and kurtosis of BOP-HH participants are close to values expected under normality. For the mean, we choose $\mu_0 \in \{0, 4, 8\}$, where a value of zero corresponds to the center of the bin definitions for the baseline group, four corresponds to the center of the definitions for the mean-shift group and eight is close to the actual inflation rate during our period of study. For the variance, we consider $\sigma_0^2 \in \{4, 9\}$ to capture settings with low and high inflation uncertainty. For each combination of μ_0, σ_0^2 and the bin definitions, we calculate the probability mass assigned to each bin and compute the histogram moments using Eqn. (1)-(4). Table 8 presents the results. To faciliate the comparison between true and estimated moments, we report variances instead of standard deviations.

While the estimated histogram moments clearly deviate across settings, they are usually fairly close to the true values. The absolute difference between the estimated histogram means across treatments is at most 0.23 percentage point in case of the setting with low uncertainty and small values of μ_0. This is much smaller than the estimated difference of 0.65 percentage point between baseline and mean-shift group in Column (4) of Table 3. Turning to the variances, we observe that the empirical variances exceed their true value in all settings. The largest difference between empirical variances across bin

Table 8: Histogram moments under stable expectations

	Baseline group	Mean-shift group	Baseline group	Mean-shift group
	$\mathcal{N}(0,4)$		$\mathcal{N}(0,9)$	
μ	0.00	−0.23	0.00	−0.14
σ^2	4.77	4.85	10.46	9.87
γ	0.00	0.16	0.00	0.05
κ	3.61	2.80	3.03	3.07
	$\mathcal{N}(4,4)$		$\mathcal{N}(4,9)$	
μ	4.23	4.00	4.14	4.00
σ^2	4.85	4.77	9.87	10.46
γ	−0.16	0.00	−0.05	0.00
κ	2.80	3.61	3.07	3.03
	$\mathcal{N}(8,4)$		$\mathcal{N}(8,9)$	
μ	8.02	8.23	8.04	8.14
σ^2	5.24	4.85	9.52	9.87
γ	0.12	−0.16	0.02	−0.05
κ	2.24	2.80	2.74	3.07

Notes: For both bin definitions, this table presents the estimated histogram moments derived under the assumption that respondents have normally distributed inflation expectations.

definitions—0.59 percentage point in absolute terms—is observed for the high-uncertainty scenario and small values of μ_0. This corresponds to an absolute difference in standard deviations of 0.09 percentage point. In contrast, Table 3 Column (5) shows that the estimated difference in standard deviations between treatment groups is more than twice as large. We conclude that our estimated treatment effects are too large to merely be the result of different discretizations across bin definitions.

5 Conclusion

For the current high-inflation environment, we find evidence that the moments of households' probabilistic inflation expectations vary with the response scale used to elicit them. This is particularly the case for the histogram mean. As a result, the wedge between point forecast and histogram mean depends on the setup used for the probabilistic expectations. We show that the histogram standard deviation is also affected. While the estimated treatment effects are too large to merely be the result of different discretizations, we do find some evidence in support of central tendency bias. Still, our results suggest that the inflation beliefs of German households have shifted upwards on average. Using the original scale to elicit expectations under these new beliefs tends to distort histogram moments as respondents have to allocate more probability mass to the higher, half-open interval in order to state their expectations. While we find the mean and variance to be affected, higher moments such as skewness and kurtosis appear to be relatively robust.

Our results have important implications for the operators of household surveys. We show that it may make sense to adjust the interval design to the current macroeconomic environment as it is commonly done in surveys of professional forecasters. A more fine-grained interval design might also be advisable to accurately capture inflation expectations once inflation surges. However, such adjustments come at the expense of the comparibility across different household surveys. A compromise, albeit costly, could be to use sample splits and include two different formulations for the probabilistic expectations at times when inflation is unusually low or high. Some of the participants receive the original design to retain consistency with previous waves and the remaining panelists are confronted with an alternative design where the center of the bin is closer to the actual inflation rate. Another option could be to center the response scale around either the most recently released inflation figure or the point forecast provided by the respondent prior to the question on the probabilistic expectation. While such flexible scales would result in heterogeneous bin definitions across time and respondents, both designs would likely limit the average probability mass assigned to the exterior bins at times when inflation rises. We have included an alternative bin design centered around the point forecast of each respondent in an upcoming wave of the BOP-HH and will explore its implications in future research.

References

Armantier, O., G. Kosar, R. Pomerantz, D. Skandalis, K. Smith, G. Topa, and W. van der Klaauw (2021). How economic crises affect inflation beliefs: Evidence from the Covid-19 pandemic. *Journal of Economic Behavior & Organization 189*, 443–469.

Armantier, O., G. Topa, W. V. der Klaauw, and B. Zafar (2017). An overview of the Survey of Consumer Expectations. Staff Reports 800, Federal Reserve Bank of New York.

Becker, C., P. Duersch, and T. Eife (2023). Measuring inflation expectations: How the response scale shapes density forecasts. available at SSRN: http://dx.doi.org/10.2139/ssrn.4323706.

Becker, C., P. Duersch, T. Eife, and A. Glas (2022). Extending the procedure of Engelberg et al. (2009) to surveys with varying interval-widths. FAU Discussion Papers in Economics No. 05/2022.

Beckmann, E. and T. Schmidt (2020). Bundesbank Online Pilot Survey on Consumer Expectations. Technical Paper.

Bruine de Bruin, W., C. F. Manski, G. Topa, and W. van der Klaauw (2011). Measuring consumer uncertainty about future inflation. *Journal of Applied Econometrics 26*(3), 454–478.

Bruine de Bruin, W., W. van der Klaauw, J. Downs, B. Fischhoff, G. Topa, and O. Armantier (2010). Expectations of inflation: The role of demographic variables, expectation formation, and financial literacy. *Journal of Consumer Affairs 44*(2), 381–402.

Coibion, O., Y. Gorodnichenko, and M. Weber (2022). Monetary policy communications and their effects on household inflation expectations. *Journal of Political Economy 130*(6), 1537–1584.

Conrad, C., Z. Enders, and A. Glas (2022). The role of information and experience for households' inflation expectations. *European Economic Review 143*, 104015.

D'Acunto, F., D. Hoang, M. Paloviita, and M. Weber (2021). Human frictions in the transmission of economic policies. NBER Working Paper 29279, National Bureau of Economic Research.

D'Acunto, F., U. Malmendier, J. Ospina, and M. Weber (2021). Exposure to grocery prices and inflation expectations. *Journal of Political Economy 129*(5), 1615–1639.

Das, S., C. M. Kuhnen, and S. Nagel (2020). Socioeconomic status and macroeconomic expectations. *The Review of Financial Studies 33*(1), 395–432.

Dovern, J. (2023). Eliciting expectation uncertainty from private households. *International Journal of Forecasting 40*(1), 113–123.

Engelberg, J., C. F. Manski, and J. Williams (2009). Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business & Economic Statistics 27*(1), 30–41.

Glas, A. (2020). Five dimensions of the uncertainty-disagreement linkage. *International Journal of Forecasting 36*(2), 607–627.

Glas, A. and M. Hartmann (2022). Uncertainty measures from partially rounded probabilistic forecast surveys. *Quantitative Economics 13*(3), 979–1022.

Goldfayn-Frank, O. and J. Wohlfart (2020). Expectation formation in a new environment: Evidence from the German reunification. *Journal of Monetary Economics 115*, 301–320.

Jonung, L. (1981). Perceived and expected rates of inflation in Sweden. *The American Economic Review 71*(5), 961–968.

Kim, G. and C. Binder (2023). Learning-through-survey in inflation expectations. *American Economic Journal: Macroeconomics 15*(2), 254–78.

Malmendier, U. and S. Nagel (2011). Depression babies: Do macroeconomic experiences affect risk taking? *The Quarterly Journal of Economics 126*(1), 373–416.

Malmendier, U. and S. Nagel (2016). Learning from inflation experiences. *The Quarterly Journal of Economics 131*(1), 53–87.

Manski, C. F. (2004). Measuring expectations. *Econometrica 72*(5), 1329–1376.

Schwarz, N. (2010). *Measurement as Cooperative Communication: What Research Participants Learn from Questionnaires*, pp. 43–61. Los Angeles: Sage.

SOEP-IS Group (2017). SOEP-IS 2015.1-PGEN: Person-related status and generated variables. SOEP Survey Papers 462: Series D - Variable Descriptions and Coding., DIW Berlin/SOEP.

Weber, M., F. D'Acunto, Y. Gorodnichenko, and O. Coibion (2022). The subjective inflation expectations of households and firms: Measurement, determinants, and implications. *Journal of Economic Perspectives 36*(3), 157–84.

Zhao, Y. (2023). Internal consistency of household inflation expectations: Point forecasts vs. density forecasts. *International Journal of Forecasting 39*(4), 1713–1735.

Appendix

Table A.1: Variable construction

Variable	BOP-HH Questionnaire	Description
	Probabilistic inflation expectations	
meanshift	*drandom2*	Equals one if the respondent belongs to the mean shift group (*drandom2*= 2), and zero for those in the baseline group (*drandom2*= 1).
bins	*infexprob_[a-j]* (CM004), *infexprob_rct1_[a-j]* (P3001A)	Number of bins to which the respondent assigns nonzero probability.
phigh	*infexprob_j* (CM004), *infexprob_rct1_j* (P3001A)	Probability mass assigned by the respondent to the rightmost available bin.
multipeak	same as *bins*	Equals one if the respondent provides a histogram with multiple peaks, and zero otherwise.
μ_i	same as *bins*	Mean of the histogram forecast for the German inflation rate over the next twelve months. We assume that the exterior bins have a width of four percentage points and that the probability mass in each bin is located at the midpoint.
σ_i	same as *bins*	Standard deviation of the histogram forecast.
γ_i	same as *bins*	Skewness of the histogram forecast.
κ_i	same as *bins*	Kurtosis of the histogram forecast.
	Point forecasts	
$\hat{\pi}_i^P$	*devinfpoint* (CQ002)	Perceived German inflation rate over the previous twelve months in percent. This question was only asked to approximately one third of the participants in Wave 30.
$\hat{\pi}_i^E$	*infdef* (CM002) and *inflexppoint* (CM003)	Expected German inflation rate over the next twelve months in percent. Equals *inflexppoint* if *infdef* equals 'Inflation' and $(-1)\cdot$ *inflexppoint* if *infdef* equals 'Deflation'.
$\lvert\hat{\pi}_i^E - \mu_i\rvert$	same as $\hat{\pi}_i^E$ and μ_i	Absolute difference between the point forecast and the histogram mean.
	Socioeconomic characteristics	
age	*age*	Age of individual. Set to 80 if *age* equals '80 years or older'.
east1989	*eastwest1989*	Equals one if eastwest1989 equals 'eastern Germany', and zero otherwise.
female	*gender*	Equals one if *gender* equals 'female', and zero otherwise.
fullemploy	*employ* (CS003)	Equals one if *employ* equals 'employed, full-time', and zero otherwise.
hhsize	*hhsize* (CS006)	Household size. Set to 6 if *hhsize* equals '6 or more'.
income	*hhinc* (CS008)	Monthly household income in €1,000 (using bin midpoints): = 0.25 if *hhinc* equals 'Less than €500', = 0.75 if *hhinc* equals '€500 to €999', = 1.25 if *hhinc* equals '€1,000 to €1,499', = 1.75 if *hhinc* equals '€1,500 to €1,999', = 2.25 if *hhinc* equals '€2,000 to €2,499', = 2.75 if *hhinc* equals '€2,500 to €2,999', = 3.25 if *hhinc* equals '€3,000 to €3,499', = 3.75 if *hhinc* equals '€3,500 to €3,999', = 4.50 if *hhinc* equals '€4,000 to €4,999', = 5.50 if *hhinc* equals '€5,000 to €5,999', = 7.00 if *hhinc* equals '€6,000 to €7,999', = 9.00 if *hhinc* equals '€8,000 to €9,999', = 11.00 if *hhinc* equals '€10,000 or more'.

Notes: This table describes the construction of the variables used in the empirical analysis. In the middle column, we refer to the names of the original variables as listed in the questionnaire for Wave 30 (June 2022) of the BOP-HH.

Table A.1: Variable construction (cont.)

Variable	BOP-HH Questionnaire	Description
yoe	*eduschool* (CS001)	Years of education of individual following SOEP-IS Group (2017): $= 7$ if *eduschool* equals 'No school-leaving certificate', $= 9$ if *eduschool* equals 'Secondary school-leaving certificate', $= 10$ if *eduschool* equals 'Other school-leaving certificate', $= 10$ if *eduschool* equals 'Intermediate secondary school certificate', $= 10$ if *eduschool* equals 'Polytechnical secondary school certificate (8th/10th grade)', $= 13$ if *eduschool* equals 'University of applied sciences entrance diploma / completed technical school', $= 13$ if *eduschool* equals 'Senior school-leaving certificate/ general or subject-specific university entrance diploma', $= 18$ if *eduschool* equals 'College / university degree'.
	Aditional characteristics	
firsttimer	*id*	Equals one if the respondent participated in the BOP-HH for the first time in Wave 30, and zero otherwise.

Notes: This table describes the construction of the variables used in the empirical analysis. In the middle column, we refer to the names of the original variables as listed in the questionnaire for Wave 30 (June 2022) of the BOP-HH.

Table A.2: Differences in inflation expectations: Wave 29 to Wave 31 participants

	(1) *bins*	(2) *phigh*	(3) *multipeak*	(4) μ_i	(5) σ_i	(6) γ_i	(7) κ_i	(8) $\hat{\pi}_i^E$	(9) $\lvert\hat{\pi}_i^E - \mu_i\rvert$
meanshift	0.09	-5.21***	-0.66	0.82***	-0.36***	0.08	0.22	-0.14	-0.97***
	(0.14)	(1.08)	(1.76)	(0.28)	(0.12)	(0.08)	(0.20)	(0.25)	(0.24)
Observations	737	737	737	737	737	404	404	721	721
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$\bar{R}^2$	0.04	0.06	0.01	0.03	0.04	0.02	0.02	0.06	0.03

Notes: This table presents the estimates from linear regressions of histogram characteristics and point forecasts on treatment status and socioeconomic characteristics for the subset of individuals that participated in Waves 29, 30 and 31. The baseline group consists of the individuals that were presented with the original bin design. In columns (6) and (7), we consider only the responses of participants who use at least three bins. For columns (8) and (9), we trim $\hat{\pi}_i^E$ by 1% from top and bottom. Heteroskedasticity-consistent standard errors are reported in parentheses. The reported coefficients and standard errors in Column (3) are the estimated ones times 100. Asterisks '*', '**', and '***' indicate significance at the 10%, 5%, and 1% critical level, respectively.

Table A.3: Differences in bin probabilities across treatments: interaction with *firsttimer*

	(1) $(-\infty, -8)$	(2) $[-8, -4)$	(3) $[-4, 0)$	(4) $[0, 4)$	(5) $[4, 8)$	(6) $[8, 12)$	(7) $[12, +\infty)$
meanshift	-1.24***	-0.11	-2.04***	-2.93***	7.81***	-2.14	0.64
	(0.39)	(0.33)	(0.38)	(0.81)	(1.36)	(1.34)	(0.80)
firsttimer	-1.23**	0.41	2.50*	-1.41	-2.66	-1.81	4.20*
	(0.54)	(0.78)	(1.35)	(1.81)	(3.20)	(3.05)	(2.17)
meanshift × *firsttimer*	1.78	-1.03	-1.69	3.72	1.65	-3.50	-0.93
	(1.24)	(0.87)	(1.52)	(2.86)	(4.68)	(4.31)	(3.46)
Observations	2,731	2,731	2,731	2,731	2,731	2,731	2,731
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$\bar{R}^2$	0.01	0.00	0.02	0.03	0.06	0.03	0.04

Notes: This table presents the estimates from linear regressions of histogram characteristics and point forecasts on treatment status and socioeconomic characteristics. The baseline group consists of the individuals that were presented with the original bin design. In columns (6) and (7), we consider only the responses of participants who use at least three bins. For columns (8) and (9), we trim $\hat{\pi}_i^E$ by 1% from top and bottom. Heteroskedasticity-consistent standard errors are reported in parentheses. Asterisks '*', '**', and '***' indicate significance at the 10%, 5%, and 1% critical level, respectively.

Download ZEW Discussion Papers:

https://www.zew.de/en/publications/zew-discussion-papers

or see:

https://www.ssrn.com/link/ZEW-Ctr-Euro-Econ-Research.html
https://ideas.repec.org/s/zbw/zewdip.html

//
IMPRINT

ZEW – Leibniz-Zentrum für Europäische Wirtschaftsforschung GmbH Mannheim

ZEW – Leibniz Centre for European Economic Research

L 7,1 · 68161 Mannheim · Germany
Phone +49 621 1235-01
info@zew.de · zew.de