Context effects in inflation surveys: The influence of additional information and prior questions

Xiaoxiao Niu and Nigel Harvey

University College London

Address for correspondence

Nigel Harvey

Department of Experimental Psychology

University College London

Gower Street

London WC1E 6BT

Email: n.harvey@ucl.ac.uk

Running head: Context in inflation surveys

Abstract

Context effects are known to affect responses to surveys. We report effects of information context and task context in surveys of inflation expectations. Information context refers to contextual information about earlier inflation rates or other economic indicators. Task context refers to judgment tasks performed prior to the inflation judgment task under consideration. In three experiments, we show that contextual information improves judgment accuracy. As this information is given in expert but not in lay surveys, its provision may partly explain why expert judgments are superior to those of lay people. In both expert and lay surveys, respondents make inflation judgments in the context of already having made other inflation judgments. We show that when different groups of people make inflation judgments either for the current year or for the upcoming year, their judgments do not differ. However, when the same people make judgments for both the current and the upcoming year, the latter are significantly higher than the former, perhaps because people expect inflation to increase over time.

Keywords: inflation surveys; inflation expectations; context effects; information context; task context

1. Introduction

The effect of different types of context on responses in both online and traditional surveys is well-documented (e.g., Reips, 2002; Smyth, Dillman and Christian, 2009; Tourangeau, Rips and Rasinski, 2000). Here we are concerned with how context influences people's judgments in inflation rate surveys. We focus on two types of context: information context and task context. Information context refers to information that people are given when they are asked to provide their judgments: for example, they may be provided with the actual inflation rate for the year before the one for which they are required to produce an estimate. Task context refers to the set of tasks in which their inflation judgment is embedded. For example, before judging the inflation rate for next year, they may be asked to judge the inflation rate for this year and, after judging the inflation rate for next year, they might be asked to judge the inflation rate for the year after that.

1.1 Inflation expectations

Central banks use surveys to monitor inflation expectations of lay people (households, consumers) and experts (economists and professional forecasters). It is important for banks to know about lay expectations because they are likely to influence future inflation levels: for example, the more that people expect inflation to increase, the more they will bring their planned purchasing of durable goods forward, thereby increasing the price of those goods by pushing up demand for them.

According to rational expectations theory (Muth, 1961), lay expectations should not differ from those of experts. The theory implies that rational economic agents form their expectations in line with what macroeconomic theories specify as rational. Thus it should not really be necessary to survey both lay people and experts: their expectations for inflation should be the same. However, they are not the same (Mankiw, Reiss and Wolfers, 2003; Palardy and Ovaska, 2015). Experts' inflation expectations are more accurate and show less heterogeneity than those of lay people. This disagreement between lay and expert forecasters may arise, in part, because they base their expectations on different types of information.

First, lay people are not exposed to or do not attend to information of the quality absorbed by experts (Binder and Rodrigue, 2018; Cavallo, Cruces and Perez-Truglia, 2017). News media comprise their main source of economic information and heterogeneity of their inflation expectations can be partly attributed to exposure to different reports (Maag and Lamla, 2009). Also, news media are likely to treat larger price rises for some items as more newsworthy than smaller rises for the majority of items: lay judgments of inflation are likely to be biased in an upward direction by this 'social amplification' process (Soroka, 2006). In contrast, experts are relatively well-informed and use similar datasets to update their beliefs (Coibion, Gorodnichenko, Kumar and Pedemonte, 2020; Gábriel, Rariga and Várhegyi, 2014).

A second difference is that only lay people draw on their own personal experience of price changes when forecasting inflation. As a result, differences in personal experience contribute to the greater heterogeneity observed in their inflation expectations (Bates and Gabor, 1986; Brachinger, 2008; Jungermann, Brachinger, Belting, Grinberg and Zacharias, 2007; Lein and Maag, 2011; Madeira and Zafar, 2015; Malmendier and Nagel, 2016; Ranyard, Missier, Bonini and Pietroni, 2018).

A third difference concerns the way in which information about certain other economic variables (e.g., inflation rates, unemployment rates) can be used to forecast inflation. Experts can use their macroeconomic models for this purpose. Lay people, without access to these models, may exploit their own naïve theories of how the economy works or use simple heuristics, such as the good-begets-good heuristic (Leiser and Krill, 2018). These lay approaches are likely to be less effective at forecasting inflation than the models used by experts.

These three factors can explain why inflation judgments by experts responding to surveys directed at them are superior to and more homogeneous than inflation judgments by lay people responding to surveys targeting them. Crucially, however, experts and lay people have been required to respond to *different* surveys. The notion that there is a difference between lay and expert judgments that is in need of explanation is predicated on the assumption that these different surveys are equally good at eliciting judgments of inflation. It is possible that this assumption is not valid. For example, if we asked experts to answer the consumer surveys normally given to lay people and lay people to respond to the surveys designed for professional respondents, we might find that the latter group are now more accurate and less homogeneous than the former one. While this outcome may not seem likely, the possibility that it could occur emphasises the importance of investigating the effects of survey format on the accuracy and homogeneity of inflation judgments. There have already been a number of studies of this issue.

1.2 Effects of survey format

Various surveys have been developed to elicit inflation expectations from lay respondents. They include the Michigan Survey of Consumers (MSC), the Federal Reserve Bank of New York's Survey of Consumer Expectations (SCE), and the Bank of England's Inflation Attitudes Survey (IAS). A different set of surveys have been designed to identify the inflation expectations of experts. These include the Federal Reserve Bank of Philadelphia's Survey of Professional Forecasters (US-SPF), and the European Central Bank's Survey of Professional Forecasters (EU-SPF).

Format varies across consumer surveys in a number of ways. In some cases, questions prompt point forecasts but, in other cases, they elicit probability density functions (Armantier, Bruine de Bruin, Potter, Topa, van der Klaauw & Zafar, 2013; Bruine de Bruin, Manski, Topa and van der Klaauw, 2011). Sometimes people are asked to estimate 'inflation' whereas, on other occasions, they are required to estimate 'general price change' (Armantier, Topa, van der Klaauw and Zafar, 2017; Bruine de Bruin, Potter, Rich, Topa and van der Klaauw, 2010; Bruine de Bruin, van der Klaauw, Topa, Downs, Fischhoff and Armantier, 2012; Bruine de Bruin, van der Klaaw, van Rooij, Teppa and de Vos, 2017). In some surveys but not others, respondents are given the opportunity to revise their answers (Bruine de Bruin et al., 2017). Main and interactive effects of these factors influence the inflation forecasts that people provide (Bruin de Bruin et al., 2017).

These studies demonstrate effects of variations in format across different surveys of lay expectations of inflation. Their findings are not directly relevant to results obtained from expert forecasters because surveys of that group (e.g., SPF) universally use the term 'inflation', always elicit pdfs (often in addition to point forecasts), and do not prompt respondents for revisions.

1.3 Information context: Differences between surveys of lay and expert forecasters

Our concerns here are with aspects of survey design that have not been previously studied. Specifically, we are interested in features that differ between lay and expert surveys. Our aim is to find out whether certain elements that are present in expert surveys but absent from lay surveys facilitate production of accurate and homogeneous inflation forecasts. The existence of such features could, at least partially, explain why lay forecasts for inflation have previously been found to be worse and less homogeneous than those of experts. In other words, the differences between the judgments made by lay and expert respondents may arise not from differences in their mental processing related to the three factors discussed above (dataset access, experience of price changes, macroeconomic knowledge) but from differences in format of the surveys they are given.

First, surveys of experts (e.g., US-SPF) provide respondents with contextual information about the level of inflation for the period immediately before the one to be forecast. Surveys to which lay people respond (e.g., SCE) do not do this. Second, surveys of experts provide respondents with contextual information about macroeconomic variables other than inflation for the period immediately before the one for which inflation must be forecast. For example, the SPF provides them with information about unemployment rate, GDP, interest rates (e.g., on treasury bonds), and various other indicators. Again, surveys to which lay people respond do not provide this contextual information.

Are these differences likely to matter? There are two studies potentially relevant to this question. First, Armantier, Nelson, Topa, van der Klaauw and Zafar (2016) asked lay people to make two successive forecasts of the one-year inflation rate either for the coming year or for three-years ahead. Between these two forecasts, there was a treatment phase: groups 1 and 2 first estimated the one-year ahead forecast made by professional forecasters and then were either told what that forecast was (group 1) or were not given this information (group 2); groups 3 and 4 estimated the change over the previous year in price of food and beverages and then were either told what that change was (group 3) or were not given that information (group 4). Analysis of point forecasts showed no significant differences in the size of the revisions made by groups 1 and 2 or by groups 3 and 4. However, analysis of the mean of one-year ahead pdf forecasts suggested a difference between groups 1 and 2 restricted to high-uncertainty respondents that was not attributable to accuracy with which professional forecasts were estimated. This implies that inflation estimates can be improved in some people by provision of information correlated with inflation.

Another potentially relevant study was reported by Cavallo et al. (2017). They asked people to estimate inflation rate over the previous year, then provided them with various types of information, and finally asked them for their inflation expectations for the following year. The types of information provided between the two estimates included statistical information about the inflation rate in the previous year and specific price changes for six supermarket products over that previous year. However, because Cavallo et al. (2017) were interested in learning rather than in the mechanisms underlying inflation expectations, they studied the effect of providing contextual information on *changes* in estimates of inflation across *different* years. In contrast, our experiments focus on the effects of providing different types of contextual information on inflation expectations for the *same* year. This is because our focus is on the effects of providing different information to experts and lay people when they asked about their inflation expectations in surveys.

1.4 Judgment heuristics used in forecasting depend on the nature of the information available

We know that the type of heuristics that people use to make judgments depends on both the nature of the information available to them and on the task demands (Gigerenzer and Selten, 2001; Payne, Bettman and Johnson, 1993). Harvey (2007) drew on the forecasting literature to show how this general finding extends to forecasting tasks. In other words, the information provided to forecasters influences the way in which they make their forecasts. This, in turn, can affect the quality of those forecasts.

When no external information is provided, judgmental forecasters must rely on relevant information held in memory. The availability heuristic is appropriate to such circumstances (Kahneman and Tversky, 1973). For consumers, extreme price changes are more salient and available to memory. Hence they have an inordinate influence on judgments of inflation or 'general price change' (Bruine de Bruine, van der Klaauw and Topa, 2011).

When contextual information about the levels of other variables is provided, people forecasting inflation first make broad assumptions about how these variables are related to inflation. For example, evidence summarised by Leiser and Krill (2018) suggests that lay people use the good-begets-good heuristic: they assume all indicators are positive when the state of the economy is good but all are negative when the state of the economy is poor. Hence, they assume that inflation is low when unemployment and interest rates are low. Making this assumption then enables them to use the representativeness heuristic (Kahneman and Tversky, 1973). For example, let us suppose that people are told that unemployment is 5% and they judge this to be one-third of the distance between its minimum (e.g., zero) and the maximum value it has reached over their lifetime (e.g., 15%). They then forecast that inflation will be one third of the distance between its minimum value (e.g., zero) and the maximum value it has reached over their lifetime (e.g., 15%); in other words, they expect inflation will be 5%.

When people are provided with contextual information about the level of inflation in the period immediately prior to the period for which inflation is to be forecast, they can use the anchoring heuristic (Kahneman and Tversky, 1973) to make their forecast. They would use the value of inflation they are given as a judgment anchor and then adjust away from that value to take account of any other information they may have about inflation (e.g., it is likely to rise) to produce their forecasts.

Tversky and Kahneman (1974, p 1131) emphasised that: “These heuristics are highly economical and usually effective, but they lead to systematic and predictable errors”. We know something about the errors associated with use of the availability heuristic when forecasting from information in memory: inflation expectations are a) too high because large price rises are more salient than smaller ones and b) heterogeneous because different people bring different price rises to mind (Bruine de Bruin et al., 2011).

Would we expect forecasts to improve if we gave people contextual information about other economic variables from the period prior to the one being forecast? While it is not unreasonable to expect that additional information will improve performance, it is possible that the two heuristics used for forecasting in this situation lead people further astray. Although the good-begets-good heuristic can be regarded as a lay version of the professional view that economies can be classified on a continuum from good to bad using a measure such as the ‘misery index’ (Barro, 1999), it is also possible to see how use of this heuristic could be misleading. For example, Phillips (1958) found an inverse relationship (the Phillips curve) between inflation rate and unemployment rate; in other words, low inflation (‘good’) begets high unemployment (‘bad’). However, since the 1970s, the relation described by the Phillips curve has become less clear, arguably because inflation expectations have had more of a role in determining inflation (Phelps, 1969). Hence, use of the good-begets-good heuristic may not lead people astray as much as it would have done in earlier times. However, use of the representativeness heuristic in the manner outlined above may also introduce error into inflation forecasts. Relations between inflation rate and other variables are subject to uncertainty and so we should expect some regression to the mean when using the latter to forecast the former. However, forecasts based on representativeness do not allow for this effect.

In summary, it far from clear whether providing contextual information about values of other variables for the period prior to the one for which an inflation forecast is required will facilitate performance. We do know that, compared to within-series forecasting, people find cross-series forecasting extremely difficult (Harvey, Bolger and McClelland, 1994). Hence it is possible that, if processing of the cross-series information dominates processing of information directly retrieved from memory, introduction of information about values of other macroeconomic variables on the period prior to the one for which inflation is forecast will actually impair performance.

Would inflation forecasts improve if we gave people contextual information about the value of inflation on the period immediately prior to the one for which a forecast is required? We think that they would. First, the information provides a ball-park figure for the forecast. Participants could even use the last known value of inflation as the forecast for the next period. This strategy, known as naïve forecasting is difficult to outperform in economic domains: Sherden (1998) found a) that the naïve forecast outperformed economists' forecasts for highly volatile variables, such as interest rates, b) that economists' forecasts outperformed the naïve forecast for highly stable variables, such as government spending, and that c) "Economists are about as accurate as the naïve forecast for a middle ground of important statistics, such as real GNP growth and inflation" (p 65). Thus, forecasters could produce inflation expectations comparable to those generated by macroeconomic models simply by using the value they had been given for the last period as a forecast for the upcoming period.

By using the last value for inflation as an anchor and adjusting towards the mean of the inflation series, they could allow for regression to the mean and potentially improve on the naïve forecast. The optimal amount of adjustment would depend on the autocorrelation in the inflation series. Without feedback, people tend to assume that there is a modest degree of positive first-order autocorrelation in series they are forecasting (Reimers and Harvey, 2011). However, for this strategy to work, they would need not only to know the last value of the series but also be able to obtain an estimate of the series mean.

2. Experiment 1

Lay people made a series of four inflation judgments either for the current year (inflation perception) or for the upcoming year (inflation expectation). Their first judgment was made without any additional information. They made their second forecast with provision of information about either the interest rate or the unemployment rate (randomly chosen) on the period prior to the one for which the inflation forecast was required. They made their third forecast with provision of information about the variable (either interest rate or unemployment rate) that had not been provided for the second forecast; again, this information pertained to the period immediately prior to the one for which the inflation forecast was required. They made their fourth forecast after additional information was provided about the level of inflation on the period immediately prior to the one for which the forecast was required.

For the first forecast, we expected to obtain results similar to those reported by Bruine de Bruin et al. (2011). Thus:

H_1: Mean value of inflation forecasts will be too high.

The above-mentioned findings of Armantier et al. (2016) and Cavallo et al. (2017), though obtained in paradigms not directly comparable to the present one, do imply that contextual information can improve inflation judgments in some circumstances. Thus, we expected that judgments that were made in the presence of contextual information would be better than those made when no such information was present. Hence:

H_2: Second, third and fourth inflation judgments will be more accurate than the first ones.

The fourth forecast that was given after we provided information about the level of inflation on the period immediately prior to the one for which the forecast required. For the reasons outlined above, we expected:

H_3: The fourth forecast will be more accurate and less variable than any of the earlier forecasts.

We mentioned above that forecasters' use of the anchoring heuristic to make the fourth forecast would benefit from them being provided with additional information from which they could estimate the mean value of recent inflation rates (assuming an absence of trend) and any sequential dependence between successive values of those rates. To test this, half our participants were provided with information about data from only the immediately preceding period when making forecasts 2-4 whereas the other half given information about the previous five periods before the one on which they were required to make a forecast. We expect:

H_4: The fourth forecast will be more accurate when people are given data about the previous five periods than when they are given data about just the immediately preceding period.

In Ranyard et al.'s (2018) model, experienced price changes, media reports and official statistics produce inflation perceptions via a nowcasting process. These inflation perceptions, together with expert forecasts and inferences produced by naïve models of the economy, then produce inflation expectations via a forecasting process. This implies that inputs to inflation perceptions (e.g., experienced price changes) then go on to influence inflation expectations. In line with this, Dräger (2015) found strong effects of structural shocks to inflation perceptions on inflation expectations.

This approach implies that information about official statistics (i.e., contextual information) will influence both inflation perceptions and expectations. For example, at the end of 2018, perceptions of inflation in that year will be influenced by information about the 2017 values of inflation and other macroeconomic variables in a similar way to that in which inflation expectations for 2019 generated at the end of 2018 will be influenced by information about the 2018 values of inflation and other macroeconomic variables. However, expectations are subject to more uncertainty than perceptions and so we should expect people to be less accurate and less confident when making them. Thus,

H_5: Effects of contextual information on inflation perceptions will be similar to its effects on inflation expectations but perceptions will be more accurate.

2.1. Method

2.1.1. Participants One hundred and forty-eight people (40 men, 108 women), all of whom had been living in the United Kingdom for at least two years, were recruited via the participant recruitment platform, Prolific.com. Table 4 in Appendix 1 shows their demographic characteristics. Each participant was paid £0.60 to complete the study. Data were collected between 7 March and 4 April 2020.

2.1.2. Design The experiment employed a mixed design with one within-participant variable and two between-participant variables. Contextual information was varied within participants: people first estimated UK inflation rate without any additional information, then with information about either the interest rate or employment rate (randomly chosen) in the year(s) before the one for which inflation rate was to be estimated, then with information about the variable from that pair (interest rate or employment rate) that had not previously been provided, and finally with information about the level of inflation in the year(s) immediately prior to the year for which inflation was to be estimated. Number of years (one or five) for which contextual information was provided was varied between participants: groups 1 & 2 were given one year of contextual information whereas groups 3 & 4 were given five years. Task (inflation expectation versus inflation perception) was also varied between participants: groups 1 & 3 were required to estimate the inflation rate for the year that had just ended (2019) whereas groups 2 & 4 were required to estimate it for the immediately upcoming year (2020).

2.1.3. Stimulus materials Participants made estimates of the UK inflation rate for 2019 or predictions of the inflation rate for 2020 by entering their judgments into empty cells of tables presented to them (Figure 1). Contextual information was supplied by entering values into appropriate cells in the tables for the last three inflation judgments and comprised UK historical data for base interest rates, unemployment rates, and CPI inflation rates for the years 2014 to 2019. All data used in the experiment were obtained from UK official reports published by the Office for National Statistics and the Bank of England.

Figure 1. Experiment 1: Summary task instructions followed by examples of tables ready for a) entry of the first inflation judgment in group 2 (upper panel) and b) entry of the fourth inflation judgment in group in group 3 (lower panel).

Task instructions

Please provide your **estimate for inflation (2019)** in this table by typing in the **one blank cell**, which should be computed at the annual- average level.

Please give your estimate using **two** figures after a decimal point: for example, 20.47, 14.66, or 0.00.

Economic indicators: Annual data (%)

	2018	2019
Unemployment rate (%)	No data	No data
Base Interest Rate (%)	No data	No data
CPI Inflation Rate (%)	No data	

Economic indicators: Annual data (%)

	2015	2016	2017	2018	2019	2020
Base Interest Rate (%)	0.50	0.40	0.29	0.60	0.75	No data
Unemployment rate (%)	4.40	3.80	4.20	3.20	3.10	No data
CPI Inflation Rate (%)	0.00	0.70	2.70	2.50	1.80	

Procedure After people had been informed about the nature of the study, been given details of the ethical permission that it had received, and been told that they could withdraw from it at any time, they gave their consent to participate. They were then supplied with simple definitions and examples of the three economic indicators involved in the study (base interest rate, unemployment rate, CPI inflation rate). They were randomly allocated to one of the four experimental groups. For each of the four judgments that they made, they were instructed to provide the inflation judgment appropriate to their group (Figure 1)[1]. After all judgments had been completed, basic demographic details were collected (gender, age, highest level of education qualification obtained, primary academic discipline, working experience related to economics, and primary country of residence over the previous two years).

2.2. Results

Participants' data were excluded from the data analysis if any of their four inflation judgments were more than three standard deviations from the mean of that judgment. As a result, the analyses were carried out on 135 people (98 women, 37 men) who had a mean age of 34 years (SD = 10 years). Of these, 35 were in Group 1, 36 were in Group 2, 30 were in Group 3, and 34 were in Group 4.

The upper panel of Table 1 shows means and standard deviations of levels of people's raw inflation judgments in the four experimental groups. To measure errors in 2019 inflation judgments, we used the 1.8% value for the year 2019 reported by the Office for National Statistics as the correct one. To measure errors in 2020 inflation judgments, we used the forecast of 1.5% for the year 2020 that was issued by HM Treasury and based on forecasts they received from many different institutions between 1st March and 17th March 2020.

Consistent with H_1, judged inflation rates were too high (Table 1, Middle panel). Directional errors were significantly above zero on the first judgment ($t(134) = 4.72$; $p < 0.001$), the second judgment ($t(134) = 3.86$; $p < 0.001$), the third judgment ($t(134) = 4.88$; $p < 0.001$) and the fourth judgment ($t(134) = 9.85$; $p < 0.001$).

A three-way mixed analysis of variance (ANOVA) on the directional errors with Task (inflation perception, inflation expectation) and Contextual Information (one year, five years) as between-participant variables and Judgment Number (first, second, third, fourth) as a within-participant

---

[1] In this and later experiments, after participants had entered each of their inflation judgments, they gave an estimate of the likelihood that it would be within 10% of the true value. These estimates showed that people were overconfident in their inflation judgments. As this phenomenon was not our present concern, we do not report data demonstrating it here. We discuss overconfidence in inflation judgments in Niu and Harvey (2021).

variable showed only an effect of Judgment Number (F (2.32, 303.79) = 3.01; p = 0.043; *ges* = 0.009)[2].

Table 1. Experiment 1: Means and standard deviations (in parentheses) of inflation judgments, their directional errors, and their absolute errors

Judgment	Inflation perceptions for 2019		Inflation expectations for 2020		means
	One year Contextual Information (Group 1)	Five years Contextual Information (Group 3)	One Year contextual Information (Group 2)	Five years Contextual Information (Group 4)	
a) Judged level of inflation					
First	2.34(1.64)	2.59(2.91)	2.61(1.83)	3.11(3.34)	2.66(2.51)
Second	2.30(1.85)	1.76(1.32)	2.41(1.74)	2.50(2.26)	2.26(1.83)
Third	2.15(1.48)	2.37(1.57)	2.36(1.77)	2.46(1.73)	2.33(1.64)
Fourth	2.49(0.72)	2.35(0.58)	2.03(0.90)	2.15(0.60)	2.25(0.72)
means	2.32(1.48)	2.27(1.95)	2.35(1.58)	2.55(2.24)	2.38(1.79)
b) Directional error					
First	0.54(1.64)	0.79(2.91)	1.11(1.83)	1.61(3.34)	1.02(2.51)
Second	0.50(1.85)	-0.04(1.32)	0.91(1.74)	0.10(2.26)	0.62(1.83)
Third	0.35 (1.48)	0.57(1.57)	0.86(1.77)	0.96(1.73)	0.69(1.64)
Fourth	0.69(0.72)	0.55(0.58)	0.53(0.90)	0.65(0.60)	0.61(0.72)
means	0.52(1.48)	0.47(1.95)	0.85(1.58)	1.06(2.24)	0.73(1.80)
c) Absolute error					
First	1.08(1.34)	1.79(2.42)	1.50(1.52)	2.16(3.00)	1.62(2.15)
Second	1.11(1.55)	1.07(0.75)	1.35(1.41)	1.56(1.90)	1.28(1.48)
Third	1.11(1.02)	1.23(1.10)	1.44(1.32)	1.45(1.33)	1.31(1.20)
Fourth	0.89(0.42)	0.67(0.43)	0.81(0.66)	0.71(0.52)	0.78(0.52)
means	1.05(1.16)	1.19(1.51)	1.28(1.26)	1.47(1.94)	1.25(1.49)

[2] When Mauchy's test showed a deviation from sphericity, Greenhouse-Geissser corrections were used to adjust degrees of freedom. Generalised eta squared (*ges*) measured effect size (Olejnik and Algina, 2003).

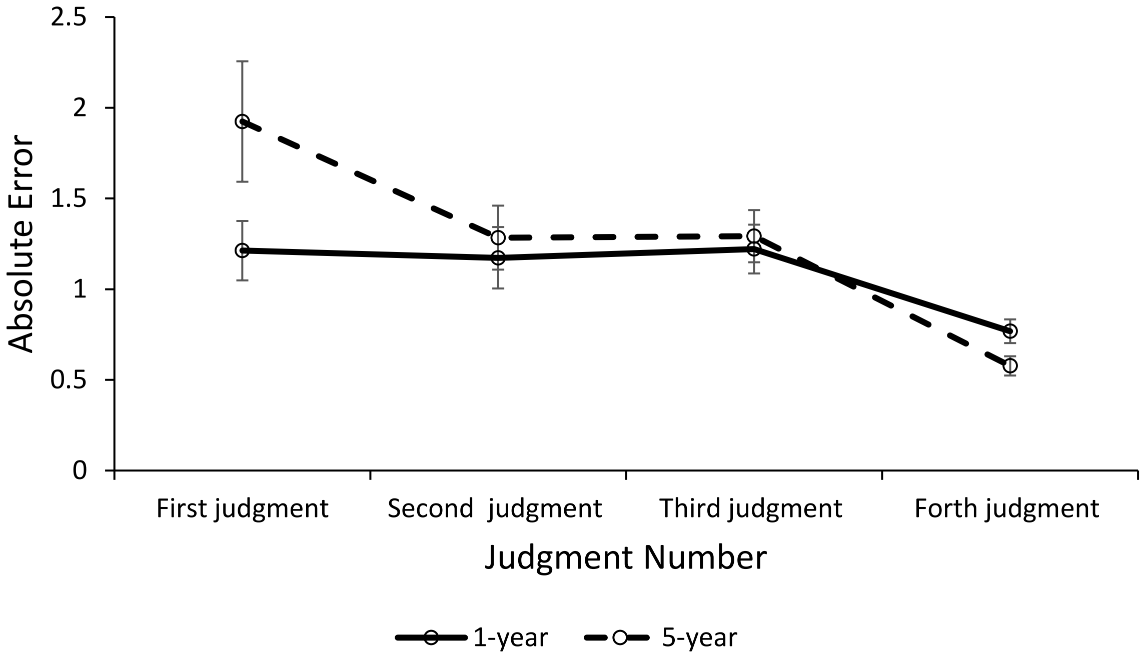
Though Bonferroni showed no differences between individual judgments, a Scheffé test showed that inflation judgments without any contextual information (Judgment 1) were higher and more biased than those with contextual information (Judgments 2, 3 and 4). The difference between these two types of judgment was -0.382 ($p < 0.032$) with a 95% family-wise confidence interval of (-0.729, -0.033). This provides evidence consistent with H_2: significantly lower judgments showing less overestimation of inflation occurred when people were given contextual information about the previous inflation rate(s).

Absolute error scores are shown in the lower panel of Table 1. A three-way mixed ANOVA using the same factors as before showed a main effect of Judgment Number (F (2.10, 275.10) = 13.16; $p < 0.001$; *ges* = 0.0434) and an interaction between Judgment Number and Contextual Information (F (2.10, 275.10) = 3.36; $p = 0.034$; *ges* = 0.011). The simple effect of Judgment Number was significant both for when there was one year of contextual information (F (2.49, 174.30) = 5.00; $p = 0.004$) and when there were five years of contextual information (F (1.79, 113.02; = 8.87; $p < 0.001$). These effects are consistent with H_2 and are shown in Figure 2.

Multiple Bonferroni pairwise comparisons showed absolute error was lower for the fourth forecast than for the first forecast (one-year information: $p = 0.02$; five-years information: $p < 0.002$), the second forecast (one-year information: $p = 0.08$; five-years information: $p = 0.006$) and the third forecast (one-year information: $p = 0.002$; five-years information: $p < 0.001$). These results provide evidence consistent with H_3: the fourth forecast was more accurate than the preceding ones.

Also consistent with H_3, provision of contextual information about previous inflation rate(s) resulted in judgments of inflation rate becoming more homogeneous. When one year of contextual information was provided, variance of the fourth judgment was significantly lower than variances of the first judgment (F (71, 71) = 4.55; $p < 0.001$), the second judgment (F (71, 71) = 4.86; $p < 0.001$) and the third judgment (F (71, 71) = 4.86; $p < 0.001$). When five years of contextual information were provided, variance of the fourth judgment was significantly lower than variances of the first judgment (F (64, 64) = 28.18; $p < 0.001$), the second judgment (F (64, 64) = 9.84; $p < 0.001$) and the third judgment (F (64, 64) = 7.84; $p < 0.001$). Furthermore, the variances of both the second judgment (F (64, 64) = 2.87; $p < 0.001$) and the third judgment (F (64, 64) = 5.46; $p < 0.001$) were lower than that of the first judgment.

Figure 2. Experiment 1: Interaction between Contextual Information and Judgment Number in the analysis of absolute error (together with standard error bars).



To test H_4, we examined the simple effect of contextual information on the fourth forecast. This showed only marginal evidence for the claim that absolute error for that forecast would be lower when five years of contextual information were provided than when just one year of context information was given (F (1, 133) = 3.02; $p = 0.085$). However, H_4 is a directional hypothesis: it can be argued that the two-tailed F-test is inappropriate for testing it. A one-tailed t-test (t (133) = 1.74; $p < .05$) suggests that, for this judgment (only), people are indeed more accurate when they are given data about the previous five periods than when they are given data about just the immediately preceding period.

2.3. Discussion

Judged inflation rates were too high (H_1) They also showed a high degree of heterogeneity. However, contextual information lowered them and made them more homogeneous (H_2). Nevertheless, they remained somewhat too high. Provision of contextual information about the preceding level(s) of inflation was more beneficial than providing contextual information about earlier levels of other macroeconomic indicators (H_3). There was also some evidence that the beneficial effect of providing information about the levels of inflation in each of the previous five years was greater than that of providing information about the level of inflation just for the immediately preceding year (H_4).

Before discussing the implications of these findings, we need to address our failure to obtain evidence consistent with H_5. We had expected that judgments reflecting people's perceptions of current inflation rate (2019) would be more accurate and be made with greater confidence than judgments reflecting their expectations of future inflation rate (2020). This was because people have more and better information about factors influencing the former (e.g., price of past purchases, reports of measured inflation and other indicators) than about those influencing the latter (e.g., price of future purchases, reports of uncertain forecasts of inflation and other indicators).

3. Experiment 2

In Experiment 1, different groups of people judged current inflation for 2019 and expected inflation for 2020. The distinction between the perception and expectation tasks was not made salient to either group. People performing these different tasks may have used very similar procedures to estimate the required inflation rate but, not being aware of the other task, may have failed to make allowances for the quality of and the uncertainty in the data on which they were basing their estimates. If we make people aware of the difference between the two tasks, they may respond differently to them. This reasoning provided the rationale for Experiment 2.

3.1. Task context

Different surveys ask people to estimate inflation for different combinations of years. The MSC asks people to estimate the percent increase in prices over the next 12 months and to estimate the average percent increase over the next five to 10 years. The SCE asks for percentage estimates of inflation over the period between the present and a date 12 months later and over the period between a date 24 months from the present and a date 36 months from the present. The IAS asks people to estimate change in prices over the last 12 months, over the next 12 months, over the 12 months after that, and over the longer term (five years). The US-SPF asks experts for their estimates of inflation rate for the current year and the two following years. The EU-SPF solicits experts' views on inflation rate for the current year and the two following years. All these surveys obtain inflation estimates for different years from the *same* respondents. This may be the reason those surveys produce different estimates from different years. Experiment 1 suggests that, had they used different respondents to obtain inflation estimates for different years, the differences between those estimates would have been much reduced.

It is easier to appreciate important differences between two options when they are evaluated jointly than when they are evaluated separately. In Hsee's (1996) task, people evaluated two dictionaries. Dictionary A was published in 1993, had 10,000 entries, and was as new with no defects. Dictionary B was published in 1993, had 20,000 entries, but had a torn cover. Participants were told that they needed a dictionary and planned to spend between $10 and $50 on one. In the separate evaluation condition, they were told that there was just one dictionary in the store, were given the details of either dictionary A or B, and decided how much they would pay for it. In the joint evaluation condition, they were told there were two dictionaries in the store, were given details of both dictionaries A and B, and decided how much they would pay for each of them. In separate evaluation, people were willing to pay $24 for A but only $20 for B. However, in joint evaluation, they were willing to pay only $19 for A but $27 for B. In joint evaluation, the difference in the important feature (i.e., number of entries) was made more salient. Other studies have replicated this evaluability effect (e.g., Hsee, Loewenstein, Blount and Bazerman, 1999).

In Experiment 1, people evaluated current and future inflation rates separately. Important differences between inflation perception and inflation expectation were not made salient. In Experiment 2, participants evaluated current and future inflation rates together by providing their estimates of inflation for 2019 and 2020 on the same screen. We anticipated that this would make the differences between the two tasks more salient and that people would better understand the different factors influencing each one. As a result, they should weight factors more heavily in perception than in expectation judgments when those factors are better predictors of current than future inflation (e.g., recent price rises). Hence,

H_6: Judgments of current inflation will be more accurate than those of future inflation.

3.2. Method

The experiment was similar to the previous one except that current and future inflation rates were jointly rather than separately evaluated.

3.2.1. Participants Eighty-seven people (24 men, 63 women), all of whom had been living in the United Kingdom for at least two years, were recruited via the participant recruitment platform, Prolific.com. Table 4 in Appendix 1 shows their demographic characteristics. Each participant was paid £0.60 to complete the study. Data were collected between 19 August and 20 August 2020.

3.2.2. Design The design was the same as that used for Experiment 1 except that Task (inflation perception versus inflation expectation) was a within-participant variable instead of a between-participant variable. Thus, Task and Judgment Number were within-participant variables and Contextual Information was a between-participant variable. Participants were randomly allocated to Group 1/2 (one year of contextual information) or Group 3/4 (five years of contextual information).

Figure 3. Experiment 2: Response tables ready for a) entry of the first inflation judgment in group 1 (upper panel) and b) entry of the fourth inflation judgment in group in group 3 (lower panel).

Task instructions

Please provide your **estimates for inflation (2019 and 2020)** in this table by typing in the **two blank cells**, which should be computed at the annual- average level.

Please give your estimates using **two** figures after a decimal point: for example, 20.47, 14.66, or 0.00.

Economic indicators: Annual data (%)

	2018	2019	2020
Unemployment rate (%)	No data	No data	No data
Base Interest Rate (%)	No data	No data	No data
CPI Inflation Rate (%)	No data		

Economic indicators: Annual data (%)

	2014	2015	2016	2017	2018	2019	2020
Base Interest Rate (%)	0.50	0.50	0.40	0.29	0.60	No data	No data
Unemployment rate (%)	5.50	4.40	3.80	4.20	3.20	No data	No data
CPI Inflation Rate (%)	1.50	0.00	0.70	2.70	2.50		

3.2.3. Materials The screen into which participants entered their responses was similar to the one used for Experiment 1 except that they filled in two empty cells, one for 2019 and one for 2020. There was no constraint on the order of responding. Examples of the response screen are shown in Figure 3.

3.2.4. Procedure The procedure was identical to that used in Experiment 1.

3.3. Results

Participants' data were excluded from the data analysis if any of their four inflation judgments were more than three standard deviations from the mean of that judgment. As a result, the analyses were carried out on 76 people (53 women, 23 men) who had a mean age of 34 years (SD = 11 years). Of these, 38 were in Group 1/2 and 38 were in Group 3/4.

Table 2. Experiment 2: Means and standard deviations (in parentheses) of inflation judgments, their directional errors, and their absolute errors

Judgment	Inflation perceptions for 2019		Inflation expectations for 2020		means
	One year Contextual Information (Group 1)	Five years Contextual Information (Group 3)	One Year contextual Information (Group 2)	Five years Contextual Information (Group 4)	
a) Judged level of inflation					
First	3.08 (3.57)	4.23(5.24)	3.54(4.37)	5.35(6.10)	4.05(3.53)
Second	2.44(1.94)	3.51(4.06)	2.88(3.03)	4.48(4.55)	3.33(2.55)
Third	2.59(2.36)	2.90(2.32)	3.16(3.39)	3.40(3.06)	3.01(2.03)
Fourth	2.44 (0.73)	2.55(0.66)	2.63(1.74)	3.00(2.21)	2.66(1.07)
means	2.64(2.38)	3.30(3.53)	3.05(3.27)	4.06(4.25)	3.26(3.47)
b) Directional error					
First	1.28(3.57)	2.43(5.24)	2.04(4.37)	3.85(6.10)	2.40(3.53)
Second	0.64(1.94)	1.71(4.06)	1.38(3.03)	2.98(4.55)	1.68(2.55)
Third	0.79(2.36)	1.10(2.32)	1.66(3.39)	1.90(3.06)	1.36(2.03)
Fourth	0.64(0.73)	0.75(0.66)	1.13(1.74)	1.50(2.21)	1.01(1.07)
means	0.84(2.38)	1.50(3.53)	1.55(3.27)	2.56(4.25)	1.61(3.48)
c) Absolute error					
First	1.53(3.47)	3.09(4.87)	2.48(4.13)	4.39(5.72)	2.87(3.32)
Second	1.04(1.76)	2.22(3.80)	1.90(2.73)	3.39(4.24)	2.14(2.36)
Third	1.27(2.13)	1.60(1.99)	2.21(3.05)	2.36(2.71)	1.86(1.80)
Fourth	0.79(0.56)	0.83(0.56)	1.51(1.42)	1.88(1.89)	1.25(0.90)
means	1.16(2.23)	1.93(3.26)	2.02(2.99)	3.00(3.92)	2.03(3.26)

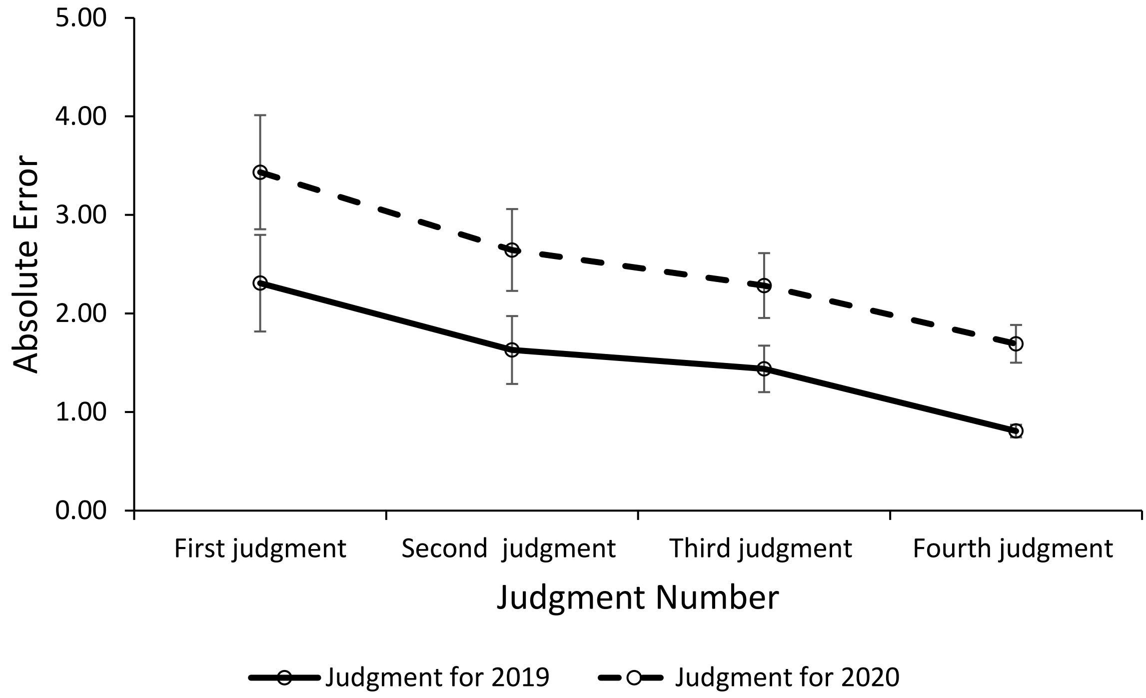
The upper panel of Table 2 shows means and standard deviations of levels of people's raw inflation judgments in the four experimental groups. To measure errors in 2019 inflation judgments, we used the same criteria for correctness as before.

As in Experiment 1, judged inflation rates were too high: directional errors were significantly above zero on the first judgment (t (75) = 3.60; p = 0.001), the second judgment (t (75) = 3.19; p = 0.002), the third judgment (t (75) = 3.53; p = 0.001) and the fourth judgment (t (75) = 8.76; p < 0.001).

A three-way mixed ANOVA on the directional errors with Contextual Information (one year, five years) as a between-participant variable and Judgment Number (first, second, third, fourth) and Task (inflation perception, inflation expectation) as within-participant variables revealed a main effect of Task (F (1, 74) = 14.04; p < 0.001; *ges* = 0.0170). In contrast to Experiment 1, overestimation was significantly greater for expected inflation in 2020 than for perceived inflation in 2019. There was also a main effect of Judgment Number (F (1.85, 136.75) = 5.31; p =0.007; *ges* = 0.0226). As the middle panel of Table 2 shows, directional error decreased over the four judgments.

Absolute error scores are shown in the lower panel of Table 3. A three-way mixed ANOVA using the same factors as before showed a main effect of Task (F (1, 74) = 19.29; p < 0.001; *ges* = 0.0235). Thus, consistent with H_7, inflation perception was more accurate than inflation expectation in this experiment. There was also a main effect of Judgment Number (F (1.80, 133.20) = 7.46; p = 0.001; *ges* = 0.0337). Post-hoc comparisons showed significant differences between the first judgment and the second judgment (p = 0.03), the third judgment (p < 0.003), and the fourth judgment (p < 0.001), between the second judgment and the fourth judgment (p = 0.002), and between the third judgment and the fourth judgment (p = 0.001). Thus, provision of contextual information again improved judgment but, in contrast to Experiment 1, this effect was shown not only by the last judgment being better than the three earlier ones but also by the first judgment being worse than the three later ones. In other words, inflation judgments were helped by providing people with past information about macroeconomic variables other than inflation but were helped even more by giving them information about previous values of inflation (Figure 4).

Figure 4. Experiment 2: Effects of Task and Judgment Number on absolute error (together with standard error bars).



The upper panel of Table 2 indicates that, as in Experiment 1, provision of contextual information resulted in judgments of inflation rate becoming more homogeneous. Mean variance of the fourth judgment was lower than that of the third judgment ($F (75, 75) = 4.00$; $p < 0.01$), the second judgment ($F (75, 75) = 6.88$; $p < 0.01$) and the first judgment ($F (75, 75) = 13.61$; $p < 0.01$), mean variance of the third judgment was lower than that of the second judgment ($F (75, 75) = 1.72$; $p < 0.025$) and the first judgment ($F (75, 75) = 3.40$; $p < 0.01$), and mean variance of the second judgment was lower than that of the first judgment ($F (75, 75) = 1.98$; $p < 0.05$).

3.4. Discussion

Use of joint evaluation was effective in rendering the difference between the inflation perception and inflation expectation tasks salient. As expected, the former was now performed more accurately than the latter. Also, as in Experiment 1, contextual information reduced absolute error in judgments. As Figure 4 shows, this was evidenced by lower judgment error when information about past values of macroeconomic variables other than inflation were provided relative to when no information was provided and lower judgment error when information about past values of inflation were provided relative to error when information about past values of macroeconomic variables other than inflation were provided.

The comparison of these two experiments shows that people's judgments of inflation for one year and the following one were influenced not just by information context (the information given to them about inflation and other macroeconomic indicators in previous years) but also by task context (asking them to provide those judgments for just one year or for more than one year).

In all major surveys, people make joint rather than separate evaluations of inflation rates in different years: estimates of inflation for one or more later years are made in the context of already having made an estimate of inflation for at least one earlier year. As a result, people's expectations about how inflation changes from one year to the next influence their judgments of inflation for later years. Our results imply that people expect inflation to increase over time, even when it does not do so. (Compare the bottoms rows of the upper panels of Tables 1 and 2.)

4. Experiment 3

There is one final issue that needs to be resolved. The experiments have shown that, relative to when no contextual information is provided, judgment error was lower when people are given information about past values of macroeconomic variables other than inflation (Figure 4). Furthermore, relative to when information about past values of macroeconomic variables other than inflation is provided, judgment error was lower when people are given information about past values of inflation (Figures 2 & 4). The issue is whether these improvements occurred a) because people had received *more* information when making later judgments than when making earlier ones, or b) because they had received *more useful* information when making later judgments than when making earlier ones. Our data already support the latter proposition. In neither experiment was judgment accuracy higher on the third judgment than on the second one. In other words, providing more information about the past values of an additional macroeconomic variable had no effect. It was only when more useful information in the form of past values of inflation was provided on the fourth judgment that an additional improvement in accuracy was observed in both experiments.

To provide additional support for this interpretation, we carried out an experiment that varied contextual information between participants. Each of three groups was given a single type of contextual information and so better accuracy in one of them could not arise because that group had more information but only because it had more useful information.

4.1. Method

The experiment was similar to the Experiment 1 except that contextual information was varied between participants. There were four groups of participants, each of which made judgments for both 2019 and 2020. Within each set, each group was given just one of four different types of contextual information: no contextual information; base interest rate information for the preceding five years; unemployment rate information for the preceding five years; inflation rate (CPI) information for the preceding five years. Thus, if accuracy is found to be higher in the fourth group than in the second and third group, it cannot be because participants in that group had more information than those in the second and third groups. It would have to be because participants in that group had more useful information than those in other groups.

4.1.1. Participants

Three hundred and fifty-two people (108 men, 244 women), all of whom had been living in the United Kingdom for at least two years, were recruited via the participant recruitment platform, Prolific.com. Table 4 in Appendix 1 shows their demographic characteristics. Each participant was paid £0.22 to complete the study. Data were collected between 4 September and 14 November 2020.

Figure 5. Experiment 3: Response tables for inflation judgments in a) group without information provided (upper panel) and b) group with the unemployment rate provided (lower panel).

Task instructions

Please provide your **estimates for CPI inflation rate (2019 and 2020)** in this table by typing in the **two blank cells**, which should be computed at the annual- average level.

Please give your estimates using **two** figures after a decimal point: for example, 20.47, 14.66, or 0.00.

Economic indicators: Annual data (%)

	2019	2020
CPI Inflation rate (%)		

Economic indicators: Annual data (%)

	2014	2015	2016	2017	2018	2019	2020
Unemployment rate (%)	5.50	4.40	3.80	4.20	3.20	No data	No data
CPI Inflation Rate (%)	No data	No data	No data	No data	No data		

4.1.2. Design Contextual information (the four types specified above) was a between-participant variable and Year (judgments for 2019 and 2020) was a within-participant variable. Participants were randomly allocated to one of the four experimental groups.

4.1.3. Stimulus materials The screen into which participants entered their responses was similar to the one used for Experiment 1 except that each of them responded to just one table by entering their judgments for 2019 and 2020. There was no constraint on the order of responding. Examples of the response screen are shown in Figure 5.

When information was provided, it was for five years starting at 2014 and ending at 2018. For interest rate information, the values were 0.50, 0.50, 0.40, 0.29, and 0.60. For unemployment rate information the values were 5.50, 4.40, 3.80, 4.20, and 3.20. For inflation rate information, the values were 1.50, 0.00, 0.70, 2.70, and 2.50.

4.1.4. Procedure The procedure was identical to that used in Experiment 1.

4.2. Results

In this between-participants experiment, Levene's test showed that the ANOVA assumption of homogeneity of variances was violated ($p < 0.05$) at each level of Year for all three dependent variables (judgment score, directional error score, absolute error score). (This was true even after outliers more than three standard deviations from the mean had been excluded.) Hence, we carried out a robust two-way mixed ANOVA (Wilcox, 2017) on each dependent variable using Information Type as a between-participants factor and Year as a within-participants factor. Data were analysed in R using robust tests on 20% trimmed means (to reduce skew) and a bootstrap procedure (*nboot* = 2000)[3] to obtain empirically-derived critical values ($p < 0.05$) against which test statistics were compared.

As the robust analyses trim means, these ANOVAs were performed on the complete data set (n = 352) with no outlier exclusion: 89 people (32 men, 57 women) with a mean age of 32 years (SD = 11.23 years) were in the group without additional information, 86 people (27 men, 59 women) with a mean age of 31 years (SD = 10 years) were in the group with interest rate information, 86 people (23 men, 63 women) with a mean age of 33 years (SD = 10 years) were in the group with unemployment rate information, and 91 people (26 men, 65 women) with a mean age of 33 years (SD = 11 years) were in the group with inflation rate (CPI) information. Means and standard deviations of the three dependent variables in each of the four conditions are shown in Table 3.

Analysis of directional error scores using the same factors as before revealed main effects of Information Type ($Q = 23.70$, $p < 0.001$) and Year ($Q = 27.53$, $p < 0.001$) but no interaction between

[3] In this section, terms in italics refer to R functions in Wilcox (2017).

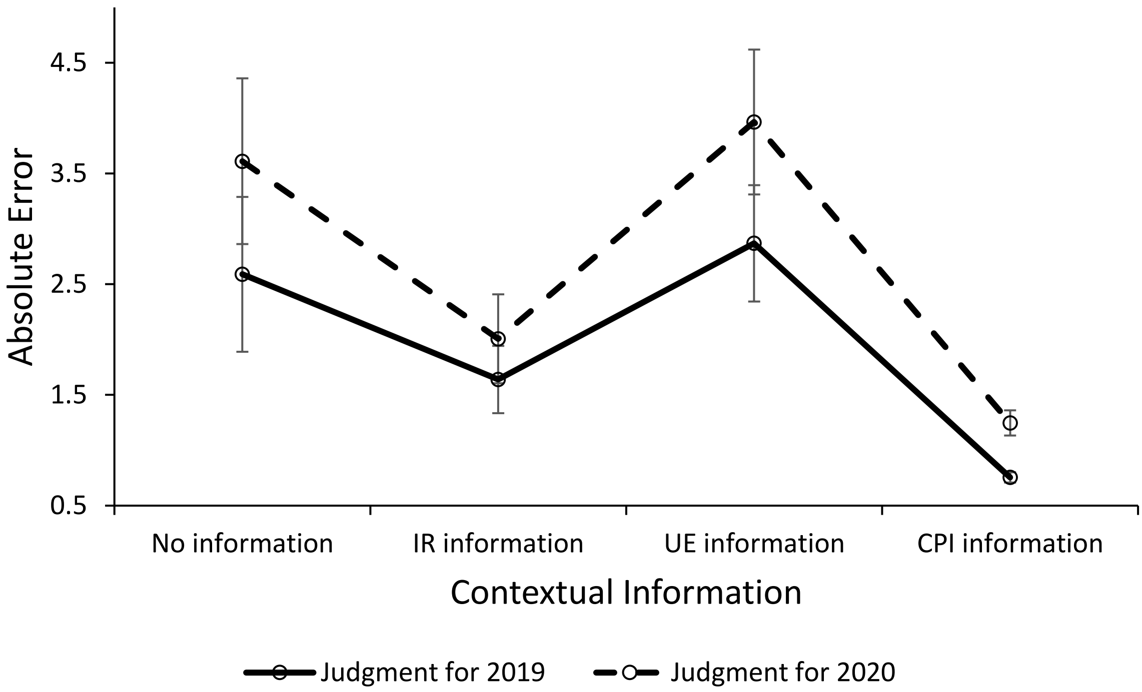
these variables. Post hoc analyses between each pair of Information Types revealed significant differences in every case (p < 0.05).

Table 3. Experiment 3: Means and standard deviations (in parentheses) of inflation judgments, their directional errors, and their absolute errors

Judgment	No information	IR information	UE information	CPI information	means
a) Judgment level of inflation					
2019	3.93(6.76)	2.09(3.25)	4.28(5.09)	2.46(0.57)	3.18(4.61)
2020	4.65(7.28)	2.50(4.12)	5.05(6.32)	2.34(1.43)	3.62(5.39)
means	4.29(7.03)	2.30(3.71)	4.67(5.74)	2.40(1.09)	3.40(5.01)
b) Directional error					
2019	2.13(6.76)	0.29(3.25)	2.48(5.09)	0.66(0.57)	1.38(4.61)
2020	3.15(7.28)	1.00(4.12)	3.55(6.32)	0.84(1.43)	2.12(5.39)
means	2.64(7.03)	0.64(3.71)	3.01(5.74)	0.75(1.09)	1.75(5.02)
c) Absolute error					
2019	2.59(6.60)	1.64(2.82)	2.87(4.87)	0.76(0.44)	1.95(4.40)
2020	3.61(7.06)	2.01(3.72)	3.97(6.07)	1.25(1.09)	2.69(5.13)
means	3.10(6.83)	1.82(3.30)	3.42(5.50)	1.00(0.83)	2.32(4.79)

The same type of analysis performed on absolute error scores (Figure 6) revealed a main effect of Information Type ($Q = 4.20$, $p < 0.02$), a main effect of Year ($Q = 30.36$, $p < 0.001$), and an interaction between these variables ($Q = 7.47$, $p = 0.006$). The simple effect of Year at each level of Information Type was examined using Wilcox's (2017) *ydbt* function to extract bootstrap confidence intervals: absolute error was significantly greater for 2020 inflation judgments than for 2019 inflation judgments when no information was supplied ($p = 0.038$), when past unemployment rates were provided (p = 0.003), and when past CPI information was given (p < 0.001). Wilcox's (2017) *t1waybt* bootstrap function showed a simple effect of Information Type for both 2019 (Ft = 6.77, p < 0.001) and 2020 (Ft = 6.01, p = 0.01). Follow-up post hoc tests using Wilcox's (2017) *linconb* function revealed that, for 2019, all paired comparisons were significant (p < 0.05) except that between no information and CPI information and, for 2020, they were all significant except for the comparisons between no information and CPI information and between no information and interest rate information.

Figure 6. Experiment 3: Effects of Year and Information Type on absolute error (together with standard error bars).



4.3. Discussion

In this between-participants experiment, absolute error was again higher for 2020 judgments than it was for 2019 judgments. This replicates the effect that we obtained in Experiment 2, where, as in the current experiment, the same participants made judgments for both 2019 and 2020. These effects of Year found in Experiments 2 and 3 contrast with the lack of such an effect in Experiment 1, where participants made judgments for just a single year – some made them for 2019 and others made them for 2020. Thus results here are consistent with a task context effect: when people make judgments for two years, their first judgment provides a context for, and thereby influences, their second one. For example, people viewing inflation as generally increasing over time will ensure that their judgment of its value for next year is higher than their judgment of its value in the current year (Tables 2 and 3). In contrast, when people make judgments for a single year, no task context effect can operate: as a consequence, inflation judgments made by people producing judgments for only next year are no different from those made by people producing judgments for just this year (Table 1).

There was again an effect of information context: the type of information given to forecasters influenced their inflation judgments. Specifically, when people were given information about past values of inflation, their estimates of the values of inflation later in the series were better than when they were given past values of other macroeconomic variables, such as interest rates or unemployment rates (Figure 6). This indicates that the effect of providing inflation rate information in the fourth judgments of Experiments 1 and 2 arose not (or not only) because people received *more* information when making later judgments in those experiments but because they receive *more useful* information when making later judgments in those experiments.

Why was information about past inflation rates more useful for judging current and future inflation rates than information about past values of interest rates and unemployment rates? Clearly it was more relevant – but how did that higher relevance impact on people’s judgments? When given past inflation rates, participants could either use the last value to produce a naïve forecast for inflation or they could extrapolate from any perceived trend in the series to produce an inflation forecast. However, when given past values of interest rates or unemployment rates, neither of these strategies would have been appropriate for producing judgments about inflation.

To use such information effectively, they would have had to make use of a mental model of the economy that was at least approximately correct. But, as Leiser and Krill (2018) have shown, they do not do this. One possibility is that they use a good-begets-good heuristic by assuming that when interest rates and unemployment rates are low, inflation is also low. This could explain why people judged inflation to be low (2.09 – 2-50 %) when they were told that interest rates were low (0.29 – 0.6%) but why the judged inflation rate to be moderate (4.28 – 5.05%) when they were told that unemployment rates were moderate (3.20 – 5.50%). This pattern of results is also consistent with Kahneman and Tversky's (1973) account of how the representativeness heuristic is used in cross-series forecasting. It may also be explained by an anchoring effect: higher judgment anchors (unemployment rates) produced higher judgments of inflation than low ones (interest rates).

5. General discussion

Inflation judgments were systematically too high, a finding that replicates what has been found in previous studies using lay participants (Bruine de Bruin, van der Klaauw and Topa, 2011; Bryan and Venkato, 2001a, b; Georganas, Healy and Li, 2014). When different people made inflation judgments for the current year or for the following year, mean values of these judgments did not differ (Experiment 1) but when the same people made judgments for both those two years, inflation judgments for 2020 were higher than those for 2019 (Experiment 2). This task context effect, triggered by joint evaluation, implies that people (wrongly) expected inflation rate to increase over time. As a result, inflation expectations for 2020 were worse than inflation perceptions for 2019 in Experiments 2 and 3.

Information context effects were found in both experiments though their nature differed somewhat. In Experiment 1, the fourth judgment, the only one that benefitted from provision of the inflation rate in the year immediately prior to the year for which inflation rate had to be estimated, was more accurate than the three earlier judgments. In Experiment 2, the fourth judgment was again superior to the previous three judgments but, in addition, the first judgment was less accurate than the three later judgments. It is likely that this difference is related to the fact that, for the fourth judgment in Experiment 2, information about the immediately preceding inflation rate could be provided only for inflation judgment for 2019; it could not be provided for the inflation judgment for 2020 because participants provided it themselves when estimating the inflation rate for 2019. In contrast, for the fourth judgment in Experiment 1, information about the immediately preceding inflation rate was explicitly provided for the inflation judgments of both 2019 and 2020. (Compare the lower panels of Figures 1 and 5.)

This suggests that Experiment 1 provides a purer comparison of the difficulties in using (and benefits arising from) the heuristics responsible for cross-series forecasting (second and third judgments) and within-series forecasting (fourth judgment). Cross-series forecasting, reliant on use of the representativeness (Harvey, 2007) and good-begets-good (Leiser and Krill, 2018) heuristics, is difficult and often ineffective (Harvey et al., 1994): comparison of the second and third judgments with the first judgment shows that it produced little improvement over memory-based forecasting. In contrast, within-series forecasting, based on the anchor-and-adjust heuristic (Harvey, 2007) or on knowledge of temporal patterns in the ecology (Harvey and Reimers, 2013; Reimers and Harvey, 2011), is more effective: comparison of the fourth judgment with the first three judgments shows the advantages it has over memory-based and cross-series forecasting.

Information context also influenced degree of judgment homogeneity. Thus, in Experiment 1, variance of the fourth judgments was lower than that of each of the three earlier judgments and, when five years of contextual information was provided, variances of the second and third judgments were lower than the variance of the first judgment. In all conditions of Experiment 2, variance of the fourth judgment was significantly lower than that of the other three judgments and variances of the second and third judgments was lower than variance of the first judgment.

5.1. Potential limitations

These experiments were conducted during a period when economic life was disrupted by the Covid-19 pandemic. It is possible that reports of its effects in the media made laypeople more aware of economic indicators than they would normally be. If so, we might expect their inflation judgments to change with the onset of the epidemic. In fact, households' inflation expectations did not exhibit a clear upward or downward change after the emergence of the pandemic (Armantier, Koşar, Pomerantz, Skandalis, Smith, Topa, & Van der Klaauw, 2020; Ebrahimy, Igan, and Peria, 2020). Furthermore, according to the Monetary Policy Report from Bank of England (2021), the Monetary Policy Committee judged that inflation expectations remained well anchored. Thus, the biased 2020 inflation rate judgments obtained from our samples are unlikely to reflect responses to economic effects of the pandemic.

It is possible, though unlikely, that participants searched the Internet for information about inflation rates. Current and past inflation rates are more easily and more quickly found on the Internet than estimates for future inflation rates. If some participants in the groups that were not provided with additional information did retrieve past inflation rate information in this way, their actions would have reduced the difference between the groups. As a result, the effects that we have reported would not have been found or would have been diminished in size. Similarly, if people had retrieved predictions for future inflation, their overestimation of future inflation rates would not have been found or would have been diminished. In summary, internet retrieval of inflation rates would not have acted to produce the effects that we obtained but would have counteracted those effects.

Demographic factors, including gender, education, and financial literacy are known to influence inflation judgments (Bruine de Bruin, van der klaauw, Downs, Fischhoff, Topa, & Armantier, 2010; Souleles, 2004). Differences in demographic characteristics could therefore potentially explain differences between results obtained in different experiments (including the task context effect revealed by the difference between the first and second experiment). In fact, as Table 4 in Appendix

1 shows, the demographic characteristics of the samples in the three experiments were highly comparable.

5.2. Implications

In surveys, lay respondents produce inflation estimates that are higher and more heterogeneous than those of experts (Mankiw et al., 2003; Palardi and Ovaska, 2015). These differences may occur because lay people and experts retain different inflation-relevant information in their memories arising from their access to different data, from variation in how much they attend to their personal experience of price changes, and from differences in their knowledge of macroeconomic processes. We agree that these factors may indeed be responsible for differences in judgments of inflation rate. However, our work leads us to question whether they have been responsible for the differences in the level and heterogeneity of inflation judgments obtained from surveys of lay and expert respondents. We have shown that lay people who are given the same type of information that experts are given in surveys produce lower, more accurate, and less heterogeneous inflation estimates. We cannot say that this information context effect would completely cancel out the lay-expert differences that have been reported but we would expect it to reduce them.

Why are surveys different for lay people and experts? Presumably, there is an assumption that lay people who are considering some economic behaviour (purchasing, saving, negotiating a pay rise) do not make reference to records of the past macroeconomic indicators that are given to experts in US-SPF, EU-SPF and other expert surveys. Instead, they are assumed to make memory-based judgments just like they are required to do in MSC, IAS, SCE and other lay surveys. In other words, surveys are designed to reflect the normal information ecology of their intended respondents. If surveys are intended as an aid to predicting behaviour of respondents in their natural environments, this design strategy has much to recommend it. However, it does mean that we should be cautious in making direct comparisons between lay and expert survey responses.

For central banks, importance of understanding inflation expectations of lay people outweighs that of experts. If, when surveying lay people, we were to provide them with the additional information that experts are given in their surveys, lay inflation expectations might become as good as those produced by experts. However, as they do not normally have that additional information when they make the economic decisions that influence inflation rates, those more accurate expectations would not then supply central banks with the information that they need to predict people's economic behaviours and the effects of those behaviours on inflation. Whether the same information should be given to respondents in expert and lay surveys remains an open question.

Task context effects also have implications. When people judged inflation rates for two successive years (Tables 2 and 3), their estimate for the later year was higher and less accurate than it was when they made a single judgment for that later year (Table 1). In other words, they did not make their judgment for the later year in the same way that they made it for the earlier year. Instead of making their judgment using only their memory and the contextual information they were given, they were also influenced by their expectation about how inflation would change from one year to the next. Expectations about how inflation is going to change over time adds another potential source of error to judgments of inflation. Currently, all major surveys require respondents to judge levels of inflation for a number of different years. Their responses, especially for later years, would be likely to be more accurate if they were asked for their estimate for a single year, with different respondents supplying estimates for different years.

5.2. Conclusions

We have shown how inflation judgments are influenced by the information context and the task context in which they are embedded. These effects have implications for how we should think about reported differences in accuracy and heterogeneity between inflation judgments made by expert and lay respondents. These differences are likely to arise at least partly from the differences in the format of the surveys designed for those different groups.

We have documented just two types of context effects effect. Our findings will not come as a surprise to those social scientists who, for some decades, have documented context effects in both traditional (e.g., McFarland, 1981; Schuman, Kalton and Ludwig, 1983; Schwarz and Sudman, 1992) and online surveys (e.g., Reips, 2002; Smyth et al., 2009). Indeed, from their work, they would expect that a number of other context effects remain to be identified in inflation surveys.

References

Armantier, O., Bruin, W. B. d., Potter, S., Topa, G., Klaauw, W. v. d., & Zafar, B. (2013). Measuring Inflation Expectations. *Annual Review of Economics, 5(1),* 273-301. doi:10.1146/annurev-economics-081512-141510

Armantier, O., Koşar, G., Pomerantz, R., Skandalis, D., Smith, K., Topa, G., & Van der Klaauw, W. (2020). *Inflation Expectations in Times of COVID-19* (No. 20200513). Federal Reserve Bank of New York.

Armantier, O., Nelson, S., Topa, G., Van der Klaauw, W., & Zafar, B. (2016). The price is right: Updating inflation expectations in a randomized price information experiment. *Review of Economics and Statistics, 98(3),* 503-523.

Armantier, O., Topa, G., van Der Klaauw, W., & Zafar, B. (2017). An overview of the survey of consumer expectations. *Economic Policy Review - Federal Reserve Bank of New York, 23*(2), 51-III.

Bank of England. (2021). The Monetary Policy Report: February 2021. https://www.bankofengland.co.uk/-/media/boe/files/monetary-policy-report/2021/february/monetary-policy-report-february-2021.pdf

Barro, R. (1999). Reagan Vs Clinton: Who's the economic champ? *Bloomberg*, 22 February.

Bates, J. M. & Gabor, A. (1986). Price perception in creeping inflation: The results of an enquiry. *Journal of Economic Psychology, 7,* 291-314.

Binder, C., & Rodrigue, A. (2018). Household Informedness and Long-Run Inflation Expectations: Experimental Evidence. *Southern Economic Journal, 85*(2), 580-598. doi:10.1002/soej.12306

Brachinger, H. W. (2008). A new index of perceived inflation: Assumptions, method, and application to Germany. *Journal of Economic Psychology, 29(4),* 433-457.

Bruine de Bruin, W., Potter, S., Rich, R. W., Topa, G., & Van Der Klaauw, W. (2010). Improving survey measures of household inflation expectations. *Current Issues in Economics and Finance, 16*(7), 1-7.

Bruine de Bruin, W., Manski, C. F., Topa, G. & van der Klaauw. W. (2011). Measuring consumer uncertainty about future inflation. *Journal of Applied Econometrics*, *26,* 454-478. Doi: 10.1002/jae.123

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

Bruine de Bruin, W., van der Klaauw, W., & Topa, G. (2011). Expectations of inflation: The biasing effect of thoughts about specific prices. *Journal of Economic Psychology, 32*(5), 834-845. doi:https://doi.org/10.1016/j.joep.2011.07.002

Bruine de Bruin, W., van der Klaauw, W., Topa, G., Downs, J. S., Fischhoff, B., & Armantier, O. (2012). The effect of question wording on consumers' reported inflation expectations. *Journal of Economic Psychology, 33*(4), 749-757. doi:https://doi.org/10.1016/j.joep.2012.02.001

Bruine de Bruin, W., van der Klaauw, W., van Rooij, M., Teppa, F., & de Vos, K. (2017). Measuring expectations of inflation: Effects of survey mode, wording, and opportunities to revise. *Journal of Economic Psychology, 59*, 45-58. doi:https://doi.org/10.1016/j.joep.2017.01.011

Bryan, M. F. & Venkato, G. (2001a). The demographics of inflation opinion surveys. *Federal Reserve Bank of Cleveland, Economic Commentary,* 10.15.2001.

Bryan, M. F. & Venkato, G. (2001b). The curiously different inflation perspectives of men and women. *Federal Reserve Bank of Cleveland, Economic Commentary,* 11.01.2001.

Cavallo, A., Cruces, G., & Perez-Truglia, R. (2017). Inflation expectations, learning, and supermarket prices: Evidence from survey experiments. *American Economic Journal: Macroeconomics*, *9*(3), 1-35.

Coibion, O., Gorodnichenko, Y., Kumar, S., & Pedemonte, M. (2020). Inflation expectations as a policy tool? *Journal of International Economics*, *124*, 103297. doi:https://doi.org/10.1016/j.jinteco.2020.103297

Dräger, L. (2015). Inflation perceptions and expectations in Sweden – Are media reports the missing link? *Oxford Bulletin of Economics and Statistics*, *77*(5), 681-700.

Ebrahimy E., Igan D., & Peria S.M. (2020). The Impact of COVID-19 on Inflation: Potential Drivers and Dynamics (Special Notes Series on COVID-19). International monetary fund. https://www.imf.org/-/media/Files/Publications/covid19-special-notes/en-special-series-on-covid-19-the-impact-of-covid-19-on-inflation-potential-drivers-and-dynamics.ashx

Gábriel, P., Rariga, J., & Várhegyi, J. (2014). *Inflation expectations in Hungary*. Magyar Nemze Bank Occasional Papers, No. 113.

Georganas, S., Healy, P.J. & Li, N. (2014). Frequency bias in consumers' perceptions of inflation: An experimental study. *European Economic Review, 67(C),* 144-158. doi: 10.1016/j.euroecorev.2014.01.014

Gigerenzer, G. & Selten, R. (Eds.) (2001). *Bounded rationality: The adaptive toolbox.* Cambridge, MA: MIT Press.

Harvey, N. (2007). Use of heuristics: Insights from forecasting research. *Thinking & Reasoning*, *13*(1), 5-24. doi:10.1080/13546780600872502

Harvey, N. and Reimers, S. (2013). Trend damping: Under-adjustment, experimental artefact, or adaptation to features of the natural environment? *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 39, 589-607.

Harvey, N., Bolger, F. & McClelland, A. (1994). On the nature of expectations. *British Journal of Psychology, 85,* 203-229.

Hsee, C.K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluation of alternatives. *Organizational Behavior and Human Decision Processes, 67(3),* 247-257.

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin, 125(5),* 576-590.

Jungermann, H., Brachinger, H.W., Belting, J., Grinberg, K. &Zacharias, E. (2007). The euro changeover and the factors influencing perceived inflation. *Journal of Consumer Policy, 30(4),* 405-419.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237–251.

Lein, S. M., & Maag, T. (2011). The formation of inflation perceptions: Some empirical facts for European countries. *Scottish Journal of Political Economy, 58(2),* 155-188.

Leiser, D., & Krill, Z. (2018). How laypeople understand the economy. In R. Ranyard (Ed.), Economic Psychology. Chichester: Wiley (pp 139-154).

Maag, T., & Lamla, M. (2009). The role of media for inflation forecast disagreement of households and professionals. *IDEAS Working Paper Series from RePEc*. https://doi.org/10.3929/ethz-a-005788384

Madeira, C. & Zafar, B. (2015). Heterogeneous inflation expectations and learning. *Journal of Money, Credit and Banking, 47(5),* 867-896.

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

Mankiw, N. G., Reis, R., & Wolfers, J. (2003). Disagreement about inflation expectations. *NBER Macroeconomics Annual*, *18,* 209-248.

McFarland, S. G. (1981). Effects of question order on survey responses. *Public Opinion Quarterly*, *45*, 208-215.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica, 29(3),* 315-335.

Niu, X. and Harvey, N. (2021). Laypeople overestimate inflation and are overconfident in their estimates: Training reduces judgment biases. Under review.

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods, 8(4),* 434-447. http://dx.doi.org/10.1037/1082-989X.8.4.434

Palardy, J., & Ovaska, T. (2015). Decomposing household, professional and market forecasts on inflation: a dynamic factor model analysis. *Applied Economics, 47*(20), 2092-2101. doi:10.1080/00036846.2014.1002889

Payne, J. W., Bettman, J. R. & Johnson, E. J. (1993). *The Adaptive Decision-Maker*. Cambridge: Cambridge University Press.

Phelps, E. S. (1969). The new microeconomics of inflation and employment theory. *The American Economic Review, 59(2),* 147-160.

Phillips, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861-1957. *Economica, 25. No.100,* 283-299.

Ranyard, R., Del Missier, F., Bonini, N.& Pietroni, D. (2018). The citizen's judgments of prices and inflation. In R. Ranyard (Ed.), *Economic Psychology*. Chichester: Wiley. (pp 155-170).

Reimers, S. and Harvey, N. (2011). Sensitivity to autocorrelation in judgmental time series forecasting. *International Journal of Forecasting*, 27, 1196-1214.

Reips, U-D. (2002). Context effects in web-surveys. In B. Batinic, U-D Reips and N. Bosnjac (Eds.), *Online social sciences*. Seattle: Hogrefe and Huber.

Schuman, H., Kalton, G. & Ludwig, J. (1983). Context and contiguity in survey questionnaires. *Public Opinion Quarterly*, 47, 112-115

Schwarz, N., & Sudman, S. (Eds.) (1992). *Context effects in social and psychological research.* New York: Springer-Verlag.

Sherden, W. A. (1998). *The fortune sellers: The big business of buying and selling predictions.* Chichester: Wiley.

Smyth, J.D., Dillman, D.A. & Christian, L. M. (2009). Context effects in internet surveys: New issues and evidence. In A. D. Joynson, K. Y. A. McKenna, T. Postmes & U-D Reips (Eds.), *Oxford Handbook of Internet Psychology*. Oxford: Oxford University Press.

Soroka, S. N. (2006). Good news and bad news: Asymmetric responses to economic information. *The Journal of Politics*, *68(2)*, 372-385.

Souleles, N. S. (2004). Expectations, heterogeneous forecast errors, and consumption: Micro evidence from the Michigan consumer sentiment surveys. *Journal of Money, Credit and Banking*, 39-72.

Tourangeau, R., Rips, L. J. & Rasinski, K.A. (2000). *The psychology of survey response.* Cambridge: Cambridge University Press.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185, 1124-1131. doi:10.1126/science.185.4157.1124

Wilcox, R. R. (2017). *Introduction to robust estimation and hypothesis testing* (4th ed.). Academic press.

Appendix 1

Table 4. Demographical statistics for participants whose data were analysed in three experiments (percentages or standard deviations in parentheses).

		Experiment 1 (n=135)	Experiment 2 (n=76)	Experiment 3 (n=352)
Age in years		34 (10)	34 (11)	32 (11)
Gender	Men	37 (27%)	23 (30%)	108 (31%)
	Women	98 (73%)	53 (70%)	244 (69%)
Education level	School leaving exam	45 (33%)	20 (26%)	120 (34%)
	Undergraduate	63 (47%)	37 (49%)	160 (45%)
	Master	23 (17%)	18 (24%)	60 (17%)
	PhD	4 (3%)	1 (1%)	12 (3%)
Primary academic discipline in Economics	No	131 (97%)	82 (93%)	326 (93%)
	Yes	4 (3%)	5 (7%)	26 (7%)
Working experience related to economics (year)		0.36 (3.11)	0.14(0.76)	0.17(0.84)

Note: one participant in Experiment 3 did not report her age.