WZB

Wissenschaftszentrum Berlin
für Sozialforschung

Sabine Kröger
Thibaud Pierrot

Comparison of different question formats eliciting point predictions

Discussion Paper

SP II 2019–213

October 2019

WZB Berlin Social Science Center

Research Area
Markets and Choice

Research Unit
Market Behavior

Wissenschaftszentrum Berlin für Sozialforschung gGmbH
Reichpietschufer 50
10785 Berlin
Germany
www.wzb.eu

Sabine Kröger, Thibaud Pierrot
Comparison of different question formats eliciting point predictions

Affiliation of the authors:

Sabine Kröger
Laval University

Thibaud Pierrot
WZB Berlin Social Science Center and Technische Universität Berlin

Wissenschaftszentrum Berlin für Sozialforschung gGmbH
Reichpietschufer 50
10785 Berlin
Germany
www.wzb.eu

Abstract

Comparison of different question formats eliciting point predictions

by Sabine Kröger and Thibaud Pierrot*

Survey questions that elicit point predictions regarding uncertain events face an important challenge as human forecasters use various statistics to summarise their subjective expectations. In this paper, we take up the challenge and study whether alternative formulations of the questions used to elicit point predictions can be successful in driving forecasters towards reporting a particular central tendency (median or mean) of their subjective expectations distribution. We set up a laboratory experiment in which the participants act as forecasters and are asked to predict the next realisation of iid random draws coming from an objectively known distribution. We elicit the subjects' point predictions in four treatments, in which we ask for either (1) a \guess" of the next draw, as is standard in survey measures, (2) a \guess" as close as possible to the next 6 draws, and (3) the mean, or (4) the median of the next six draws. We then compare the predictions reported in the different treatments and their proximity to the three main central tendencies (mean, median, mode) of the objectively known distributions. We also investigate the cognitive process that affects the production of point predictions.
We find that the majority of predictions in the two guess treatments, (1) and (2), are close to the mode. In treatment (2) (\one guess for the next six draws"), the forecasters report the mean and the median more often in comparison to (1) (\guess for the next draw"), but the mode remains the central tendency around which most of the predictions are located. In treatments (3) and (4), we find that forecasters adjust the point they report in the direction of a particular central tendency when specifically asked to report the mean or the median. Adjustments are more precise for forecasters with higher measures of numeracy and for those who have more experience. However, numeracy has no explanatory power when the forecasters are asked to report a \guess for the next draw" in treatment (1) which suggests that forecasters have different ways to summarise a distribution.

Keywords: subjective expectations, forecasting, eliciting point predictions, experiment

JEL classification: C91; C72; D84

* E-mail: sabine.kroger@ecn.ulaval.ca, thib.pierrot@gmail.com.

1 Introduction

Data on subjective expectations is an important input for macro and microeconomic models of decision-making under uncertainty. The measurement of expectations is a challenging exercise for survey designers. There is an ongoing discussion regarding how to elicit the subjective probability distributions over all possible outcomes of a random variable of interest that the respondents have in mind (Manski, 2004). Even though an increasing number of surveys elicit subjective distributions in whole(Delavande, 2014), the most common practice remaining is to ask the survey panel to provide a summary of their subjective distribution in the form of a point forecast regarding the future realisation of the variable of interest without further indication(Pesaran and Weale, 2006).

While it is generally assumed that the point forecast obtained is the mean of the probability distribution that the respondents have in mind, there is empirical evidence that forecasters summarise their subjective expectations in various ways (Engleberg et al., 2009). Not being able to observe the individual summary rules makes it difficult to interpret the forecasts, use them in theoretical models, or even to compare between forecasters. For example, two persons with exactly the same expectations distribution regarding an event may still report different forecasts because one chooses the mean of his subjective probability distribution while the other uses its mode (Kröger and Pierrot, 2019).

The challenge for survey designers is to develop a question that elicits the same point of a distribution. The literature on scoring rules suggests using incentives to reach this goal (Savage, 1971). However, giving incentives in general and professional surveys on macroeconomic or personal variables is rather complex. In this paper, we investigate how to elicit a particular point of the subjective expectations by using different question formats for the target audience. Our laboratory study explores the effect of different ways of formulating the question on individuals' point forecasts. The participants in the experiment are asked to predict the realizations of an iid random variable that are drawn from an objectively-known distribution displayed in the form of a histogram.

The contribution of this research to the literature is that we employ four different question formats to elicit the forecasts. We ask for either (1) a “guess” regarding the next realisation of the variable of interest, (2) a “guess” as close as possible to the next six realisations, (3) a point forecast as close as possible to the mean of the next six realisations, or (4) a point forecast as close as possible to the median of the next six realisations.

In our first treatment, the participants are asked to forecast the next realisation of the random variable. This question format is the most commonly used in surveys nowadays. In the second treatment, we extend the forecasting horizon. In principle, while the first treatment may encourage a respondent to report the most likely realisation, we expect a respondent to “smooth” his summary and take other moments of the distribution into account when asked to summarise several realisations with a single forecast. This treatment also serves as a baseline to interpret the effects of our final two treatments. In the third and fourth treatments subjects are directly asked to report a point corresponding to the particular central tendency of the underlying distribution. In the third treatment we ask them to report the mean of the next six draws and in the fourth we ask for the median.

Our analysis reveals that, when asked to predict the next draw (in the first treatment), the majority of the forecasters in our experiment report the mode of the underlying distributions. In the second treatment, the mean and the median are more often reported in comparison to the first treatment but the pre-eminence of the modal report remains. The third and fourth treatments are successful in modifying the participants’ forecasting strategy. In these two treatments a majority of the predictions reflect respectively the mean or median of the underlying distribution. The forecasters are led to adjust their forecast in the direction of a particular central tendency when they are specifically asked to report the corresponding tendency for the following six realizations.

To complement the estimation of the treatment effects, we develop a model explaining the distance between the forecasts and the three main central tendencies of the underlying distributions. This model integrates individual characteristics such as cognitive ability, effort and experience to better understand the benefits and costs of the different question formats.

Our results can be summarised as follows. We observe that (i) the forecasters who exert more effort report points that are closer to the mean or the median; (ii) better cognitive ability has a similar effect as effort but only in the treatments with a longer forecasting horizon; and (iii) experience has only a marginal impact on the distances between the forecasts and the central tendencies. Finally, we note that asking for the mean of the next six realizations significantly reduces the distance between the forecasts and the mean of the distribution at the cost of more effort to produce the forecast, whereas asking for the median is cognitively less demanding but suffers from a large effect of individual ability, i.e. only the subjects with a high numeracy score report a point forecast close to the median.

The article is organised as follows. Section 2 explains the experimental design and procedure. In section 3 we present the descriptive statistics of the experiment. Section 4 presents our empirical model and section 5 interprets its results. We close with a discussion in section 6.

2 Experiment

2.1 Experimental design

During the experiment the participants were asked to report point predictions regarding the future realisations of a random iid variable. They received the following explanation about the experimental task: *"You work as a professional forecaster. You are asked to predict the future profits of 10 different firms. For each firm, you have access to its "profit distribution" and must report predictions regarding the future profits of the firm."* A profit distribution was displayed in the form of a histogram with eight possible realisations for the company's profit: $\{-35, -25, -15, -5, 5, 15, 25, 35\}$. Forecasters had to make 10 consecutive predictions for each of the 10 companies.

The 10 distributions used in the experiment to represent the 10 companies are shown in Figure 1.[1] In total, participants were asked to make 100 predictions in

[1] Appendix B explains in detail how they were constructed

each treatment.

A point prediction could be any real number between -35 and 35. After each prediction, a forecaster saw the profit realisations drawn from the displayed distribution. These realisations were provided alongside all the past predictions and realisations in the form of a history table. Figure 2 shows a screenshot of the prediction task presenting the profit distribution on the top left side. The companies were presented to the participants sequentially in random order. Once a forecaster had provided the 10 predictions for a company, a new company (i.e. another profit distribution, and an empty history table) was displayed on the screen and the forecasters started all over again to provide the point predictions for this company.

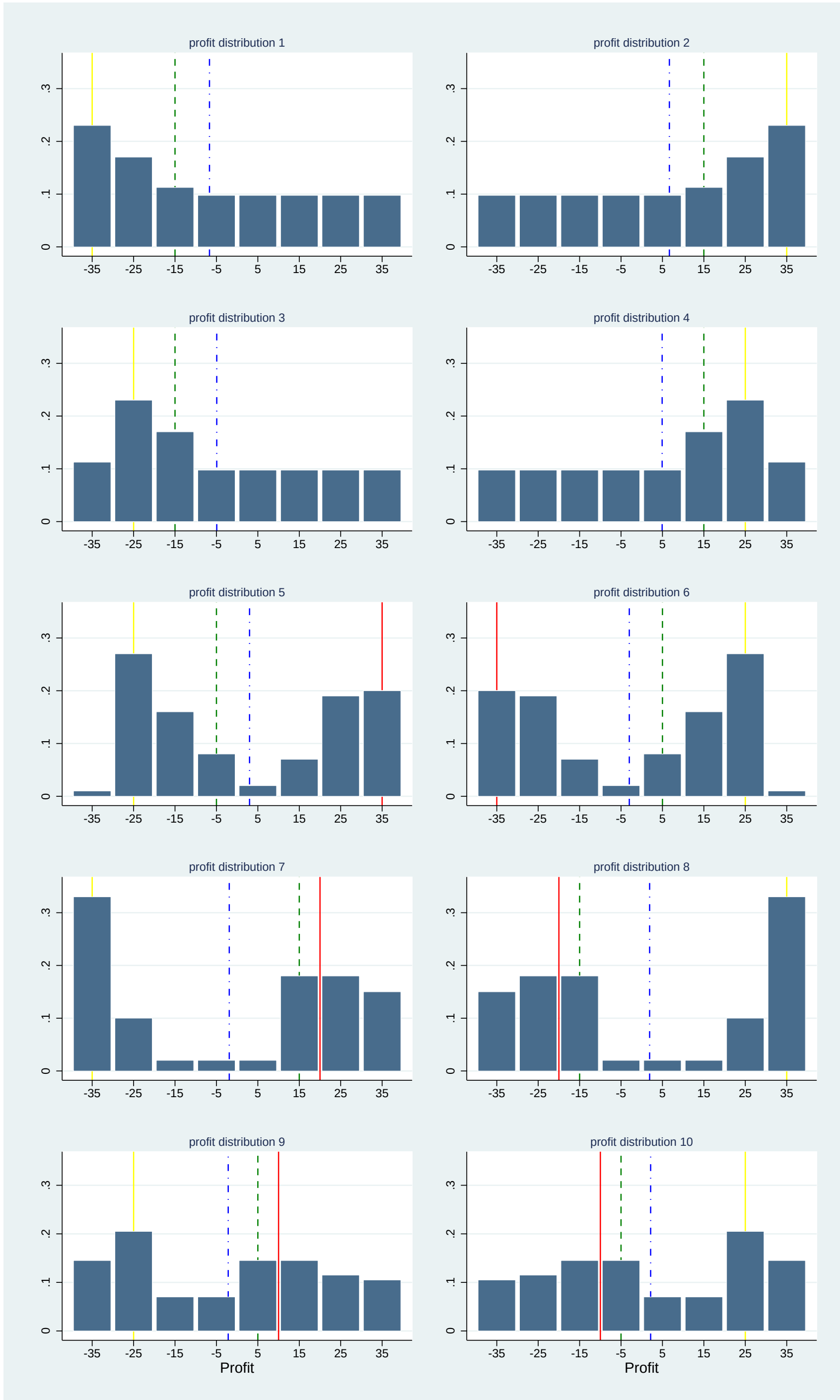

Figure 1: Ten profit distributions that participants saw in random order and for which they predicted the realisation of the next or the next six draws.
Even distributions are the "mirror" counterpart of odd distributions, i.e. all profits are multiplied by -1, e.g. distribution 1 has the mode at -35, distribution 2 has the mode at 35.
—— —— Full lines indicate the mode (yellow) or the second highest mode (red).
- - - - Dashed lines (green) indicate the location of the median.
-.-.-.- Dashed-dotted lines (blue) indicate the location of the mean.

We implemented a total of four treatments. The treatments varied in whether the participants had to predict the realisation of the next period's profit or to make one prediction that was close to the profit realisations of the following six periods. The four treatments were as follows:

(1) In the first treatment, "***Guess-1***", forecasters were asked to predict the next period's profit realisation without further instructions. We only told them to report "their best guess" and to "be as close as possible to the next period's realisation of profit". This treatment was a replication of Kröger and Pierrot (2019).

(2) In the second treatment, "***Guess-6***", we increased the length of the forecast interval from one draw to six consecutive draws. The forecasters had to provide a **unique** forecast that was compared to the next six consecutive draws. Again, the participants were told to be as "close as possible to each of the six profit realisations", but with one single forecast.

(3) In the third treatment, "***Mean-6***," the participants were asked to report a forecast as close as possible to the mean of the next six profit realisations. The following explanation of the concept of "Mean" was displayed on the screen during the forecasting task (translated from French): "*The average profit is a number which, multiplied by the number of periods, gives the total of the profits realised for these periods. For example, if the profits of six periods are 15, -5, -15, 5, 25 and 15, the total profit is 40 and the average profit is 6.67, because 6.67x6 = 40.*"

(4) In the fourth treatment, "***Median-6***", the participants were asked to report a forecast as close as possible to the median of the next six profit realisations. The following explanation of the concept of "Median" was displayed on the screen during the forecasting task (translated from French): "*The median is the number for which half, i.e. three of the realised profits from the next six periods are lower and the other half, thus the other three profits are higher. For example, if profits were 15, -5, -15, 5, 25 and 15, the median would be 10.*"[2]

[2]Dominitz and Manski (1996) are to the best of our knowledge the first to explicitly elicit the median of a distribution in the context of students' subjective expectations about returns to schooling. We formulated our explanation for the median very closely to theirs.

Figure 2 presents a screenshot from the baseline treatment "***Guess-6***". A full set of instructions can be found in appendix A.

"***Guess-6***" is our baseline treatment in this study. We can investigate changes in the forecasts when increasing the length of the interval of realisations that the point forecast covers, comparing ***Guess-6*** to ***Guess-1***. And we can observe the effect of asking for a particular point, either mean or median, comparing ***Guess-6*** to ***Mean-6*** and ***Median-6***.

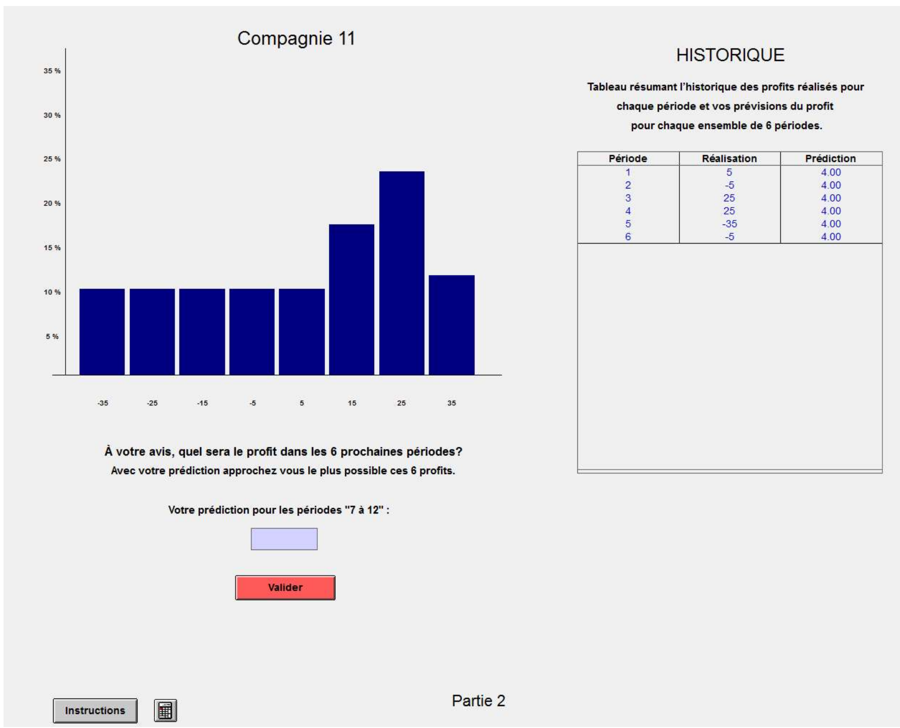

Figure 2: Screenshot of baseline treatment ***Guess-6***.

The experiment was conducted in September 2016 at the LEEL (Laboratory of Economic Experiments at Laval University, Quebec, Canada). A total of 41 forecasters participated in two different treatments. Forecasters faced one of three different treatment orders. Forecasters in Group 1 (N=12) first took part in ***Guess-1*** before continuing with ***Guess-6***. Forecasters of the second group, Group 2 (N=13), and the third, Group 3 (N=16) participated first in the baseline treatment, ***Guess-6***, before continuing respectively in the ***Mean-6*** or ***Median-6*** treatment.

After the prediction task, forecasters filled in a post-experimental questionnaire including the Berlin Numeracy Test (BNT hereafter, Cokely et al. 2012) as a measure of cognitive ability, standard socio-economics questions, and they were asked to explain the strategy employed when producing their point predictions.

We use the BNT to measure the numeracy of our subjects, i.e. the cognitive ability to represent, store and accurately process mathematical operations.[3] The BNT is particularly powerful for measuring the cognitive ability of individuals to understand and manipulate ratio concepts, proportions, probabilities and percentages.

2.2 Experimental procedure

The experiment was conducted in the laboratory of experimental economics of University Laval in Quebec using the software z-Tree (Fischbacher, 2007). We recruited our participants on the campus of Laval University via an online recruitment system. Upon arriving at the laboratory, participants were randomly seated at one of the terminals. After having signed the consent forms, the experiment started.

A video presented the instructions at the beginning of the experiment and demonstrated the forecasting task. The instructions were also accessible in a written form during the entire experiment. Participants could click on the "Instructions" button on the bottom left of the screen and read the instructions again in a pop-up window on the screen.

A session lasted an average of two hours including watching the video instructions and responding to the post-experimental questionnaire. We conducted a total of five sessions and collected 8,200 point predictions.

[3] We employed the computerised version of the BNT, that is adaptive such that follow-up questions depend on previous answers. The test contains 4 questions.

Participants received a 5 CAD show-up fee, a fixed payment of 30 CAD for the completion of the forecasting task and an additional 5 CAD for the completion of the post-experimental questionnaire.

We did not incentivise the point predictions for various reasons. First, survey data collectors often do not and cannot incentivise survey questions, such as income , inflation, GDP, life expectancy or others. Second, using a particular scoring rule to incentivise beliefs would force a profit-maximising respondent to report a particular point of the distribution (e.g. the mean for the quadratic scoring rule, the median for the absolute scoring rule) and would not allow us to study the specific point at which their subjective distribution respondents report when asked to guess.

3 Descriptive Statistics

In this section, we present some descriptive statistics on forecasters background characteristics, on their point predictions and on the effort they put in their predictions measured by the amount of time they take to produce a prediction.

3.1 Individual Characteristics

A total of 41 subjects participated in the experiment. They were an average of 32 years old and were mostly students and personnel from Laval University. Students came mainly from economics, environmental science and engineering. Almost half of all participants were women (40%). Table 1 reports the descriptive statistics of their background characteristics (age, gender and *numeracy*) by treatment.

The variable "numeracy" is a measure for forecasters individual cognitive ability. It is the score of the Berlin Numeracy Test (BNT), a psychometric instrument that assesses and represents the statistical and risk literacy on a scale from 1 to 4. A value of 1 in the BNT represents the lowest level of numeracy and 4 the highest. Participants in our experiment have an average numeracy of 2.4. It varies across participants with a standard deviation of 1.2. These figures are similar to the findings reported by Cokely et al.(2012). They obtained a mean numeracy of 2.6 with a variance of 1.13 from a subject pool of students and former students.

As everyone participated in the baseline treatment ***Guess-6***, Column (2) of Table 1 contains the average statistics of all variables for the entire sample. These statistics are comparable across treatments even though the ratio of women is slightly higher in the ***Median-6*** treatment (0.5) and the participants are a little older in the ***Mean-6*** treatment (34.4).

For the analysis in this paper we will focus only on the first 10 predictions per company. In future research we plan to study how feedbacks affect forecasts in cases where the same event is repeatedly predicted. Thus the analysis that follows is based on the point predictions reported as a first forecast for each of the 10 distributions that the respondents summarised in two treatments.

	(1) ***Guess-1***	(2) ***Guess-6***	(3) ***Mean-6***	(4) ***Median-6***
Group	1	1, 2 & 3	2	3
# of subjects	12	41	13	16
age -in years-	32.6	32.5	34.4	31.7
	(10.3)	(9.8)	(10.1)	(9.0)
gender (0=male)	0.33	0.41	0.38	0.5
numeracy	2.4	2.4	2.3	2.4
	(1.1)	(1.1)	(1.1)	(1.0)

Table 1: Descriptive statistics by treatment. Standard deviations are in parentheses.

3.2 Point predictions in relation to central tendencies

Forecasters could report any real number between -35 and 35 as a prediction. For each individual $i = (1, ..., N)$ we observe a point prediction $pp_{i,d}$ for the displayed distribution $d = (1, ..., D)$. In this section, we report the number of predictions that are close to the mean, median and mode of each distribution in each treatment. We also report the absolute distances between each point prediction and the three central tendencies of its underlying distribution, i.e. the mean, the median and the mode.

3.2.1 Correspondence between forecasts and tendencies

We define that a point prediction “is close” to a central tendency when it lays within the interval, or “bin”, containing this central tendency. Each interval has a length of 10 units and is centred around a possible profit. For example, a prediction of -2 for a distribution with mean -1.9, median 15 and mode -35 would be coded as corresponding to the mean because it is contained within the same bin ($-2, -1.9 \in [-10; 0]$). For the same distribution, a prediction of -31 would be coded as corresponding to the mode ($-35, -31 \in [-40; -30]$) and a prediction of 25 would be coded as not corresponding to any of the three central tendencies. We have constructed the distributions such that the central tendencies are apart from one another and may never lie in the same bin. Therefore, according to our definition, a prediction cannot correspond to more than one central tendency at a time.[4]

Group	Central Tendency	***Guess-1***	***Guess-6***	***Mean-6***	***Median-6***	p-value
Group 1	at least one	0.73	0.76			0.64
	Mean	**0.11**	**0.31**			**0.00**
	Median	0.18	0.21			0.65
	Mode	**0.71**	**0.48**			**0.00**
Group 2	**at least one**		**0.72**	**0.82**		**0.03**
	Mean		**0.27**	**0.54**		**0.00**
	Median		0.29	0.37		0.15
	Mode		**0.44**	**0.09**		**0.00**
Group 3	at least one		0.77		0.75	0.67
	Mean		0.33		0.43	0.12
	Median		**0.20**		**0.38**	**0.00**
	Mode		**0.47**		**0.19**	**0.00**

Table 2: Shares of point predictions falling into the same bin as one of the three central tendencies, and also for those point predictions the proportion by central tendency.
Information in bold letters indicates significant results of a pairwise sign test at the 5% level.

[4] See appendix B for details on the construction of the distributions.

In Table 2, we report the share of point predictions that are close to the central tendencies for each experimental group and each treatment. For example, in the ***Guess-1*** treatment, 73% of the predictions of Group 1 are close to one of the central tendencies. Among these predictions, 11% correspond to the mean, 18% to the median and 71% to the mode. The p-values displayed in the last column are obtained by testing the null hypothesis that the treatment had no effect on the shares in a panel model with fixed effects for the subjects.

The ***Guess-1*** treatment can be seen as a laboratory replication of the empirical analysis of Engleberg, Manski and Williams (2009) and our findings are in line with theirs. In their investigation of the survey of professional forecasters, they found that 80% of the point predictions recorded correspond to one or more of the three central tendencies, a share that is very close to and only slightly higher than the 73% that we observe here. This difference might be due to the ability of the professional forecaster to report points that are more consistent with their subjective distributions. Another potential explanation lies in the shape of the underlying distributions chosen for our experiment which are more complicated than the uni-modal and often Gaussian-looking ones that were reported by the forecasters in the SPF.

Looking at the responses of all three groups under the baseline treatment (***Guess-6***), we find that a comparable proportion of point predictions are close to one of the central tendencies, i.e., between 72% and 77%. Also, for each different tendency, the share within those predictions that are close to that tendency is comparable across group 1, 2, and 3 with respectively 31%, 27%, and 33% of the predictions being close to the mean; 21%, 29%, and 20% to the median; and 48%, 44% and 47% to the mode. These similarities between the predictions made by the three groups in the treatment ***Guess-6*** indicate that the predicting rules used by the subjects in this treatment are unaffected by the order in which the treatment was conducted, either as first (groups 2 and 3) or second (group 1) treatment.

Table 3 also reports how the proportions of point predictions that correspond to a central tendency change when the groups transfer from one treatment to another. In Group 1, that starts with ***Guess-1*** before continuing in ***Guess-6***, the share of predictions that are close to any tendency is the same for both treatments.

Moreover, in both treatments most of the forecasters seem to favour a prediction close to the mode of the distribution. Nevertheless, the share of predictions that are close to the mean increases significantly, by 20 percentage points, when it is made for six instead of one realisation. Meanwhile, the share of predictions close to the mode decreases by 23 percentage points.

The within-variations of groups 2 and 3 show that modifying the question formats can successfully change the reporting rules. Forecasters in groups 2 and 3 chose to report the requested point, either the mean or the median, twice as often when they were asked to make a prediction for the next six realisations compared to a simple guess without further information. Notably, in group 2, the share of predictions close to the mean increase from 27% to 54% when asked to report the mean of the following 6 draws instead of a guess that is close to the six draws. Similarly, predictions close to the median increased in Group 3 from 20% to 38%. At the same time, the share of predictions close to the mode of the distribution decreases substantially from 44% to 9% and from 47% to 19% in Group 2 and Group 3 respectively.

To sum up, the analysis on the aggregate level suggests that (i) people report the mode most prominently as a point forecast when they are asked to predict the outcome of one random draw; (ii) expanding the prediction's horizon by asking for a forecast close to multiple draws drives some of the participants to choose a point prediction close to the mean even though the modal report remains the most popular; and (iii) asking for a forecast close to a particular central tendency of the next six realizations, i.e. the mean or median, increases the share of point predictions that are close to the corresponding central tendency of the distribution and decreases dramatically the share of those that are close to the mode.

3.2.2 Distance between individual forecasts and central tendencies

In this section, we report the absolute distances $y_{i,d}^{ct} = |pp_{i,d} - ct_{i,d}|$ between each point prediction and the three central tendencies of its underlying distribution, i.e. the mean $y_{i,d}^{mean} = |pp_{i,d} - mean_d|$, the median $y_{i,d}^{median} = |pp_{i,d} - median_d|$ and the mode $y_{i,d}^{mode} = |pp_{i,d} - mode_d|$ of forecaster i for distribution d.

Table 3 contains the average distances between the predictions and the central tendencies of the distributions.[5] In the ***Guess-1*** treatment, the shortest distance is between the forecasts and the mode, $\overline{y}_{i,d}^{mode} \leq \overline{y}_{i,d}^{ct}, ct = \{mean, median\}$. Nevertheless, it remains relatively large compared to $\overline{y}_{i,d}^{mean}$ in the ***Mean-6*** and $\overline{y}_{i,d}^{median}$ in the ***Median-6*** treatment. This average distance, $\overline{y}_{i,d}^{mode} = 14.5$, and the large standard deviation associated with it ($\sigma_{i,d}^{mode} = 18.3$) indicate the presence of heterogeneity amongst the reporting rules of forecasters when they are asked for a prediction of one draw without further instructions.

	Guess-1	***Guess-6***	***Mean-6***	***Median-6***
$\overline{y}_{i,d}^{mean}$	20.4 (10.3)	15.6 (10.9)	8.1 (8.2)	11.3 (9.6)
$\overline{y}_{i,d}^{median}$	21.2 (15.5)	17.3 (13.9)	10.6 (10.1)	12.8 (12.3)
$\overline{y}_{i,d}^{mode}$	14.5 (18.3)	17.7 (16.6)	24.3 (11.9)	22.4 (14.4)
# obs	120	410	130	160

Table 3: Distances between forecasts and central tendencies (mean, median and mode).

In the ***Guess-6*** treatment we observe a shift in the reported predictions. The shortest distance is with the mean, $\overline{y}_{i,d}^{mean} = 15.6$, and the standard deviation associated with that measure is smaller than in the ***Guess-1*** treatment, with $\sigma_{i,d}^{mean} = 10.9$. When they are asked to make one prediction for the next six draws some of the participants change their reporting rule. This leads to an increase of the distance between the point forecasts and the mode form $\overline{y}_{i,d}^{mode} = 14.5$ to $\overline{y}_{i,d}^{mode} = 17.7$ and a decrease of the distance with the mean. Nevertheless a high heterogeneity in the forecasting choices remains.

In the ***Mean-6*** and ***Median-6*** treatments, the distance between the forecasts and the mode $\overline{y}_{i,d}^{mode}$ is substantially larger than in the two ***Guess*** treatments, with $\overline{y}_{i,d}^{mode} = 24.3$ and $\overline{y}_{i,d}^{mode} = 22.4$ respectively. The points reported are much closer to the mean, with $\overline{y}_{i,d}^{mean} = 8.1$ and $\overline{y}_{i,d}^{mean} = 11.3$ respectively, and to the median, with $\overline{y}_{i,d}^{median} = 10.6$ and $\overline{y}_{i,d}^{median} = 12.8$ respectively. Notably, the associated standard deviations are the lowest in these two treatments. The participants respond to the treatment by changing their forecasting rule towards the central

[5] With bimodal distributions, we only consider the mode with the highest probability mass.

tendency corresponding to the question format. They report predictions closer to the mean in the ***Mean-6*** treatment and to the median in the ***Median-6*** treatment.[6] Moreover, asking for a specific tendency such as the mean or the median of the next six distributions reduces the distance between the forecasts and the corresponding central tendency of the underlying distribution in a way that suggests a lower heterogeneity among predictions.

In short, we observe that not indicating a precise point of the distribution as in our ***Guess-1*** and ***Guess-6*** treatments, leads to a large distance between the point predictions reported by the forecasters and all the central tendencies of the displayed distributions. Contrarily, the distances measured when the mean or the median of the next realisations were indicated as the desired point predictions are much shorter, especially with the corresponding tendency (mean or median). So far, our descriptive analysis suggests that the treatments ***Mean-6*** and ***Median-6*** were successful guiding forecasts towards a desired point prediction. Moreover, they reduced the heterogeneity of the recorded forecasts in terms of distance with the target central tendency.

3.2.3 Effort measured by the time taken to produce a forecast

The time taken by the subjects to report a prediction varies quite intensively between the different treatments - it ranges from 20" to over a minute[7]. This may be an indication that producing a point forecast can be more or less cognitively demanding depending on the question to answer. Interestingly, producing a prediction for one random draw without further indication and predicting the median of six random draws were the two fastest tasks for the participants. It took them 24.7" (***Guess-1***) and 21" (***Median-6***) whereas making one prediction for six draws without any indication of the point to report or asking for the mean took an average of 43.6" (***Guess-6***) and 63" (***Mean-6***).

[6] By construction, if the predictions are closer to the mean, they are also closer to the median and vice versa as the mode is either on the left or right of the other two central tendencies (see Table 8). The median in the ***Mean-6*** treatment, however, has a 2.5 units larger distance compared to the mean, but the mean in the ***Median-6*** treatment has only a very small 1.5 units difference compared to the median.

[7] We measure the time participants took from first seeing the distribution until they submit their prediction. The visualisation of the draws is not included in the time measure to keep it comparable across treatments.

	Guess-1	***Guess-6***	***Mean-6***	***Median-6***
$\overline{time}$	24.7 (46.8)	43.6 (75.9)	63.0 (108.7)	21.0 (27.0)
$time_2$	46.7 (39.6)	57.2 (88.6)	66.4 (77.8)	22.7 (28.7)
$time_6$	12.4 (9.5)	38.0 (67.4)	67.0 (109.0)	15.9 (15.3)
$time_{10}$	12.3 (10.4)	25.1 (36.3)	37.8 (44.9)	15.8 (23.3)
# obs	120	410	130	160

Table 4: Time taken to produce a forecast in seconds

Of course, one might expect the prediction time to shorten over the experiment because of learning effects. To investigate this effect, in Table 4 we provide multiple measures of the prediction time at different moments in the experiment, i.e. for the distributions seen in order 2, 6 and 10.[8] These measures give a more precise idea of how the prediction time evolves for each treatment. As expected, it generally decreases as the subjects become more familiar with the task. Nevertheless, the cross-treatment comparison holds for every point in time. The prediction time is, on average, shorter in the treatments ***Guess-1*** and ***Median-6*** compared to the treatments ***Guess-6*** and ***Mean-6***.

4 Empirical model

In this section we present our empirical model with the aim of capturing the prediction task. We want to account for the benefits and costs associated with each question format. We model the distance between a point prediction and a central tendency, $y^{ct}_{i,d} = |pp_{i,d} - ct_{i,d}|$ as a function of the forecaster *numeracy*, the *effort* that he exerts when producing his prediction, captured by the time taken to provide the prediction, and the forecaster's *experience*, captured by the order in which the forecaster saw the distribution. The order of distributions was randomised for each forecaster separately and takes values from 1 to 10 with higher values corresponding to distributions seen later in the experiment. We

[8]We report this time from the distribution presented in order 2 onward with equidistant intervals of 4, thus at 6 and at 10, to give an idea of the evolution of reporting time over the course of the experiment. We excluded the first distribution as it might just reflect the time it took for the subjects to understand the task.

interpret this variable as the *experience* of the forecaster.

4.1 Panel data model

To ensure that our variable of interest, the absolute distance between a prediction and the central tendency, is always larger than zero, we model this distance with an exponential function.

$$y^{ct}_{i,d} = exp[\alpha + x'_{i,d}\beta + \epsilon_{i,d}]$$

Where $x_{i,d}$ contains a binary variable for each treatment as well as the individual characteristics of the forecasters and their interaction.[9] Individual characteristics include variables such as the *effort* measured in seconds as time to produce a prediction ($effort_{i,d}$), the *experience* measured by the order in which the distribution appeared from 1 to 10 ($experience_{i,d}$), and the forecaster's level of numeracy measured by an integer index between 1 (low) and 4 (high) ($numeracy_{i,d}$).

We also allow for an interaction between *effort* and *experience* ($effort{*}experience_{i,d}$) as forecasters might produce predictions faster at the end of the experiment because they are more familiar with the prediction task. The error term $\epsilon_{i,d} = \mu_i + \lambda_d + e_{i,d}$ consists of an individual (random) component $\mu_i \sim N(0, \sigma_\mu)$, a fixed effect for the distribution λ_d and a random component $e_{i,d} \sim N(0, \sigma_e)$.

We can write the model in matrix notation as follows:

$$\mathbf{Y^{ct}} = \mathbf{exp}[\alpha \quad \mathbf{1_{ND}} + \mathbf{X}'\beta + \mathbf{G_1} \quad \mu + \mathbf{G_2} \quad \lambda + \mathbf{e}]$$

with: $\quad \mathbf{G_1} = \mathbf{I_N} \otimes \mathbf{1_D}$ and $\quad \mathbf{G_1} = \mathbf{1_D} \otimes \mathbf{I_N}$

We use a two-way error component panel data model. The error component

[9] $x_{i,d} = \{FirstTreatment_{i,d},\ effort_{i,d},\ experience_{i,d},\ numeracy_{i,d},\ D_{\textbf{Guess-1}},\ D_{\textbf{Mean-6}},\ D_{\textbf{Median-6}},\ D_{\textbf{Guess-1}} * effort_{i,d},\ D_{\textbf{Guess-1}} * experience_{i,d},\ D_{\textbf{Guess-1}} * numeracy_{i,d},\ D_{\textbf{Mean-6}} * effort_{i,d},\ D_{\textbf{Mean-6}} * experience_{i,d},\ D_{\textbf{Mean-6}} * numeracy_{i,d},\ D_{\textbf{Median-6}} * effort_{i,d},\ D_{\textbf{Median-6}} * experience_{i,d},\ D_{\textbf{Median-6}} * numeracy_{i,d}\}$

attached to the distribution is captured by fixed effects λ_d and the error component attached to the subjects is captured by random effects μ_i in order to avoid multicollinearity with the numeracy scores.

$y_{i,d}^{ct}$ follows a Gamma distribution ($y_{i,d}^{ct} \sim \Gamma(a,b)$). We estimate the model using the GEE (Generalised Estimating Equation) procedure. This method allows us to compute the parameters of the following relationship:

$$log(E[y_{i,d}^{ct}]) = \alpha + x_{i,d}'\beta$$

It is important to note that the GEE procedure directly estimates the relationship between the expectation of the variable of interest and the explanatory variables. This feature allows us to keep the observations where $y_{i,d}^{ct} = 0$ (where the logarithm is undefined). Nevertheless, it implies that the model that we specified at the beginning of this section ($y_{i,d}^{ct} = exp[\alpha + x_{i,d}'\beta + \epsilon_{i,d}]$) is an approximation - it is not true when $y_{i,d}^{ct} = 0$. The GEE procedure requires a third assumption on the working correlation matrix of Y^{ct}. We assume that this matrix is exchangeable - i.e. it allows for a unique subject-specific correlation parameter γ. In other words, we assume that all the predictions made by a person are correlated with each other in the same way.

4.1.1 Parameters' estimates

In this section, we present the results obtained by estimating the model, more precisely, the marginal effects of *effort*, *experience* and *numeracy* computed at the sample mean. We examine the effects of these variables both for the entire experiment as well as for each specific treatment.

$y_{i,d}^{ct}$	$D_{\textit{\textbf{Guess-1}}}$	$D_{\textit{\textbf{Mean-6}}}$	$D_{\textit{\textbf{Median-6}}}$	$numeracy_{i,d}$	$effort_{i,d}$	$experience_{i,d}$
$y_{i,d}^{mean}$	3.91***	-7.00***	-3.40***	-2.34***	-2.34***	-0.05
$y_{i,d}^{median}$	2.82***	-5.13***	-5.24***	-2.00***	-1.75***	-0.37***
$y_{i,d}^{mode}$	-2.64***	8.10***	2.63***	0.32	4.16***	-0.28

Table 5: Marginal Effects in GEE Panel Data Model with ***Guess-6*** as baseline.

Table 6 presents the marginal effects for three models that have as dependent variables the distance between the point prediction and either the mean ($y_{i,d}^{mean}$), the median ($y_{i,d}^{median}$) or the mode ($y_{i,d}^{mode}$) of the distributions. The estimates of the treatment effects reported in Table 5 are perfectly aligned with our descriptive findings in section 3. Taking the ***Guess-6*** treatment as the baseline, we find that all the dummy variables attached to the other treatments in our model affect the distances just as expected.

Questions asking to guess the next realisation of a random variable decrease the distance with the mode of the distribution by 2.64 units and increase the distance with the mean and median by 3.91 and 2.82 units, respectively. Those effects are captured by the dummy variable $D_{\textbf{\textit{Guess-1}}}$.

Meanwhile, asking for the mean of the next six realisations decreases the distance between the point prediction and the mean by 7 and the median by 5.13. However, forecasters report 8.10 units further away from the mode. Asking directly for the median decreases the distance between point predictions and median by 5.24 and the mean by 3.4, whereas the distance between point predictions and mode increases by 2.63. In short, $D_{\textbf{\textit{Guess-1}}}$ increases the distances with the mode and median and decreases the distance with the mode while $D_{\textbf{\textit{Mean-6}}}$ and $D_{\textbf{\textit{Median-6}}}$ have the exact opposite effects.

Regarding the other explanatory variables, Table 5 reports that the points chosen by both forecasters who put more *effort*, i.e. who took more time, and those who obtained a higher *numeracy* score are, on average, closer to the mean and the median of the distributions and further away from the mode. Meanwhile, the *experience* a forecaster has with the prediction task does not substantially affect the distance between predictions and the central tendencies.[10]

In Table 6 we present the treatment-specific marginal effects computed by using the interaction effects of the treatment dummies and the three main explanatory variables. They allow us to explore how *effort*, *experience* and *numeracy* relate to the point predictions reported in each question format. *Experience* significantly reduces the distance between the predictions and the mode in the ***Guess-1***

[10]With the exception of the distance to the median that decreases significantly, however the decrease is very small.

treatment.

	$y_{i,d}^{ct}$	Marginal effects *numeracy*	*effort*	*experience*
	$y_{i,d}^{mean}$	-2.02**	-3.10***	-0.22
Guess-6	$y_{i,d}^{median}$	-2.14***	-0.76	-0.23
(Baseline)	$y_{i,d}^{mode}$	-0.16	2.76***	0.026
	$y_{i,d}^{mean}$	0.16	-12.2***	-0.66*
Guess-1	$y_{i,d}^{median}$	0.59	-5.95*	-0.83*
	$y_{i,d}^{mode}$	-0.24	10.40***	-1.19**
	$y_{i,d}^{mean}$	-2.56***	-0.41	0.015
Mean-6	$y_{i,d}^{median}$	-1.73***	-0.13	-0.076
	$y_{i,d}^{mode}$	1.00	1.66	-0.10
	$y_{i,d}^{mean}$	-3.41***	1.5	0.36*
Median-6	$y_{i,d}^{median}$	-3.13***	-3.02*	-0.38*
	$y_{i,d}^{mode}$	1.76	2.80	-0.33

Table 6: Treatment-specific marginal effects in GEE panel data model

First, we observe that forecasters with higher *numeracy* scores report point predictions closer to the mean or the median when asked to summarise the following six predictions with a single point prediction (***Guess-6***,***Mean-6***, and ***Median-6***). The distance between these two central tendencies and the predictions decreases by 2 units for each unit increase in numeracy when asked for a guess in general (***Guess-6***). This distance decreases even further when being asked to predict either the mean or the median. For example, when asked to report the mean (***Mean-6***) one more unit of the numeracy score decreased the distance of predictions to both the median and the mean by 3.13 and 3.41, respectively. Notably, the level of numeracy of a forecaster is not indicative of the point that this forecaster reports when only asked to guess a single realisation of a random variable without further indication in the ***Guess-1*** treatment.

Second, *effort* amplifies this effect in the ***Guess-6*** treatment. There, a forecaster who takes one more minute to report a prediction would choose a point that is, on average, 3.12 units ($p < 0.01$) closer to the mean and 2.76 units further away from the mode. Thus, a person with two points more on the numeracy scale who think one minute longer reports a point 7 units closer to the mean. The magnitude of the effect of *effort* is almost four times larger in the ***Guess-1*** treatment where taking one minute longer to produce the point forecast translates into being 12 units ($p < 0.01$) closer to the mean. This effect is substantial. For example, a person taking 30 seconds longer than the average time of 24.7 seconds would produce a prediction that is, on average, 6 units closer to the mean - equalizing the average distance to the mean ($\overline{y}_{i,d}^{mean}$: $20.4 - 6 = 14.4$) to the distance to the mode ($\overline{y}_{i,d}^{mode} = 14.5$). Providing more *effort* also significantly decreases the distance to the median when asked to report the median while not affecting at all the distances to the other central tendencies. There is no improvement from providing more *effort* on closeness to the central tendencies when asked to report the mean.

Finally, forecasters with more *experience* report closer to any central tendency when asked to guess one future realisation. Thereby, the distance with the mode declines the fastest, followed by the distance with the median and then that with the mean. Just as a comparison, predictions of the 10th company would be 11 units closer to the mode, by 7 units closer to the median and by 6 units closer to the mean compared to those of the first. We observe further learning effects when the instructions indicated to report the median. In this treatment, more experienced forecasters predicted points closer to the median and further away from the mean. These opposites are of almost the same magnitude, i.e. a forecaster will for his 10th forecast predict a point that is 3 units further from the mean and 3 units closer to the median.

5 Comparison of the question formats

Table 7 regroups all the relevant results that we have gathered in the analysis in order to give an overview of how the different questions modify the forecasters' predictions. Taking ***Guess-6*** as the baseline, we now compare the respective benefits and drawbacks of the different question formats.

Asking to guess one future realisation (treatment ***Guess-1***) requires less effort, i.e. the time is reduced by half to produce a prediction. With this question, the forecasters who put more effort, i.e. who take more time, or more experienced forecasters make predictions that are closer to the mean and the median. However, reporting the mode is by far the most preferred strategy.

Asking to report the mean of six future realisations (treatment ***Mean-6***) is, on average, the most successful in eliciting homogeneous forecasts that reflect a central tendency of the objective distribution - its mean. The average distance between the point predictions and the mean of the distribution is the smallest ($\overline{y}_{i,d}^{mean} = 8.1$ units) and the variance around that measure is only $\sigma_{i,d}^{mode} = 8.2$. Nevertheless, with an average time of 63 seconds, this question is the most time-consuming indicating that it is likely to be the most cognitively demanding for the participants. Taking more time to think does not change the subjects' behaviours in this treatment. Everybody already seems to be investing a great many cognitive resources in order to respond to the task to report a point close to the mean.

Asking to predict the median of six future realisations (treatment ***Median-6***) results in predictions that are, on average, very close to both the mean and the median - with respectively $\overline{y}_{i,d}^{mean} = 11.3$ and $\overline{y}_{i,d}^{median} = 12.8$. According to our results, the questions used in the ***Mean-6*** and ***Median-6*** treatments seem to have very similar properties. However, they differ in two important dimensions. Firstly, even though the cognitive ability plays an important role in both, it is even more crucial in the ***Median-6*** treatment where each point in the numeracy score reduces the distance $\overline{y}_{i,d}^{median}$ by 3.13 unit. This measure indicates that forecasters with high numeracy skills seem to better understand the summary statistics that we ask them to use. Secondly, the effort, measured by the average prediction time is much lower in the ***Median-6*** than in the ***Mean-6*** treatment. It took forecasters only 21 seconds on average to produce a point prediction in the former compared to 63 seconds in the latter. In addition, the gains from providing more effort in the ***Median-6*** treatment substantially reduce the distance between the point prediction and the median. This indicates that encouraging respondents to take more time and to think carefully might actually be very beneficial for the precision of the prediction. Furthermore, experience has a positive effect on the precision in ***Median-6*** but none in ***Mean-6***. This result suggests that respondents in survey panels might actually be able to learn

		(1)	(2)	(3)	(4)			(5)
		Descriptive Statistics		Treatment	Marginal effects			
	$y_{i,d}^{ct}$	$\overline{y}_d^{ct}$	$\overline{time}$	effects	*numeracy*	*effort*	*experience*	# obs
	$y_{i,d}^{mean}$	15.6 (10.9)	43.6		-2.02**	-3.10***	-0.22	
Guess-6	$y_{i,d}^{median}$	17.3 (13.9)	(75.9)		-2.14***	-0.76	-0.23	410
(Baseline)	$y_{i,d}^{mode}$	17.7 (16.6)			-0.16	2.76***	0.026	
	$y_{i,d}^{mean}$	20.4 (10.3)	24.7	3.91***	0.16	-12.2***	-0.66*	
Guess-1	$y_{i,d}^{median}$	21.2 (15.5)	(46.8)	2.82***	0.59	-5.95*	-0.83*	120
	$y_{i,d}^{mode}$	14.5 (18.3)		-2.64***	-0.24	10.40***	-1.19**	
	$y_{i,d}^{mean}$	8.1 (8.2)	63.0	-7.00***	-2.56***	-0.41	0.015	
Mean-6	$y_{i,d}^{median}$	10.6 (10.1)	(108.7)	-5.13***	-1.73***	-0.13	-0.076	130
	$y_{i,d}^{mode}$	24.3 (11.9)		8.10***	1.00	1.66	-0.10	
	$y_{i,d}^{mean}$	11.3 (9.6)	21.0	-3.40***	-3.41***	1.5	0.36*	
Median-6	$y_{i,d}^{median}$	12.8 (12.3)	(27.0)	-5.24***	-3.13***	-3.02*	-0.38*	160
	$y_{i,d}^{mode}$	22.4 (14.4)		2.63***	1.76	2.80	-0.33	
	$y_{i,d}^{mean}$	14.3 (10.8)	39.5		-2.34***	-2.34***	0.05	
All	$y_{i,d}^{median}$	15.9 (13.8)	(73.4)		-2.00***	-1.75*	-0.37***	820
Treatments	$y_{i,d}^{mode}$	19.2 (16.1)			0.32	4.16	-0.28	

Table 7: Descriptive statistics of experimental measures: average distance of prediction and all three central tendencies (mean "$\overline{y}_{i,d}^{mean} = \sum_{\forall i,d} |pp_{i,d} - mean_d|$", median "$\overline{y}_{i,d}^{median} = \sum_{\forall i,d} |pp_{i,d} - median_d|$" and mode "$\overline{y}_{i,d}^{mode} = \sum_{\forall i,d} |pp_{i,d} - mode_d|$"; average time taken per prediction, treatment effects and marginal effects of the panel data model at the sample mean.

to predict the median, while there are no effects for the mean.

Finally, providing more effort decreases the distance to the mean and median and increases the distance to the mode in all the treatments. Numeracy also has a strong effect on reporting a point closer to the mean and median, but no effect on the distance to the mode.

6 Conclusion

In this work we present an experimental study on the properties of different question formats that aim to elicit point predictions. We study how the questions alter the selection of points when people are asked to predict the realisations of an *iid* random variable with a known distribution. We hope that the results of this study are informative for survey designers and researchers using point predictions.

We find that forecasters with a higher cognitive ability report predictions closer to the mean and median, particularly when they were asked to report those central tendencies. This result has several implications. First, measures of numeracy can indicate the closeness to the mean or median of a prediction. Second, questions asking specifically for the mean can be asked in groups with a high numeracy score but should be avoided in groups where this score is low. Third, there might be benefits from educating respondents by helping them to achieve higher numeracy scores.

Furthermore, we observe that for some simple question formats such as guessing a single future realisation or the median of six future realisations forecasters who provide more effort, i.e. who take more time to produce a prediction, report point predictions that are closer to the mean, when a particular point is not being asked for, or the median when specifically asked to summarise the median. These results suggest that additional hints when answering a question, such as the encouragement for taking time to think or providing financial incentives to compensate people for their effort, might be effective tools for providing responses that improve the outcome when respondents do not put in much effort when producing a forecast. Thus, learning opportunities and examples before eliciting point predictions are very likely to decrease this distance with the central tendencies.

In the different question formats that we study here, we observe a clear trade-off between asking precisely for the mean of a distribution or for a summary of the distribution without further instructions. Asking for the mean of a distribution is cognitively demanding and takes much more time to respond. But as a result, this question format yields predictions that are closer to the desired central tendency with less variation in the responses between forecasts. Asking for a guess of what the next realisation of a random variable might be seems at first glance easier to respond to - as response times are much shorter. However, responses are much more dispersed and strongly affected by whether respondents reflect on the task and take the time to produce a prediction or not.

References

[1] Adeline Delavande. Probabilistic expectations in developing countries. *Annu. Rev. Econ.*, 6(1):1–20, 2014.

[2] Jeff Dominitz and Charles F. Manski. Eliciting student expectations of the returns to schooling. *Journal of Human Resources*, 31:1–26, 1996.

[3] Joseph Engelberg, Charles Manski, and Jared Williams. Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business and Economic Statistics*, 165:146–158, 2009.

[4] Urs Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178, June 2007.

[5] Charles F. Manski. Measuring expectations. *Econometrica*, 72(5):1329–1376, 2004.

[6] M. Hashem Pesaran and Martin Weale. *Handbook of Economic Forecasting*, volume 1, chapter 14: Survey Expectations, pages 715–776. 2006.

[7] Leonard J Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1971.

A Instructions

The instructions are very similar in all treatments. The differences between treatments are indicated in bold.

Page 1

Welcome!
Please listen carefully to the instructions. The experiment lasts approximately 120 minutes. During the experiment, we ask that you do not communicate with your neighbours. If you have any questions, please raise your hand and we will answer your question in private.

Page 2

Before starting the experiment, we will present the instructions. We will explain the progress of the experiment in detail.
The experiment consists of two parts that vary slightly in the tasks you will be asked to fulfil. First, we will present the task of the first part and after completing the first part, we will present the task of the second part.
Once you have started, an on-screen summary of the instructions will be available for the duration of the experiment.
We will also provide you with a printed copy of the instructions.
When you have finished the experiment, please, stay seated! We will come to your cubicle and will also give you a cash compensation for your participation.

Page 3

Imagine that you are working for a consulting firm as a forecaster. Your task is to predict the future profits of different companies.

Page 4

All companies make a profit between -35 and 35. More specifically, the profit of each company is one of the following numbers: -35, -25, -5, -5, 5, 15, or 25, 35. To the left, you can see an example of the chances associated with making these eight profits in the form of a histogram. The vertical axis of this histogram shows the probability that each of the eight potential profits (-35, -25, -15, -5, 5, 15, 25 or 35) realizes.
This histogram is from Company X.

Page 5

You follow the company for 10 periods. In each period, the company makes a new profit. The next period's profit is not known in advance. It will be determined by the computer drawing randomly one from eight profits (-35, -25, -15, -5, 5, 15, 25 or 35) and this is according to the chances that each particular profit is realised. The probabilities that a particular profit is realised are shown in the histogram. The profit of the company in a particular period is independent of other past and future profits.

In the Guess-1 Treatment : Your task is to provide a quantified forecast of the company's profit earned in the next period.
In the Guess-6 Treatment : Your task is to provide a quantified forecast of the company's profits earned in the next six periods.
In the Mean-6 Treatment : Your task is to provide a quantified forecast of the company's average profit earned in the next six periods.
In the Median-6 Treatment : Your task is to provide a quantified forecast of the company's median profit earned in the next six periods.

In the Mean-6 Treatment : The average profit is a number which, multiplied by the number of periods, gives the total of the profits realized for these periods. For example, if the profits of six periods are 15, -5, -15, 5, 25 and 15, the total profit is 40 and the average profit is 6.67, because 6.67x6 = 40.

In the Median-6 Treatment : The median is the number for which half, i.e. three of the realised profits from the next six periods are lower and the other half, thus the other three profits are higher. For example, if profits were 15, -5, -15, 5, 25 and 15, the median would be 10.

As a forecast you can specify any number between -35 and 35.
In the Guess-1 Treatment : You can see below an example, where you can enter your prediction for period 4.
In the Guess-6, Mean-6 and Median-6 Treatments : Below, you can see an example, where you can enter your prediction for periods 1 to 6.
Once you have provided your forecasts, you can validate them by pressing the "Validate" button.

Page 6

In the Guess-1 Treatment : Once your prediction has been validated, you can move on and the profit of the next period will be realised in the following way:
In the Guess-6, Mean-6 and Median-6 Treatments : Once your prediction has been validated, you move on and the profit of the next six periods are realised as follows:

The computer draws a number at random while respecting the probabilities visible in the histogram. For example, the histogram on this screen indicates that the number "-5" has a chance of ...% being drawn. That is, in a large sample of random draws (e.g. 1,000,000 draws) about ...profits would be "-5".
The computer draws a new number at each period. The chance that a number will be drawn again corresponds to the percentages indicated in the histogram.

Page 7

The "History" table at the right of the screen summarises for each period the realisations of the profits and your predictions. The "Period" column indicates the period in which the profit was realised. Your prediction for this period is in the ”Predictions” column.

In the Guess-6, Mean-6 and Median-6 Treatments : Note that you will be asked to give one prediction for six consecutive periods, which the table saves as six separate predictions.

The profit made in that period appears in the “Realisation” column after each period.

Page 8

In the Guess-1 Treatment : Test the program and make a prediction for period 4!

In the Guess-6, Mean-6 and Median-6 Treatments : Test the program and make a prediction for periods 1 to 6!

Page 9

In the Guess-1 Treatment : Note that your prediction appears in the row “Predictions” where period is equal to 4.

In the Guess-6, Mean-6 and Median-6 Treatments : Note that your prediction appears in the row ”Predictions” for the periods of your prediction, periods 1 to 6.

In total, you will be asked to make predictions for 10 companies. This completes the first part. In the second part, you will also make forecasts, under other conditions that will be explained to you after completing the first part.

Page 10

A summary of the instructions will be available throughout the experiment. You can consult it by pressing the “Instructions” button at the bottom left. In addition, if necessary, you can access a calculator throughout the experiment by pressing the button with the symbol of a calculator, next to the instructions button.

Page 11

At the end of the experiment, you will receive a summary for all companies that you have evaluated. This information summarises your predictions and profit realisations separately for both parts. For making these predictions, you will receive 35 CAD and 5 CAD for your participation in the experiment - a total of 40 CAD.

Page 12

This is the end of the instructions. If you have any questions, please raise your hand. We will answer your question in private at your station. If you do not have any questions, you can start the experiment.

B Probability Distributions

The distributions used for the experiments $\{p_n\}$ have a discrete support with probability mass on eight equidistant points $n \in \{1, \dots, 8\}$ with distance d. We have chosen a symmetric support around zero with $d = 10$, resulting in the set of points $n \in \{-35, -25, -15, -5, 5, 15, 25, 35\}$ as support for the distributions. The distributions vary in the location of their central tendencies, i.e. the mean $\mu = \sum_{n=\underline{n}}^{\overline{n}} np_n$, median $m = \{\min n : \sum_{i=\underline{n}}^{n} p_i \geq 0.50\}$ and mode $M = \{n : \max p_n\}$. The distributions are chosen in such a way that their central tendencies each lie in a different interval around a point $n \pm d/2$.

The distributions used in the experiment are presented in Table 8 and Figure 1. In its first three column, the table summarises the three central tendencies and presents the probability distribution and cumulative distribution. We use each distribution twice: once in its original form (odd lines in Table 8 and left column in Figure 1) and once its mirror image (the following even line and the right column, resp.). For example, distribution 2 is the mirror image of 1. We use two unimodal distributions (no 1 and 3 and their corresponding mirror images 2 and 4).[11] For both distributions $M < m < \mu$, and $\mu < m < M$ for the corresponding mirror image. To allow for the order $m < \mu < M$ (and its mirror image $M < \mu < m$) and to increase the distance between the central tendencies, we additionally use distributions that are "slightly" bimodal. For those distributions, we define as the mode the point n with the most probability mass which for distributions 11 and 12 (5 and 6) [7,8 and 9,10] has at least 6 (8) [10] percentage points more than the point with the second highest probability mass. We use distribution 13 as an example in the instructions, but not in the experiment.

In total, counting each distribution and its mirror image separately, we use 12 distributions for the experiment. For half of the distributions, the order of the central tendencies is $M < m < \mu$ (distributions 1, 3 and 5) and $\mu < m < M$ for their mirror distribution. For the other half and their mirror distribution, the order is $M < \mu < m$ (distributions 7, 9 and 11) and $m < \mu < M$, respectively.[12]

[11] A discrete distribution is unimodal with the integer M as mode if $p_n \geq p_{n-1}, n \leq M$ and $p_n \leq p_{n-1}, n > M$. Distributions 1 and 3 (2 and 4, resp.) are not strong unimodal distributions, in the sense of case $p_n^2 \geq p_{n-1}p_{n+1}$. (Encyclopedia of Statistical Sciences, 2006)

[12] Nevertheless, the order $\mu < M < m$ and $m < M < \mu$ is theoretically possible for discrete distributions, we would have needed a support with more equidistant points to implement such order. We did not extend the support in order to have meaningful probability mass at each point of the support.

No interval/bin realization	Central tendencies			Probability and cumulative distribution							
	μ mean	m median	M mode	$-35 \leq x < -30$ -35	$-30 \leq x < -20$ -25	$-20 \leq x < -10$ -15	$-10 \leq x < 0$ -5	$0 \leq x < 10$ 5	$10 \leq x < 20$ 15	$20 \leq x < 30$ 25	$30 \leq x \leq 35$ 35
1	-6.675	-15	-35	0.23 0.23	0.17 0.40	0.1125 0.5125	0.0975 0.61	0.0975 0.7075	0.0975 0.805	0.0975 0.9025	0.0975 1
2	6.675	15	35	0.0975 0.0975	0.0975 0.195	0.0975 0.2925	0.0975 0.39	0.0975 0.4875	0.1125 0.6	0.17 0.77	0.23 1
3	-4.925	-15	-25	0.1125 0.1125	0.23 0.3425	0.17 0.5125	0.0975 0.61	0.0975 0.7075	0.0975 0.805	0.0975 0.9025	0.0975 1
4	4.925	15	25	0.0975 0.0975	0.0975 0.195	0.0975 0.2925	0.0975 0.39	0.0975 0.4875	0.17 0.6575	0.23 0.8875	0.1125 1
5	3	-5	-25	0.01 0.01	0.27 0.28	0.16 0.44	0.08 0.52	0.02 0.54	0.07 0.61	0.19 0.8	0.2 1
6	-3	5	25	0.2 0.2	0.19 0.39	0.07 0.46	0.02 0.48	0.08 0.56	0.16 0.72	0.27 0.99	0.01 1
7	-1.9	15	-35	0.33 0.33	0.1 0.43	0.02 0.45	0.02 0.47	0.02 0.49	0.18 0.67	0.18 0.85	0.15 1
8	1.9	-15	35	0.15 0.15	0.18 0.33	0.18 0.51	0.02 0.53	0.02 0.55	0.02 0.57	0.1 0.67	0.33 1
9	-2.15	5	-25	0.145 0.145	0.205 0.35	0.07 0.42	0.07 0.49	0.145 0.635	0.145 0.78	0.115 0.895	0.105 1
10	2.15	-5	25	0.105 0.105	0.115 0.22	0.145 0.365	0.145 0.51	0.07 0.58	0.07 0.65	0.205 0.855	0.145 1
11	1.8	-15	-25	0.01 0.01	0.3 0.31	0.2 0.51	0.02 0.53	0.03 0.56	0.04 0.6	0.2 0.8	0.2 1

Table 8: Central tendencies (mean, median and mode), probability and cumulative distribution of distributions used for the experiment (1-10) and the distribution used as illustration in the instructions (11).

Discussion Papers of the Research Area Markets and Choice 2019

Research Unit: **Market Behavior**

Azar Abizada, Inácio Bó Hiring from a pool of workers	SP II 2019-201
Philipp Albert, Dorothea Kübler, Juliana Silva-Goncalves Peer effects of ambition	SP II 2019-202
Yves Breitmoser, Sebastian Schweighofer-Kodritsch Obviousness around the clock	SP II 2019-203
Tobias König, Sebastian Schweighofer-Kodritsch, Georg Weizsäcker Beliefs as a means of self-control? Evidence from a dynamic student survey	SP II 2019-204
Rustamdjan Hakimov, Dorothea Kübler Experiments on matching markets: A survey	SP II 2019-205
Puja Bhattacharya , Jeevant Rampal Contests within and between groups	SP II 2019-206
Kirby Nielsen, Puja Bhattacharya, John H. Kagel, Arjun Sengupta Teams promise but do not deliver	SP II 2019-207
Julien Grenet, Yinghua He, Dorothea Kübler Decentralizing centralized matching markets: Implications from early offers in university admissions	SP II 2019-208
Joerg Oechssler, Andreas Reischmann, Andis Sofianos The conditional contribution mechanism for repeated public goods – the general case	SP II 2019-209
Rustamdjan Hakimov, C.-Philipp Heller, Dorothea Kübler, Morimitsu Kurino How to avoid black markets for appointments with online booking systems	SP II 2019-210
Thibaud Pierrot Negotiation under the curse of knowledge	SP II 2019-211
Sabine Kröger, Thibaud Pierrot What point of a distribution summarizes point predictions?	SP II 2019-212
Sabine Kröger, Thibaud Pierrot Comparison of different question formats eliciting point predictions	SP II 2019-213

Research Unit: **Economics of Change**

Kai Barron, Steffen Huck, Philippe Jehiel Everyday econometricians: Selection neglect and overoptimism when learning from others	SP II 2019-301
Marta Serra-Garcia, Nora Szech The (in)elasticity of moral ignorance	SP II 2019-302
Kai Barron, Robert Stüber, Roel van Veldhuizen Motivated motive selection in the lying-dictator game	SP II 2019-303
Maja Adena, Steffen Huck Can mass fundraising harm your core business? A field experiment on how fundraising affects ticket sales	SP II 2019-304
Maja Adena, Rustamdjan Hakimov, Steffen Huck Charitable giving by the poor: A field experiment on matching and distance to charitable output in Kyrgyzstan	SP II 2019-305
Maja Adena, Steffen Huck Personalized fundraising: a field experiment on threshold matching of donations	SP II 2019-306
Kai Barron Lying to appear honest	SP II 2019-307

Research Unit: **Ethics and Behavioral Economics**

Daniel Parra, Manuel Muñoz-Herrera, Luis Palacio The limits of transparency as a means of reducing corruption	SP II 2019-401

All discussion papers are downloadable:
http://www.wzb.eu/en/publications/discussion-papers/markets-and-choice