International Journal of Forecasting 41 (2025) 517–531

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Cognitive reflection, arithmetic ability and financial literacy independently predict both inflation expectations and forecast accuracy☆,☆☆

David A. Comerford

University of Stirling, United Kingdom of Great Britain and Northern Ireland

ARTICLE INFO

Article history:
Dataset link: https://osf.io/bpcjm/

Keywords:
Expectations
Inflation
Financial literacy
Numeracy
Cognitive reflection

ABSTRACT

Cognitive reflection is defined as the tendency to detect and check intuitive errors and has been found to predict forecast accuracy in a range of domains. The current research demonstrates in a purpose-designed survey that a question in the Survey of Consumer Expectations serves as a test of cognitive reflection. Using this measure, I demonstrate for the first time in a time-series of inflation expectations that cognitive reflection is associated with greater forecast accuracy. I then apply this insight to interrogate the spike in inflation expectations that occurred over the year 2021. The data rule out that the spike was driven by respondents low in cognitive reflection, who are most vulnerable to overreacting to recent news. These results are insightful for the use of survey data not only in forecasting inflation but in forecasting more generally.

1. Research question

While serving as governor of the US Federal Reserve, Janet Yellen observed "I see the standard framework that sees inflation as being largely determined by inflation expectations… as a reasonably sound framework for thinking about inflation" (Yellen, 2018). Given the primacy of inflation expectations in policymakers' models of the inflation process, it is a useful enterprise to minimize noise and extract as much signal as possible from expectations measures.

The current research contributes toward that project. It reports the results of a purpose-designed survey experiment that tests whether one of the numeracy items collected as part of a state-of-the-art survey of inflation expectations measures cognitive reflection. Cognitive reflection is described as the tendency to check and detect intuitive errors (Sinayev & Peters, 2015). Recent research finds that cognitive reflection measures are highly predictive of the sort of reasoning that would promote the accuracy of inflation forecasts (D'Acunto, Hoang, Paloviita, & Weber, 2023). More generally, high cognitive reflection is a reliable characteristic of the "superforecasters" identified by the Better Judgment Project (Budescu, Himmelstein, & Ho, 2021; Himmelstein, Atanasov, & Budescu, 2021; Mellers et al., 2015; Tetlock & Gardner, 2016). The first contribution of this paper is to show that the state-of-the-art Survey of Consumer Expectations (SCE) has within it a measure of cognitive reflection. The second contribution demonstrates for the first time in a time-series of inflation expectations that cognitive reflection is associated with greater forecast accuracy. This gain in forecast accuracy is independent of respondents' arithmetic ability or financial literacy, two traits that have previously been found to predict forecast accuracy and that are also found

☆ The numerical results presented in the manuscript were reproduced by CASCaD, up to some minor discrepancies, on the 17th September 2024.

☆☆ Thanks to anonymous reviewers and to seminar participants at the behavioural economics seminar series at University College Dublin, especially Karl Whelan, for insightful comments.

E-mail address: david.comerford@stir.ac.uk.

https://doi.org/10.1016/j.ijforecast.2024.06.011

to do so in these data. The third contribution investigates the spike in inflation witnessed over the summer of 2021 with the goal of discerning whether it was concentrated among those who are well-informed and reflective or whether it was concentrated among those who are inclined to overreact to recent events.

This last contribution answers a question raised by European Central Bank executive board member Philip Lane, when he asks *whose* expectations are driving the uptick in inflation expectations that were observed over the years 2020–2021. In one story, "more attentive traders, experts and individuals may identify more quickly a persistent shift in inflation dynamics". Another story is that the uptick is driven by those who "over-react to high spot inflation" (Lane, 2022, p.4).

Philip Lane's question is material not only for interpreting survey expectations in the current situation but for testing their validity more generally. One explanation for the uptick that would cause survey data to *negatively* predict trends in realized inflation is if survey respondents wrongly infer transitory price increases from the recent past to foreshadow larger price increases to come. This explanation would give rise to extrapolation bias i.e. positive bias when a variable is at its peak (Da, Huang, & Jin, 2021; Hirshleifer, Li, & Yu, 2015; for inflation forecasts specifically see Pfajfar & Santoro, 2010). Evidence that suggests overreaction to recent news might drive the uptick is that survey respondents overweight large price increases they observe on shopping trips when answering inflation expectations questions (D'Acunto, Malmendier, Ospina, & Weber, 2021; D'Acunto, Malmendier, & Weber, 2021).

Cognitive reflection is an especially insightful characteristic for answering Philip Lane's question. Overreaction to recent news is the flipside of underweighting base rates. Those high in cognitive reflection are more likely to use base rates appropriately than are their less reflective peers (e.g. Alós-Ferrer & Hügelschäfer, 2016; Hoppe & Kusterer, 2011; Vartanian et al., 2018). Accordingly, recent research suggests that cognitive reflection reduces extrapolation bias (Barahona, Cassella, & Jansen, 2021). In short, we would expect those who are high in cognitive reflection to be less likely than the average respondent to overweight high spot inflation when formulating their inflation forecast.

On the other hand, if we were to observe that the uptick in inflation expectations had been driven by respondents who are highest in cognitive reflection, arithmetic ability and financial literacy then it would be hard to dismiss the uptick in expectations data as mere noise. After all, these are the most informed and considered respondents from across the country. If these respondents show an uptick in their reported expectations then it would be worth attending to that uptick.

In what follows I find that the uptick in inflation expectations over the year 2021 was greatest among respondents who are highest in each of cognitive reflection, financial literacy and arithmetic ability. In exploratory analysis, I identify a subgroup of respondents who are highest in all three of these capabilities and I test the accuracy of their forecasts. Whereas the previous literature has found that in general households are positively biased in their inflation forecasts (Coibion, Gorodnichenko, & Weber, 2022), the mean squared error of these respondent's forecasts is not reliably different to zero.

The next section provides a background context that motivates the specific hypotheses that follow. Section 3 reports the empirical methods and analyses. Section 4 concludes with implications for forecasting.

2. Background

2.1. The uses of household survey measures of inflation expectations

In principle, household surveys on inflation expectations ought to provide data that is unique in its usefulness to policymakers. If household surveys do what they purport to do then the inflation expectations they return would serve two distinct functions. First, they would give an accurate snapshot in real time of the inflation expectations that the population of US households is bringing to bear on their economic decisions. The second function to offer a leading indicator that forecasts future inflation.

At time of writing, the Federal Reserve Bank of New York's website is reporting median year ahead inflation expectations of 3.3%. This figure comes from the Survey of Consumer Expectations (SCE), which recruits a representative sample of US residents each month. This section summarizes evidence that we should be wary of inferring that the median US resident will act as though they expect prices to rise by 3.3 percent. In light of that evidence, the section then considers whether inflation expectations reports can still serve a useful function. I conclude that their function as a snapshot of the population's inflationary beliefs is called into doubt but that there is no reason why they could not function as a valuable leading indicator.

The inflation expectations data collected in the SCE represents the state-of-the-art because it was designed by a collaborative team of economists and psychologists specialising in expectations elicitation (Armantier, Topa, Van der Klaauw, & Zafar, 2017; Bruine de Bruin & Fischhoff, 2017; Manski, 2017). The published measure derives from a series of questions that ask respondents for the percent chance that inflation over the next 12 months will fall in certain ranges e.g. −4% to −8%; 2%–4% etc. It is termed a density forecast measure because the answers to these individual questions are collated by survey administrators to impute a forecast density function.

The density forecast relies on respondents being able to translate their likelihood beliefs onto a percentage scale. But a substantial proportion of the population cannot (Comerford, 2019; Comerford & Robinson, 2017). It is not merely that the data returned is a noisy approximation of beliefs; it gives a biased impression of population beliefs and it causes biased inference (Bell, Comerford and Douglas, 2020; Comerford, 2021).

The headline inflation expectations measure published by the New York Fed is missing data from those who failed to deliver a usable density forecast, which is over 2 percent of respondents. Those respondents are not random – they are disproportionately less educated and low in arithmetic ability. Hence, the inflation expectations data published on the New York Fed's website are no longer representative of the population.

Of those that do return a usable forecast, a non-trivial minority deliver density forecasts that are directionally wrong. There is a question in the SCE that simply asks "will there be inflation or deflation over the next 12 months?" About one out of every twenty respondents imply deflation when the response to the verbal question shows that they meant inflation (Comerford, 2023).

So one concern is that many respondents fail to articulate a coherent response when asked for their inflation [3] expectations. But separately, there is a concern regarding those responses that appear coherent. Although these purport to capture the inflationary beliefs that the respondent brings to bear on their economic decisions, they demonstrate a number of anomalies.

One anomaly is that the inflation expectations that respondents report evolve with respondent experience; respondents form their inflation expectations with reference to a different mental model after they have been through several waves of the survey than they did when they first came into the survey (Kim & Binder, 2023). This implies that the very act of answering the survey shapes the inflation expectations that respondents form, which means that these survey respondents hold inflation expectations that are no longer representative of the population from which they were randomly drawn.

A second anomaly is that a substantial subset of respondents fail to act in line with the survey measures of subjective expectations they report (Drerup, Enke, & Von Gaudecker, 2017). It is not that these respondents misreport in the survey their best judgment of future outcomes; rather it is that they feel insufficiently confident in their individual judgments and so they instead rely on social cues to inform their real-world economic behaviour e.g. they might mimic the economic choices of friends whose judgment they esteem.

Whereas the concerns expressed in previous paragraphs suggest that household surveys will give a biased impression of the beliefs that inform the household decisions, they do not necessarily compromise household survey data as a valid leading indicator. When it comes to forecasting inflation, all that matters is predictive power. A valid leading indicator need not capture data from a representative sample of the population nor need it capture the beliefs that inform respondents' real-world decisions. Notwithstanding the manifest deficiencies of household surveys of expectations, there is still reason to believe that they could prove uniquely useful as a tool for forecasting inflation.

Household surveys are well placed to harness the wisdom of the crowd. The wisdom of the crowd refers to the tendency for the average judgment of a diverse group of observers to accurately predict an outcome. It results from the tendency for independent estimates to be unbiased and distributed with random error so that the mean estimate matches the true value (Da et al., 2021; Surowiecki, 2005). Inflation has good properties to be forecast by a wisdom of the crowd approach. Inflation is a continuous numerical variable and so there is scope for over- and under-estimates to cancel one another out, which is the crucial mechanism through which the wisdom of the crowd operates. Inflation is an outcome about which respondents are likely to have observed some diagnostic information e.g. they will have recently seen gas and grocery prices and they have incentives to attend to these prices. They may additionally have private information that predicts future inflation e.g. they might know that workers in their sector will receive a pay-rise in the coming quarter. Importantly, household surveys tap into a broad swathe of the population and hence capture differing perspectives informed by distinct and relevant private information.

Still, the anomalies reviewed at the beginning of this section might also undermine survey data on expectations as a leading indicator. We are left wondering whether data showing co-movements of inflation expectations and realized inflation result from reverse causation i.e. "survey respondents adjusting their forecasts in response to past inflation rates" (Mehra, 2002, p.23).

An additional concern is that the wisdom of the crowd is undermined by bias. Given a tendency in population surveys for inflation to be overestimated (Coibion et al., 2022), we might worry that the mean estimate returned by the population will be an overestimate. If we could identify those respondents who are not biased in their forecasts, however, it might be that their forecasts are a particularly powerful leading indicator. The current research is a step towards this goal.

The wisdom of the crowd operates best in a subsample of respondents who are adept in probabilistic reasoning, well informed on the subject area and who are high in cognitive reflection (Mellers et al., 2015). The SCE allows us identify such people. Because it collects data on numeracy, financial literacy and – we shall see in the next section – cognitive reflection, the SCE allows us identify which respondents are well equipped to return informed forecasts.

2.2. Explaining heterogeneity in forecasts

The current research contributes to a literature on heterogeneity in forecasts. There is variety across individuals in their inflation expectations (Mankiw, Reis, & Wolfers, 2003). Some variety would be expected because of idiosyncrasies in the information individuals encounter and in how they process that information. Indeed, this sort of variety is conducive to the successful operation of the wisdom of the crowd – disparate information sets and processing styles would lead errors to be less correlated with one another and hence less likely to lead to bias when aggregated.

But some variation in forecasts appears to capture a bias. For instance, households use the prices observed in their grocery shopping as a basis for their inflation updates (Cavallo, Cruces, & Perez-Truglia, 2017; D'Acunto, Malmendier, Ospina, & Weber, 2021; D'Acunto, Malmendier, & Weber, 2021; Niu & Harvey, 2023). To the extent that grocery price increases are non-representative of changes in the general price level, this approach to inflation forecasting will lead to bias.

A separate mechanism is variation in how groups interpret news. Coibion et al. (2022) and D'Acunto, Hoang, and Weber (2022), D'Acunto, Malmendier, Ospina, and Weber (2022) show that policy communication impacts differently the inflation expectations of various demographic groups. Prior experience of inflation also informs expectations (Cavallo et al., 2017; Malmendier & Nagel, 2016). Returning to the role of observed price changes in expectations formation, D'Acunto, Malmendier, Ospina, and Weber (2021), D'Acunto, Malmendier, and Weber (2021) find that inflation expectations are disproportionately informed by large price changes and that this result is especially strong among those who go grocery shopping less frequently.

The previous paragraph highlights the importance of information processing and memory in expectations formation. This links to research on cognitive characteristics that predict inflation expectations. Forecast errors are predicted by financial literacy (Bruine de Bruin, Vanderklaauw, Downs, Fischhoff, Topa, Armantier, 2010; Burke & Manz, 2014) and by IQ, particularly arithmetic intelligence (D'Acunto, Hoang, Paloviita, & Weber, 2019; D'Acunto et al., 2023). These results suggest that, given a set of data, certain respondents will formulate more accurate forecasts than will others. Supporting this claim, Armantier, Bruine de Bruin, Topa, Van Der Klaauw, and Zafar (2015) find in an experimental setting that those higher in financial literacy and numeracy are especially likely to update their inflation expectations in response to new information and are especially likely to alter their choice behaviour in line with their updated expectations.

One way to read the current research is as a conceptual replication of the experimental result reported in Armantier et al. in a real-world context: where Armantier et al. (2015) found that financial literacy and arithmetic ability were associated with larger reactions to experimentally-assigned information, the current research decomposes an aggregate response to real-world events, the inflation spike of summer 2021, to assess whether those same cognitive types were most responsive.

A second contribution is to investigate the role of cognitive reflection in inflation expectations. Cognitive reflection is described as the tendency to check and detect intuitive errors (Sinayev & Peters, 2015). Frederick (2005) introduced a multiple item scale to measure cognitive reflection. As befits a test for detection and correction of intuitive errors, each item in the scale asks a question that calls to mind an intuitive answer that, on closer inspection, turns out to be wrong. For example, one asks "A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? (in cents)". While the intuitive answer that a ball costs 10 cents jumps readily to mind, a respondent who demonstrates a close and attentive reading of the question and who is willing to think past whatever initial answer spontaneously jumps to mind will detect that an answer of 10 cents would lead the bat to cost just 90 cents more than the ball and so cannot be correct. D'Acunto et al. (2023) find that successfully answering Frederick's cognitive reflection questions predicts inflation expectations as well as various skills that are related to rational expectation formation e.g. understanding of inflation as a concept; level-*k* reasoning; higher than average accuracy in forecasting of two simulated data sequences.

The results in D'Acunto et al. (2023) echo research documenting that people who are high in cognitive reflection perform better at forecasting in a range of domains. Studies that ask about the probability of a large range of future events reliably find that, controlling for other individual differences, respondents who are high on cognitive reflection deliver more accurate and better calibrated forecasts than do other respondents (Budescu et al., 2021; Himmelstein et al., 2021; Mellers et al., 2015).

One reason why cognitive reflection would be associated with more accurate forecasts concerns extrapolation bias. Various models converge to explain extrapolation bias as a symptom of overweighting top-of-the-mind information and of underweighting base rates (e.g. Bordalo, Gennaioli, & Shleifer, 2018; Comerford, 2019). Those high in cognitive reflection are more likely to use base rates appropriately than are their less reflective peers (e.g. Alós-Ferrer & Hügelschäfer, 2016; Hoppe & Kusterer, 2011; Vartanian et al., 2018). Recent research suggests that higher cognitive reflection reduces extrapolation bias (Barahona et al., 2021). We would expect those who are high in cognitive reflection to be less likely than the average respondent to overweight high spot inflation when formulating their inflation forecast.

Taken together, the previous literature offers a profile of which types of respondents are likely to be most accurate in their inflation forecasts and which are most likely to overreact to high spot inflation. If we see that the uptick in inflation expectations is driven by respondents who are low in financial literacy, arithmetic ability or cognitive reflection then we urge caution in taking the uptick in these survey measures of inflation expectations as a useful leading indicator.

3. Empirics

3.1. Pretest demonstrating that the SCE measures cognitive reflection

The SCE data include a variable called Num_Cat. It categorises each respondent as either high in numeracy or low in numeracy on the basis of that respondent's answers to five numeracy questions. Four of those questions are listed in Table 1 as measures of arithmetic ability.

The remaining item is QNUM2. It reads: "Let's say you have $200 in a savings account. The account earns ten per cent interest per year…how much will you have in the account at the end of two years?" There are two answers to this question that demonstrate arithmetic ability: $240 and $242. The correct answer of $242 additionally demonstrates a tendency to check and detect intuitive errors, which is the defining characteristic of cognitive reflection (Sinayev & Peters, 2015). Hence, Comerford (2023) advocates that QNUM2 captures cognitive reflection.

Here I present a purposely-designed survey to formally determine whether QNUM2 measures cognitive reflection. The first is a correlational test: if QNUM2 measures cognitive reflection independent of arithmetic ability then correct responses to QNUM2 should be predicted in a

Table 1
Data measuring cognitive capabilities in the SCE.

Variable	Measure	% correct
Inflation expectation	In your view, what would you say is the percent chance that, over the next 12 months… the rate of inflation will be [12%, 8%, 4%] or higher	
	Cognitive Capacities	
Financial literacy	QNUM8: Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After one year, how much would you be able to buy with the money in this account?	83
	QNUM9: Please tell me whether this statement is true or false: Buying a single company's stock usually provides a safer return than a stock mutual fund	92
Arithmetic ability	QNUM1. In a sale, a shop is selling all items at half price. Before the sale, a sofa costs $300. How much will it cost in the sale?	94
	QNUM3. In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1000 people each buy a single ticket from BIG BUCKS?	82
	QNUM5. If the chance of getting a disease is 10 percent, how many people out of 1000 would be expected to get the disease?	89
	QNUM6. The chance of getting a viral infection is 0.0005. Out of 10,000 people, about how many of them are expected to get infected?	78
Cognitive Reflection	QNUM2. Let's say you have $200 in a savings account. The account earns ten per cent interest per year. Interest accrues at each anniversary of the account. If you never withdraw money or interest payments, how much will you have in the account at the end of two years?	51

multivariate regression by respondents' scores on a numeracy scale and independently by their scores on a cognitive reflection scale. The second test is an experimental test of a hypothesis rooted in a conceptual distinction between arithmetic ability and cognitive reflection. Arithmetic ability is very stable across contexts – ask a respondent to a survey to calculate 10% of 200,000 and, if the person is high on arithmetic ability, you will get the correct answer regardless whether the question is placed among questions relating to arithmetic ability or financial literacy or cognitive reflection or health behaviours or any other module of questions.

Cognitive reflection is different. There are contexts in which a respondent will be on the alert for intuitive errors e.g. when answering a set of brain teasers. This conceptual distinction between arithmetic ability and cognitive reflection leads to the following hypothesis: if a respondent has just encountered the set of questions that form the cognitive reflection task, then they will be more likely to answer any question that measures cognitive reflection correctly.

Harnessing this insight, I experimentally vary the location of QNUM2 in a survey experiment. For half the sample, QNUM2 is located among the numeracy questions. For the other half of the sample, it is located among the cognitive reflection questions. If QNUM2 captures cognitive reflection over and above arithmetic ability then we would expect the rates of answering QNUM2 correctly to be higher when it is placed among the cognitive reflection questions.

3.1.1. Procedures used for pretest

In this preregistered pretest,[1] I aimed to recruit 500 respondents via Prolific.co to answer a websurvey. Ultimately, 512 complete responses were delivered (mean age = 42; 59% female).

On entering the survey, respondents were randomly assigned to either a CRTprime condition or to a control condition.

In the CRTprime condition, QNUM2 appeared on the same screen as the four items that comprise the Cognitive Reflection Task (Thomson & Oppenheimer, 2016; sample item: "Emily's father has three daughters. The first two are named April and May. What is the third daughter's name?"; full scale reported as appendix A). In the control condition, QNUM2 appeared among the items that are reported in Table 2 measuring arithmetic ability, just as in the SCE.

3.1.2. Results

Each respondent is scored out of four for their arithmetic ability and their cognitive reflection by counting how many questions they got correct on the relevant scale (responses to QNUM2 are not included in scoring either arithmetic ability or cognitive reflection). The correlation between arithmetic ability and cognitive reflection is of medium size ($r = 0.21$).

One piece of evidence that correct response to QNUM2 can be considered a measure of cognitive reflection is that it is predicted independently by arithmetic ability score and by cognitive reflection score. Table 2 presents the results of a binary logistic regression of correct response to QNUM2. It shows that, holding constant arithmetic ability, respondents higher on cognitive reflection were more likely to correctly answer QNUM2.

We earlier saw that the correlation between arithmetic ability and cognitive reflection was less than $r = 0.21$, and hence far short of the threshold for collinearity of $r = 0.80$ (Dormann et al., 2013). Further evidence that cognitive reflection and arithmetic ability are two distinct

[1] https://aspredicted.org/4LM_RGG

Table 2
Binary logistic regression of correct response to QNUM2.

Arithmetic ability score	1.022**
	(0.148)
Cognitive reflection score	0.714**
	(0.105)
Age	0.012
	(0.007)
Female	−0.246
	(0.216)
Constant	−5.852**
	(0.681)
Pseudo R-squared	.2130
N	503

* p<0.05; ** p<0.01, standard errors in parentheses. Arithmetic ability scored out of 4 as a count of how many of the questions reported in Table 1 the respondent got correct. Cognitive Reflection score is measured out of 4 as a count of how many of the Thomson and Oppenheimer (2016) CRT questions the respondent got correct.

constructs is that there is a deterioration in model fit if we remove cognitive reflection from the model detailed in Table 2. The model that excludes cognitive reflection scores delivers a pseudo r-squared of 0.1321, which compares to a pseudo r-squared of 0.2130 when cognitive reflection scores are included. This result demonstrates that, independent of arithmetic ability scores, cognitive reflection scores predict responses to QNUM2.

A second piece of evidence that QNUM2 measures cognitive reflection comes from my survey experiment. A binary logistic regression of survey condition shows that random assignment was successful in delivering two groups whose average scores for arithmetic ability and cognitive reflection were similar (arithmetic ability: $n = 512$, $z = 0.18$, $p = 0.83$; cognitive reflection: $n = 512$, $z = 0.63$, $p = 0.58$). In other words, respondents who would later be assigned to the control condition showed capacities that suggest they would be no less able to answer QNUM2 correctly than those who would later be assigned to the priming condition. (There was however a marginally significant difference in age across conditions 32 [$z = 1.69$, p = .096]).

Fig. 1 depicts the result of the experiment. The key result is that respondents who had been randomly assigned to the CRTprime condition were more likely to have answered QNUM2 correctly: 35.2% of respondents answered $242 in the control condition compared to 43.6% in the CRTprime condition (df = 512, $z = 1.95$, p = .051; controlling for age, df = 511, $z = 1.96$, p = .050).

3.1.3. Insight

This result is experimental evidence that the question QNUM2 captures cognitive reflection in addition to arithmetic ability. If QNUM2 were solely measuring arithmetic ability then there is no reason why varying its place in a survey would cause it to be answered correctly. The fact that QNUM2 was more likely to be answered correctly when it was placed among cognitive reflection items signals that a reflective mindset is conducive to answering QNUM2 correctly and hence that QNUM2 measures whether the respondent has a reflective mindset.

This result provides experimental evidence that aligns with the correlational results described in Table 2. It documents that, independent of arithmetic ability, cognitive reflection predicts correct response to QNUM2.

3.2. Does cognitive reflection improve the accuracy of inflation expectations?

3.2.1. Methods

The SCE collects data each month from over 1,000 randomly selected households, stratified only by Census division (Armantier et al., 2017). Respondents deliver a probability distribution regarding inflation over the next 12 months.

In their first wave answering the SCE, a respondent is asked a range of questions that measure cognitive capacities (Table 1). If a respondent answered later waves, I impute their cognitive capacities from their first wave responses. Financial literacy and arithmetic ability are measured as a count of how many questions respondents got correct (out of 2 for financial literacy; out of 4 for arithmetic ability).

3.2.2. Results from 2013–2020

Table 3 describes the mean inflation expectations reported by respondents highest in each cognitive capability compared with those in the next highest category up to January 2020. This period is selected as a baseline when inflation was low and relatively stable (inflation never exceeded 4% over this period). June 2013 is the first wave of data collection for the SCE and data on cognitive reflection and arithmetic ability were collected in every wave since. Data on financial literacy started to be collected in May 2015 and so the results on financial literacy in Table 3 come from a shorter time period than do the results on the other cognitive capabilities. The Covid pandemic introduced volatility following January 2020 and so we leave the period after January 2020 to later analyses.

In OLS regressions of mean inflation reports (the SCE variable FDmean; central column of Table 3) that control for month-year dummies, each of financial literacy, arithmetic ability and cognitive reflection are negative predictors of inflation expectations. A key insight from Table 3 is that in this period of low and stable inflation, those who measured highest in each of these cognitive capacities reported significantly lower inflation expectations than did other respondents.

The final column of Table 3 shows that this difference in expectations translated into higher levels of accuracy. The question respondents were asked sought their expectations of inflation for the coming twelve months. When looking for accuracy, therefore, we look to realized inflation rates over the twelve-month period that followed their survey response and test for the mean squared error between that figure and the expectations reported by each respondent. We take as our inflation measure the seasonally-adjusted median CPI (FRED, 2022). Median CPI was chosen because this particular measure of inflation is recommended by the literature (Verbrugge & Zaman, 2024; Meyer, Venkatu, & Zaman, 2013; Tables A1 and A2 in the appendix report the results of Tables 3 and 4 using

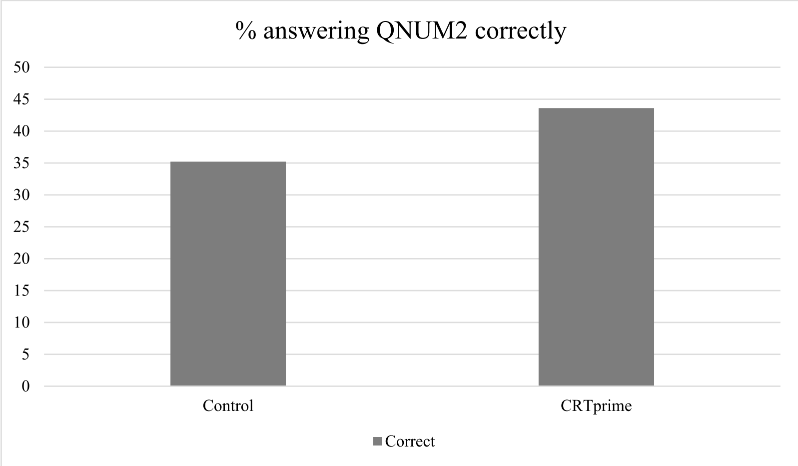

Fig. 1. Likelihood of answering correctly by experimental condition.
Note: Control condition saw QNUM2 among questions testing arithmetic ability. CRTprime condition saw QNUM2 presented among questions testing cognitive reflection.

Table 3
Inflation expectations by type (2013 – 2020).

Type	Mean inflation forecast (FDmean)	FDmean, t-stat, p – value	MSE, t-stat, p – value
High in Cognitive reflection vs. not ($n = 103{,}126$)	3.3% vs. 4.0%	24.44, < .001	67.21, < .001
Highest financial lit. vs. next highest ($n = 65{,}557$)	3.4% vs. 3.8%	9.46, < .001	43.21, < .001
Highest arithmetic vs. next highest ($n = 88{,}886$)	3.4% vs. 3.9%	13.81, < .001	37.23, < .001

Notes: The left-most column summarizes the mean inflation expectation reported by respondents in the top category and second-from-top category for each cognitive characteristic. Inflation expectation is measured as the mean of the density forecast (FD mean). The central column reports results from an OLS regression of FDmean that controls only for month / year where the sample is limited to those in the highest and next highest categories on each cognitive capability. The right-hand column reports results from an OLS regression of Mean Squared Error that controls only for month / year where the sample is limited to those in the highest and next highest categories on each cognitive capability. The arithmetic ability and cognitive reflection results use data from monthly waves of the SCE from June 2013 to January 2020. Financial literacy data are from May 2015 to January 2020.

an alternative measure of inflation recommended by an anonymous reviewer, CPI all items: total for the United States).

The Mean Squared Error is calculated as follows

$$\mathrm{MSE}_{i,t} = [\mathrm{FDmean}_{i,t} - \mathrm{MeanCPI}_{ttot+11}]_2$$

The hypothesis tested in the last column of Table 3 is that respondents $_i$ who are highest in financial literacy, arithmetic ability or cognitive reflection will, relative to respondents $_j$ who are in the next highest category on each of these capabilities, deliver more accurate forecasts i.e. the mean of their density forecast (FDmean) will more closely match the mean of realized inflation over the coming twelve months (Mean $\mathrm{CPI}_{t\,to\,t+11}$). The mean of realized inflation over the coming twelve months was calculated by summing median CPI for each of the 12 months from the moment of survey response through to 12 months later and then dividing that sum by 12 to deliver a mean. Table 3 reports the results of separate OLS regressions on each of the three cognitive capabilities.

We earlier saw that QNUM2 is capturing cognitive reflection independent of arithmetic ability. We might worry that financial literacy is collinear with cognitive reflection or, which would also trouble interpretation of the results reported in Table 3, that there may be multicollinearity. One immediate test we can run to assess whether QNUM2 is capturing something independent of financial literacy and arithmetic ability is a regression of getting QNUM2 correct on financial literacy and arithmetic ability. Specifically, we code a binary variable to indicate that the respondent got QNUM2 correct. Then we regress that binary variable simultaneously on financial literacy score and on arithmetic ability score. If QNUM2 is merely another measure of arithmetic ability and/or financial literacy then this regression will be highly predictive i.e. the resultant r-squared will be very high. Dormann et al. (2013) suggests that if the resultant r-squared exceeds 0.80 then whatever QNUM2 is picking up is so collinear as to be functionally indistinguishable from what has already been picked up by numeracy and/ or financial literacy. In fact, the resultant r-squared is just 0.15. This indicates that there is something captured by QNUM2 that is not captured by either financial literacy or arithmetic ability. Our experiment shows that what is picked up by

QNUM2 is cognitive reflection. In summary, the data recommend treating financial literacy, arithmetic ability and getting QNUM2 correct as three separate cognitive constructs. They further suggest that QNUM2 is measuring cognitive reflection. All told, these results recommend including financial literacy, arithmetic ability and QNUM2/ cognitive reflection as independent predictors of inflation expectations in a multiple regression model. Model 1 of Table 4 presents the results of this multiple regression model.

A core claim of this paper is that coding QNUM2 as a measure of cognitive reflection adds forecasting power. Model 2 of Table 4 reports the results that obtain if we treat QNUM2 as just another numeracy item instead of treating it is an independent variable. The SCE data administrators treat QNUM2 as one of 5 questions that measure numeracy (see discussion of the Num_Cat variable in the Section 3.3.1). With this in mind, users of SCE data would be expected to code QNUM2 as in Model 2. Hence, Model 2 is a useful baseline against which to compare Model 1. The data show that treating QNUM2 as an independent variable results in an improvement in model fit – the r-squareds rise from .2412 in Model 2 to .2419 Model 1. It is more appropriate here to report adjusted r-squareds because a model that includes a higher number of covariates enjoys a mechanical advantage over a model that includes fewer variables. The results are similar: the adjusted r-squared increases from .2357 in Model 2 to .2363 in Model 1. In summary, there is an improvement in model fit that results from treating QNUM2 as a measure of cognitive reflection relative to treating it as a measure of numeracy. Though it is slight in these data, this improvement in model fit indicates that QNUM2 is more usefully coded as a measure of cognitive reflection than as a measure of arithmetic ability.

The gain in model fit is larger if we restrict the sample to just those respondents who were answering the SCE for the first time. Table A3 in the online appendix shows that among these respondents the r-squared rises from .2704 in the baseline Model 2 to .2720 in Model 1 that treats QNUM2 as a measure of cognitive reflection (adjusted r-squareds: Model 1: .2269; Model 2: .2253). This pattern of results suggests that a respondent's level of cognitive reflection is especially predictive of their inflation expectations in the first wave of the SCE that they answer and that, with experience answering the SCE, those low in cognitive reflection come to deliver inflation expectations that more closely resemble those given by respondents high in cognitive reflection. This result is consistent with Kim and Binder's finding that, with experience of answering the SCE, respondents become more attentive to inflation.

The models reported in Table 4 control for the raw forecast made by each respondent, which is insightful for ruling out an uninteresting mechanism that could explain the univariate results. Coibion et al. (2022) shows that the average respondent in the US reported an inflation expectation that was positively biased over this time period. Reporting an inflation expectation that is less upward biased than that of the average respondent would mechanically deliver a reduction in MSE. In fact,

Table 4
Regression of mean square error of inflation forecasts.

	Model 1	Model 2
Raw inflation forecast (FDmean)	0.148**	0.151**
	(0.002)	(0.004)
Cognitive reflection	−0.451**	
	(0.018)	
Arithmetic ability	−0.245**	
	(0.011)	
Numeracy w. QNUM2		−0.314**
		(0.025)
Financial literacy	−0.232**	−0.234**
	(0.016)	(0.016)
Female	−0.403**	−0.415**
	(0.017)	(0.017)
Age	−0.004**	−0.004**
	(0.001)	(0.001)
Highschool or less	0.433**	0.440**
	(0.027)	(0.027)
$50–100k	−0.184**	−0.189**
	(0.021)	(0.021)
> $100k	0.366**	0.365**
	(0.020)	(0.020)
Month year dummies	X	X
Area dummies	X	X
Constant	2.152**	2.316**
	(0.157)	(0.156)
R^2	0.2419	0.2412
N	92,874	92,874

Notes: * p<0.05; ** p<0.01. Results of an OLS regression of Mean Squared Errors of respondent's inflation expectations, where errors are calculated by subtracting the mean of respondent's density forecast (FDmean) from the average seasonally-adjusted median CPI over the 12 months following the survey. The data used is from monthly waves of the SCE from May 2015 to June 2021.

you would have appeared more accurate than the average respondent to the SCE if you had simply reported the Fed's inflation target of 2% each and every month over the period 2013–2020. So, in order to rule out that it is merely low inflation expectations that account for the results of Table 3, the regressions in Table 4 control for the respondents' raw inflation forecast.

Since the dependent variable in Table 4 is mean squared error (MSE) rather than raw forecast, it is insightful to test on all available waves rather than merely those waves included in Table 3, when inflation was always low and stable. Hence Table 4 presents results from the start of financial literacy data collection in the SCE in May 2015 through to June 2021. It finds that each cognitive capacity is a distinct contributor to reducing the mean squared error of forecasts (financial literacy: $t = 18.26$, arithmetic ability $t = 37.04$; cognitive reflection $t = 36.56$; all $ps < .001$).

3.2.3. Insight

The foregoing shows that respondents who are high in cognitive reflection reported inflation expectations that turned out to be significantly more accurate than those of less reflective respondents. This gain in forecast accuracy is independent of the gains in forecast accuracy that result from being high in arithmetic ability or high in financial literacy.

3.3. Whose expectations drive the uptick in 2020–2021?

ECB board member Philip Lane asks whose expectations account for the spike in inflations expectations in 2021. In this analysis I focus on the question of whether cognitive reflection predicts an especially large uptick in inflation expectations. If those low in cognitive reflection are especially responsive, this would be worrisome because low cognitive reflection predicts extrapolation bias. If those high in cognitive reflection are especially responsive, this would be encouraging for the use of household survey data as a leading indicator because cognitive reflection is associated with higher forecast accuracy.

3.3.1. Methods

For this analysis, my primary dependent variable is the reported probability that inflation will exceed 4%, which is calculated by summing for each respondent the probabilities they assign to three ranges asked about in the SCE: 4%–8%; 8%–12%; >12%.[2] I choose this measure because it cleanly captures the outcome of core interest. If expectations respond to the jump in inflation that occurred during the summer of 2021 then we would expect to see a substantial increase in the answers respondents give when asked what is the percent chance that inflation will be exceptionally high i.e. greater than 4%.

The reader may wonder why I do not use as my dependent variable more standard measures of inflation expectations: the average of each respondent's density forecast (henceforth FDmean) or a point estimate collected in the SCE survey by asking respondents, "What do you expect the rate of inflation or deflation to be over the next 12 months?" The reason is that each of these measures is less sensitive to picking up individual differences in belief updates than is my preferred dependent variable. The next few paragraphs detail why this is so.

The FDmean suffers systematic non-response bias, with respondents who are lower in cognitive skills being less likely to deliver usable data (Comerford, 2023; D'Acunto and Weber, 2024). This selective non-response would be especially damaging to the current research, which is focussed on the role of cognitive skills in inflation expectations.

As for the point estimate, it delivers data that are very vulnerable to outliers because it allows respondents to provide unbounded values. During the decade that the SCE has been collecting inflation expectations, 1.2% of responses suggested inflation/ deflation would be exactly 50%. Taking the point estimates at face value implies that roughly 1 out of every 100 US residents was expecting a catastrophic upset to prices. A more plausible interpretation of these data is that they reflect the fact that respondents who want to express "I have no idea" tend to answer 50% (Fischhoff & Bruine De Bruin, 1999). This interpretation is supported when we investigate who answers 50%. Looking to the Num_Cat variable coded by the SCE administrators, those respondents categorised as "low numeracy" were 6.8 times more likely to answer that inflation/ deflation would be 50% than the high numeracy group ($z = 36.84$, $p < .001$). Some of these 50% responses are mere noise. This matters because these reports of 50% have an outsized influence on the mean point estimate returned by the sample as a whole, which is 0.3 percentage points higher when the 50% responses are included than when dropped. One consequence is that they will bias the mean point estimate returned by the sample as a whole. A more insidious consequence is that they will obscure updates to beliefs. Mean point estimates will be less sensitive to changing macroeconomic conditions because month after month there is a cohort of respondents answering with noise responses that have outsize influence on the mean.

The FDmean measure is also noisy. The complexity of providing a density forecast leads the FDmean to give a biased indication of respondents' inflation expectations. About 6% of respondents deliver an FDmean that indicates deflation despite reporting the opposite when asked the simple question "will there be inflation or deflation?" (Comerford, 2023). This bias is also especially common among those respondents categorized by the SCE administrators as low in numeracy.

Additionally, the FDmean measure is spuriously constrained. For inflation outcomes close to zero, respondents are asked to report the percent chance that inflation will fall within ranges that span 2 percentage points e.g. 0%–2%; 2%–4%. But for higher inflation outcomes, the bin size increases – respondents are asked for the percent chance that inflation will fall in the range 4%–8% and that inflation will exceed 12%. This feature of survey design guides the FDmean towards zero (Weber et al. 2022; Becker, Duersch, & Eife, 2023). This matters because it attenuates variation in reported inflation expectations, which inhibits our capacity to identify the factors that predict variation in inflation expectations.

For all of these reasons, the FDmean and point estimate measures obscure meaningful updates to beliefs that inflation will be especially high in the future and hence inhibit our ability to identify what type of respondent makes such updates. For completeness Table 6 reports the results that obtain with each of these measures as the dependent variable. As it happens, those results deliver contradictory results across the FDmean and point estimate measures, demonstrating that at least one of these conventional measures of inflation expectations delivers a misleading answer to the research question of interest.

Returning to our analyses, a second hypothesis concerns the uncertainty around respondents' inflation forecasts. The SCE reports for each density forecast its interquartile range (IQR). The type of respondents who feel confident that they can predict inflation with a high degree of precision will return density forecasts with the mass of probabilities assigned to a narrow range of inflationary outcomes, resulting in a small IQR. Respondents who are very uncertain will distribute their probabilities across a wider range of outcomes and hence their IQR will be larger. By looking to respondent's IQRs we can identify which type of respondents feel especially confident in their forecasts and which feel more equivocal.

Ideally, I would test for differences in updates to forecasts at the level of individual respondents. The SCE collects data longitudinally for just 12 months and so it is not

[2] We are constrained by the ranges asked about in the SCE and so cannot code the dependent variable to be reported probability that inflation will exceed 3% or 5%.

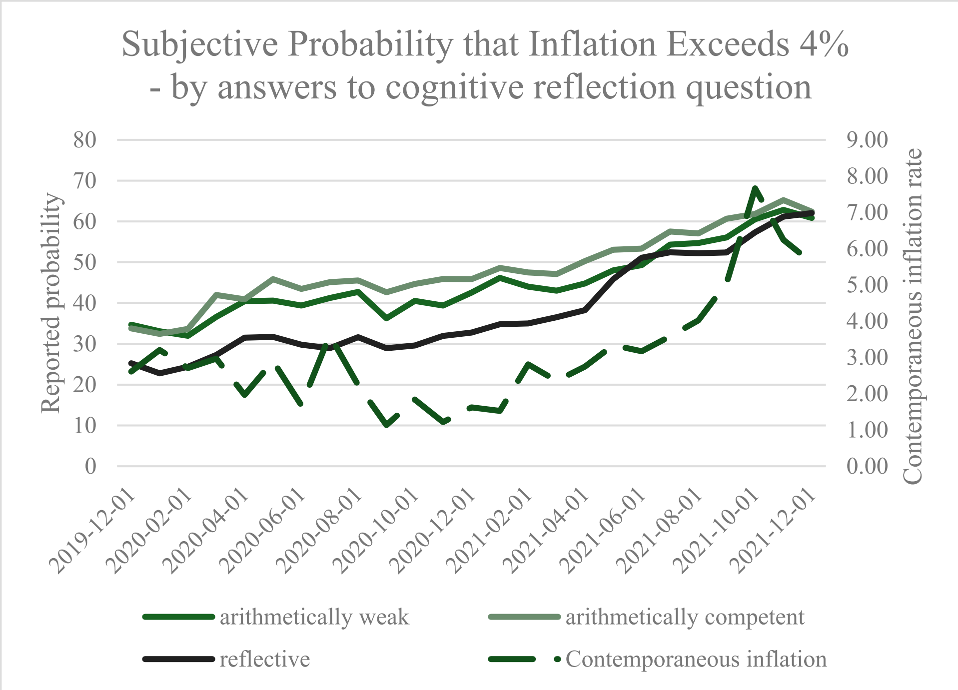

Fig. 2. Reported percentage chance that inflation exceeds 4%, by responses to the cognitive reflection question, QNUM2.
Note: Reflective answered QNUM2 with the compound interest answer of \$242. *Arithmetically competent* answered with the simple interest answer, \$240. *Arithmetically weak* answered neither \$242 nor \$240.

possible to follow the updates of individual respondents over longer periods. Instead of testing at the level of individual respondent, I exploit the random sampling process of the SCE and make comparisons across different *types* of respondent e.g. a reflective type vs. a less reflective type. This repeated cross-sectional approach is equivalent to a longitudinal analysis as long as the composition of respondents is not expected to change month-on-month. That assumption is satisfied because the SCE draws a random sample each month (Armantier et al., 2017).

I run an OLS regression of the form described in Model a:

$$\%{>}4\%\ \textit{inflation} = b1(2021) + b2(\textit{reflect}) + b3(\textit{reflect} * 2021) + b4(\textit{time trend}) + b5(\textit{reflect} * \textit{time trend}) \tag{a}$$

In this model, *2021* is the event indicator that distinguishes the high inflation period from an earlier baseline period. The spike in inflation occurred in the Summer of 2021; as depicted in Fig. 2, CPI had been 2.8% in July but jumped to 5.0% by September. The variable *2021* takes a value of 1 for forecasts made in the months October – December 2021 and takes a value of 0 for all other time periods. Independent of that event, I include a *time trend* variable that is coded from 1 = June 2013 for each monthly wave of data through 107 = April 2022. It captures any time trend in inflation expectations that is independent of the 2021 inflationary shock.

Reflect is an indicator variable that takes a value of 1 if the respondent answered QNUM2 correctly and a value of zero otherwise. The *reflect*time trend* interaction term captures any time trend that was unique to reflective respondents versus the rest of the sample.

The key test in model *a* concerns the interaction term *reflect*2021*, as measured by *b3.* If *b3* is positive and significant then the upshift in inflationary expectations from before the inflationary event to after the event was greater among the reflective types than among less reflective types. We would conclude then that cognitive reflection is associated with greater expectations updating in response to this inflationary shock.

H1. Respondents who are high in cognitive reflection display larger updates to their inflation expectations than do respondents lower in cognitive reflection.

I then run analogous models to that testing for an effect of cognitive reflection using financial literacy and arithmetic ability respectively as independent variables.

H2. Respondents who are high in financial literacy display larger updates to their inflation expectations than do respondents lower in financial literacy.

H3. Respondents who are high in arithmetic ability display larger updates to their inflation expectations than do respondents lower in arithmetic ability.

After testing the effect on updating of each cognitive capability in turn, I then run a multivariate regression that tests whether cognitive reflection has an independent effect, over and above that of financial literacy or arithmetic ability. As a baseline for comparison, I report the results of a model in which I treat QNUM2 as an additional numeracy item instead of treating it as a measure of cognitive reflection. Finally, I test for the effects of each cognitive trait on uncertainty, as measured by the interquartile range of respondents' forecasts.

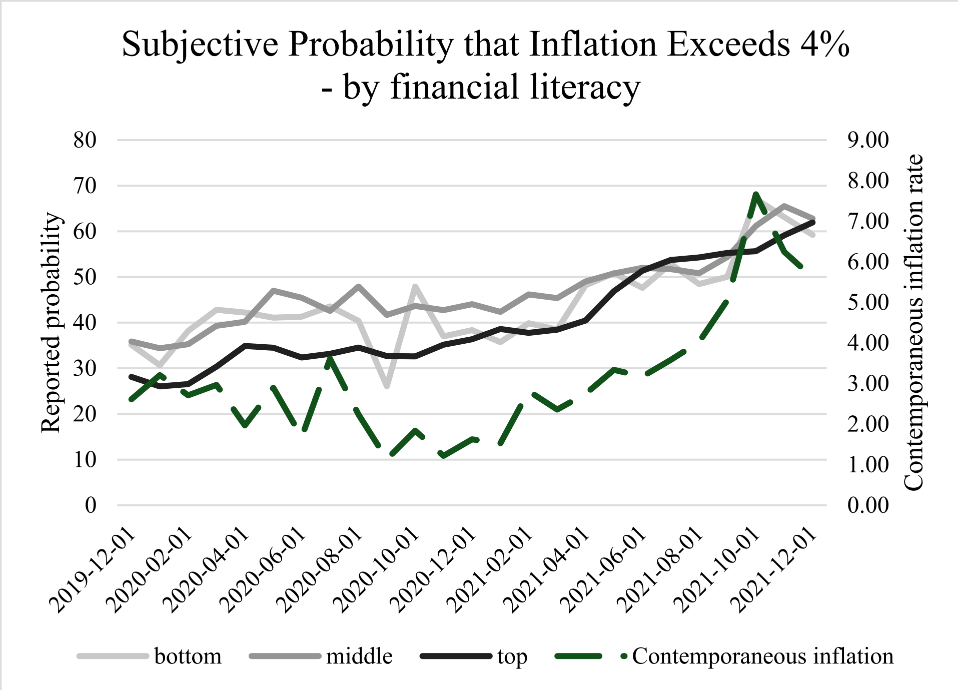

Fig. 3. Reported percentage chance that inflation exceeds 4%, by financial literacy.
Note There were 2 financial literacy questions – *bottom* answered neither correct; *middle* answered one correct; *top* answered both correct.

3.3.2. Results

Each figure depicts the mean probability assigned to >4% inflation over time alongside the contemporaneous rate of inflation. Fig. 2 distils the key contribution. It depicts the uptick in inflation expectations as a function of how respondents answered QNUM2. *Reflective* respondents are those who answered correctly, $242 (51% of the sample). *Arithmetically competent* respondents are those who answered $240 (23% of the sample). *Arithmetically weak* respondents failed to answer either $242 or $240 (26% of the sample). The uptick in inflation forecasts was similar among those who are *arithmetically competent* and those who are *arithmetically weak*. These groups both begin and end the data series with virtually identical forecasts. In contrast, the r*eflective* group began the data series by reporting a much lower chance of high inflation than either the *arithmetically competent* or the *arithmetically weak* groups. All three groups end the data series with similar forecasts. Hence the uptick in inflation forecasts was especially large among *reflective* respondents.

The crucial parameter *b3* in the OLS regression reported as equation *a* above is positive and significant. There is a reliable interaction effect of *reflect*2021* after controlling for time trends and main effects of 2021 and cognitive reflection ($b = 7.54$, $t = 6.45$, p <.001).

Fig. 3 presents the results for financial literacy. An OLS regression equivalent to that above, but replacing cognitive reflection with financial literacy at each point, finds a positive and statistically significant interaction effect of financial literacy and 2021 ($b = 7.82$, $t = 6.59$, p =.001).

The regression reported in Table 5 extends Model *a* to incorporate arithmetic ability score, financial literacy and their respective interactions with *2021*. To take account of preexisting trends in forecasts by different types, each cognitive capacity is interacted with a time variable in Table 5. Additionally, Table 5 takes account of variation in the random samples recruited across waves of the SCE by controlling for the demographic and geographic characteristics of respondents. Because we have already seen that QNUM2 captures something independent of financial literacy and of arithmetic ability, we can consider the coefficient on *reflect* in Model 1 of Table 5 to measure the effect on forecasts of being high in cognitive reflection holding constant all other effects, including those of financial literacy and arithmetic ability. A side-effect of controlling for financial literacy is that we cannot analyse data prior to May 2015, when the SCE started collecting financial literacy data.

Model 1 of Table 5 shows that the coefficient on the *cognitive reflection*2021* interaction is positive and statistically significant after controlling for all other effects and the time trend. Hypothesis H1 is supported: Over this period when inflation spiked, especially large increases in inflation expectations were observed among those high in cognitive reflection.

Model 2 of Table 5 is reported as a baseline against which to measure the added value of treating QNUM2 as a measure of cognitive reflection. Model 2 treats QNUM2 as merely another numeracy item. In Model 2, the r-squared is .1138, which rises to .1142 in Model 1. The conclusion is unchanged if we look at adjusted r-squareds, which take account of the number of covariates included in each model. Here model fit rises from .1084 in the baseline Model 2 to .1087 in Model 1.

This improvement in model fit is unsurprising given the results depicted in Fig. 2. It showed considerably less updating among respondents who answered QNUM2 with the arithmetically competent but inattentive response of $240 than among those respondents who showed cognitive reflection by correctly identifying QNUM2 as asking for compound interest.

The gain in model fit from treating QNUM2 as measuring cognitive reflection is even larger among those respondents who were answering the SCE for the first time. Table A4 in the online appendix shows that among these respondents the r-squared rises from .1182 in the baseline Model 2 that treats QNUM2 as a measure of numeracy to .1210 in Model 1 that treats QNUM2 as a measure of cognitive reflection (adjusted r-squareds: Model 1: .0753; Model 2: .0727). Just as we saw when we examined the mean squared error of respondents' forecasts, cognitive reflection is especially predictive of the forecasts made by first time respondents to the SCE.

Model 3 of Table 5 analyses respondent IQRs, which measure forecast uncertainty. The key result from Model 3 is captured by the interaction term, *cognitive reflection *2021*. The non-significant coefficient indicates that the change in uncertainty of forecasts in the last quarter of 2021 was neither smaller nor larger among reflective respondents than it was among less reflective respondents.

Reflective respondents continued to be more precise in their forecasts than less reflective respondents; the IQRs returned by reflective respondents in the final quarter of 2021 were substantially smaller than those returned by less reflective respondents who answered QNUM2 with the intuitive but incorrect response of $240 ($IQR_{reflect} = 4.3$ vs. $IQR_{arithmeticallycompetent} = 6.5$; $t = 12.42$, $p < .001$). The key result of Model 3 is that, after accounting for time trends and other observables, those high in cognitive reflection did not become especially confident (or lacking in certainty) in late 2021.

Additional results from Model 3 that are instructive concern the role of the other cognitive characteristics of respondents. The positive coefficient on *Financial literacy*2021* suggests that in the high inflation period those high in financial literacy became more uncertain in their forecasts than did otherwise similar respondents. The negative coefficient on *Arithmetic score*2021* result suggests that in the high inflation period those high in arithmetic ability became more certain (or less uncertain) in their forecasts than did otherwise similar respondents. Table A4 in the online appendix shows that all the significant directional results reported in this paragraph and in the earlier paragraph that discussed Model 1 replicate when we reduce the sample to just those respondents who were answering the SCE for the first time.

A final result concerns the alternative measures of inflation expectations, the mean of the forecast density (FDmean) and respondents' point estimates when asked "What do you expect the rate of [inflation/deflation] to be over the next 12 months?". Table 6 reports the results from OLS regressions that apply the same specification as in Model 1 of Table 5 to each of these measures.

Table 6 shows no significant effect of the *cognitive reflection*2021* interaction on either measure of inflation expectations (FDmean: $b = -1.00$, $t = 0.52$, $p = .605$; point estimate: $b = -0.38$, $t = 0.18$, $p = .853$). This null result is at odds with the results presented in Table 5. This discrepancy might be explained by noise in the dependent variables reported in Table 6. As highlighted in Section 3.3.1, this noise makes it difficult to discern reliable patterns of updating across types of respondent. A graphic illustration of how noisy measures would give a misleading test of our hypothesis comes from looking at the coefficient on financial literacy. In the model of inflation expectations where FDmean is the dependent variable, financial literacy is a significant negative predictor ($b = -0.153$, $t = 2.41$, $p = .016$). But the same variable has a significant positive coefficient in the model

Table 5
Cognitive reflection predicts updates to forecasts and IQRs.

	Forecast	Forecast	IQR
Reflect	−6.326**		−1.044**
	(0.694)		(0.090)
Reflect*2021	3.000*		−0.125
	(1.272)		(0.164)
Numeracy (incl. QNUM2)		−3.996**	
		(0.304)	
Numeracy*2021		1.903**	
		(0.564)	
2021	−1.145	−2.485	−0.292
	(2.911)	(2.566)	(0.375)
Financial literacy	−1.711**	−1.756**	0.494**
	(0.415)	(0.415)	(0.054)
Financial literacy*2021	9.357**	9.403**	1.071**
	(1.274)	(1.274)	(0.164)
Arithmetic score	−2.865**		−1.102**
	(0.432)		(0.056)
Arithmetic score*2021	1.373		−0.210*
	(0.790)		(0.102)
Time trend	0.141**	0.122**	0.046**
	(0.023)	(0.020)	(0.003)
Reflect*time trend	0.024*		0.003*
	(0.011)		(0.001)
Financial literacy*time trend	−0.001	−0.000	−0.025**
	(0.008)	(0.008)	(0.001)
Arithmetic score*time trend	0.000		0.002*
	(0.007)		(0.001)
Numeracy*time trend		0.008	
		(0.005)	
Female	−5.199**	−5.313**	−0.752**
	(0.223)	(0.222)	(0.029)
Age	0.128**	0.128**	−0.010**
	(0.007)	(0.007)	(0.001)
Highschool or less	2.401**	2.468**	1.295**
	(0.356)	(0.356)	(0.046)
$50–100k	−3.496**	−3.554**	−0.307**
	(0.269)	(0.268)	(0.035)
> $100k	3.696**	3.687**	0.965**
	(0.258)	(0.258)	(0.033)
Month Year dummies	X	X	X
Area dummies	X	X	X
Constant	41.467**	50.216**	8.849**
	(2.302)	(1.833)	(0.297)
R^2	0.1142	0.1138	0.21
N	100,352	100,352	100,352

* $p<0.05$; ** $p<0.01$, standard errors in parentheses. Models 1 and 2 present OLS regressions of respondents' reported probability that inflation exceeds 4%. Model 3 presents an OLS regression of the interquartile ranges of respondents' forecast densities for inflation a year hence. Table A3 replicates the results looking only at respondent's first wave of SCE data to take account of panel conditioning (Binder and Kim, 2022). Data range May 2015 to December 2021.

Table 6
Tests of alternative measures of inflation expectations.

	FDmean	Point estimate
Cognitive reflection	−0.494**	−0.383
	(0.106)	(1.002)
Cognitive reflection *2021	−0.100	−0.339
	(0.194)	(1.835)
2021	1.585**	6.129
	(0.444)	(4.197)
Financial literacy	−0.153*	1.217*
	(0.063)	(0.599)
Financial literacy*2021	0.826**	0.874
	(0.194)	(1.837)
Arithmetic score	0.037	−1.042
	(0.066)	(0.622)
Arithmetic score*2021	−0.190	−1.251
	(0.121)	(1.140)
Time trend	0.023**	0.116**
	(0.004)	(0.033)
Cognitive reflection*time trend	−0.000	−0.023
	(0.002)	(0.015)
Financial literacy*time trend	0.001	−0.050**
	(0.001)	(0.011)
Arithmetic score*time trend	−0.002*	0.006
	(0.001)	(0.009)
Female	−0.543**	−0.624
	(0.034)	(0.321)
Age	0.017**	0.011
	(0.001)	(0.010)
Highschool or less	−0.005	2.666**
	(0.054)	(0.514)
\$50–100k	−0.340**	−0.682
	(0.041)	(0.387)
> \$100k	0.352**	1.902**
	(0.039)	(0.373)
Month year dummies	X	X
Area dummies	X	X
Constant	3.401**	4.453
	(0.351)	(3.316)
R^2	0.07	0.01
N	100,129	100,129

* p<0.05; ** p<0.01, standard errors in parentheses. Model 1 presents an OLS regression of the mean of respondents' forecast densities. Model 2 presents an OLS regression of the point estimate given by respondents. Data range May 2015 to December 2021.

where inflation expectations are measured by the point estimate procedure ($b = 1.22$, $t = 2.03$, $p = .042$). In other words, one measure of inflation expectations suggests that those high in financial literacy expect significantly higher inflation than others and another measure of the same inflation expectations suggests those same respondents expect significantly lower inflation than others. Since these two measures purport to be capturing the same beliefs from the same respondents, it can only be that at least one of these measures has delivered a statistically significant result that is misleading.

3.4. Exploratory analysis - Can we find a subsample who reliably deliver accurate forecasts?

One motivation for this paper is to test the validity of survey measures of inflation expectations as a leading indicator. Whereas the results just presented call into question the validity of inflation expectations returned by the sample as a whole, the results reported in Tables 3 and 4 suggest that respondents high in financial literacy, arithmetic ability and cognitive reflection might be especially accurate. Perhaps this subsample of respondents deliver inflation expectations that can serve as an especially useful leading indicator?

Table 7
OLS regressions of bias and error in forecasts made by those high in cognitive capabilities (left) and the full sample (right).

	High capability subsample		Complete sample	
	Bias	MSE	Bias	MSE
Month year dummies	X	X	X	X
Constant	0.414	0.388	−1.799**	2.517**
	(0.319)	(0.234)	(0.146)	(0.078)
R^2	0.11	0.15	0.04	0.07
N	34,477	34,477	93,429	93,429

Notes: * p<0.05; ** p<0.01 Results of an OLS regression of Mean Squared Errors of respondent's inflation expectations, where errors are calculated by subtracting the mean of respondent's density forecast (FDmean) from the average seasonally-adjusted median CPI over the 12 months following the survey. The analyses uses data from SCE waves from May 2015 through to June 2021.

This exploratory analysis tests the accuracy of the inflation forecasts made by those individuals who score highest in all of financial literacy, arithmetic ability and cognitive reflection. Just over a quarter of the sample (28%) answered all of the questions in Table 1 correctly and those respondents constitute the subsample for the results reported in the two leftmost columns of Table 7.

Here I use FDmean as the measure of inflation expectations. I do so because the issues I earlier highlighted with the FDmean – selective non-response and misreporting of beliefs – are of no consequence for answering the current research question, which is whether survey expectations are a biased a leading indicator. As pointed out in Section 2.1, a representative sample is not required of a leading indicator. So it does not matter for this use case that there is selective non-response to the forecast density procedure. Also, these respondents who answered all the questions in Table 1 correctly are best equipped to translate their inflationary beliefs onto a percentage scale and so are the least likely to misreport their beliefs by the forecast density procedure.

Model 1 runs an OLS regression of bias controlling only for the month in which data was collected, where bias is measured as:

$$\text{Bias}_{i,t} = \text{FDmean}_{i,t} - \text{MeanCPI}_{ttot+11}$$

The key variable to look at is the constant. Among the high cognitive capacity subsample, bias is not significantly different from zero ($b = 0.41$, $t = 1.30$, $p = .195$). The rightmost columns present the same analyses for the same time period from the sample as a whole. It shows significant bias ($b = -1.80$, $t = 14.00$, $p < .001$).

It is a necessary condition for the operation of the wisdom of the crowd that estimates are not biased. If the wisdom of the crowd operates among this high cognition group then we should additionally see that Mean Squared Error, a measure of forecast noise, is also low. In Model 2 the dependent variable is Mean Squared Error, which is calculated as it was in Table 3:

$$\text{MSE}_{i,t} = [\text{FDmean}_{i,t} - \text{MeanCPI}_{ttot+11}]_2.$$

Model 2 shows that among those high in all cognitive capacities, MSE is not significantly different from zero ($b = 0.388$, $t = 1.66$, $p = .098$).

4. Discussion

4.1. Summary of results and limitations

We first saw that the Survey of Consumer Expectations includes a measure that captures cognitive reflection. That measure is useful because cognitive reflection is a characteristic that predicts accurate forecasting in general (Mellers et al., 2015). The first insight gained from our investigation of the SCE data is that respondents higher in cognitive reflection made lower and less error-prone inflation forecasts than did the full sample. We then focussed specifically on the uptick in inflation expectations that occurred in 2021 and asked was that uptick disproportionately large among respondents low in cognitive reflection, who are prone to overreacting, or among those high in cognitive reflection, whose forecasts are generally more accurate? None of tests above suggest that the uptick was driven by those most likely to fall victim to extrapolation bias, those low in cognitive reflection.

One limitation of these data is that they are not longitudinal. Ideally, we would observe the same individuals throughout the sample period so we could quantify how they update their beliefs. Given the constraints of SCE recruitment and retention process, that was not possible. Instead, I relied on repeated cross-sections for my inference. Reliance on repeated cross-sections will add noise to the data as the sample composition from one month to the next fluctuates. Still, the SCE recruits a representative sample of over 1000 US residents and so, by design, there should be nothing systematic that changes about sample composition. To the extent that wave-on-wave fluctuations in sample composition are a source of noise, some of that noise is controlled for in Tables 4, 5 and 6 by including demographic characteristics as regressors.

A related point is that survey respondents become more attentive to inflation with experience of answering the survey (Kim and Binder, 2022) i.e. after answering a wave of the SCE, respondents cease to form their inflation expectations in the same manner as the population from which they were randomly drawn. In principle, it could be then that cognitive reflection is only predictive of inflation expectations in a sample of experienced respondents but is not predictive of inflation expectations in the wider population. Tables A3 and A4 in the appendix report detailed results on this point. They restrict the sample to just those respondents answering the SCE for the first time. Even in those much smaller datasets, the same results reliably obtain. Indeed, when we look to the gain in r-squareds, the predictive power of cognitive reflection is even greater in the subsample of respondents who are answering the survey for the first time. To the extent that survey-induced learning biases the estimates in these analyses, all the available evidence suggests that bias is to *underestimate* the role of cognitive reflection in the formation of inflation expectations among the general public.

A second limitation of these data is that they come from a relatively short time-period, less than a decade. As such, readers should interpret with caution the finding that there was no systematic error in forecasts made by respondents who score highest in all of arithmetic ability, financial literacy and cognitive reflection (Table 7). It remains to be seen whether future inflation can be accurately predicted by looking to the responses of this subgroup of respondents.

A final limitation of these data is that the measure of inflation expectations is noisy. To the question "whose inflation expectations spiked in 2021?" one response suggested by these results is "it depends which measure of inflation expectations you look to". Notwithstanding the inconsistencies in results across Tables 5 and 6, there is no evidence at all to suggest that the uptick in inflation expectations in 2021 was concentrated among those low in cognitive reflection.

4.2. Implications for forecasting more generally

These results confirm in inflation data that cognitive capabilities are conducive to forecast accuracy. That result is not surprising but embedded within it are some novel contributions that have implications for forecasting in domains beyond inflation. First, cognitive reflection is meaningfully captured by a single question from the SCE, QNUM2. Since cognitive reflection is an especially useful trait for forecasting and it costs little to add QNUM2 to surveys, a recommendation is that forecasting surveys include this question.

Second, we can use existing household surveys to identify subsets of respondents whose forecasts are likely to be especially powerful leading indicators (Tetlock & Gardner, 2016). Taking the SCE as an example, it includes other measures beyond cognitive capabilities that are likely to predict forecast accuracy e.g. how many previous waves of the survey the respondent has answered. I look forward to future research that further tests the accuracy of forecasts made by the group identified in this research as high in cognitive capacities.

Inflation expectations are exceptionally influential – they are a leading indicator that informs monetary policies. Given their policy influence – and hence their impact on the everyday lives of the population – it is important to reduce forecast error. The overall contribution of the paper is to show that we can better predict inflation by focusing on the forecasts of certain respondents e.g. those high in cognitive reflection.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data and code availability

The data and syntax files are available here https://osf.io/bpcjm/.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2024.06.011.

References

Alós-Ferrer, C., & Hügelschäfer, S. (2016). Faith in intuition and cognitive reflection. *Journal of behavioral and experimental economics*, *64*, 61–70.

Armantier, O., Bruine de Bruin, W., Topa, G., Van Der Klaauw, W., & Zafar, B. (2015). Inflation expectations and behavior: Do survey respondents act on their beliefs? *International Economic Review*, *56*(2), 505–536.

Armantier, O., Topa, G., Van der Klaauw, W., & Zafar, B. (2017). An overview of the survey of consumer expectations. *Economic Policy Review*, *23*(2), 51–72.

Barahona, R., Cassella, S., & Jansen, K. A. (2021). *Do teams alleviate or exacerbate behavioral biases? Evidence from extrapolation bias in mutual funds*: *Working paper*.

Becker, C. K., Duersch, P., & Eife, T. (2023). Measuring inflation expectations: How the response scale shapes density forecasts. Available at SSRN 4323706.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2018). Diagnostic expectations and credit cycles. *The Journal of Finance*, *73*(1), 199–227.

Bruine de Bruin, W., & Fischhoff, B. (2017). Eliciting probabilistic expectations: Collaborations between psychologists and economists. *Proceedings of the National Academy of Sciences*, *114*(13), 3297–3304.

Bruine de Bruin, W., Vanderklaauw, 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.

Budescu, D. V., Himmelstein, M., & Ho, E. (2021). Boosting the wisdom of crowds with social forecasts and coherence measures. In *Presented at annual meeting of Society of Multivariate Experimental Psychology (SMEP), Online, October 2021*.

Burke, M. A., & Manz, M. (2014). Economic literacy and inflation expectations: evidence from a laboratory experiment. *Journal of Money, Credit and Banking*, *46*(7), 1421–1456.

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

Comerford, D. A. (2019). Asking for frequencies rather than percentages increases the validity of subjective probability measures: Evidence from subjective life expectancy. *Economics letters*, *180*, 33–35.

Comerford, D. A. (2021). Apparent age and gender differences in survival optimism: To what extent are they a bias in the translation of beliefs onto a percentage scale? *Judgment and Decision Making*, *16*(4), 1072–1096.

Comerford, D. A. (2023). Response bias in survey measures of expectations: Evidence from the survey of consumer expectations' inflation module. *Journal of Money, Credit and Banking*, http://dx.doi.org/10.1111/jmcb.13003.

Comerford, D. A., & Robinson, J. (2017). Die-by framing both lengthens and shortens life: Further evidence on constructed beliefs in life expectancy. *Journal of Behavioral Decision Making*, *30*(5), 1104–1112.

Da, Z., Huang, X., & Jin, L. J. (2021). Extrapolative beliefs in the cross-section: What can we learn from the crowds? *Journal of Financial Economics*, *140*(1), 175–196.

D'Acunto, F., Hoang, D., Paloviita, M., & Weber, M. (2019). Cognitive abilities and inflation expectations. *Vol. 109*, In *AEA papers and proceedings* (pp. 562–566).

D'Acunto, F., Hoang, D., Paloviita, M., & Weber, M. (2023). IQ, expectations, and choice. *Review of Economic Studies*, *90*(5), 2292–2325.

D'Acunto, F., Hoang, D., & Weber, M. (2022). Managing households' expectations with unconventional policies. *The Review of Financial Studies*, *35*(4), 1597–1642.

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

D'Acunto, F., Malmendier, U., Ospina, J., & Weber, M. (2022). Managing households' expectations with unconventional policies. *The Review of Financial Studies*, *35*(4), 1597–1642.

D'Acunto, F., Malmendier, U., & Weber, M. (2021). Gender roles produce divergent economic expectations. *Proceedings of the National Academy of Sciences*, *118*(21), Article e2008534118.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ... Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46.

Drerup, T., Enke, B., & Von Gaudecker, H. M. (2017). The precision of subjective data and the explanatory power of economic models. *Journal of Econometrics*, *200*(2), 378–389.

Fischhoff, B., & Bruine De Bruin, W. (1999). Fifty–fifty=50%? *Journal of Behavioral Decision Making*, *12*(2), 149–163.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42.

Himmelstein, M., Atanasov, P., & Budescu, D. V. (2021). Forecasting forecaster accuracy: Contributions of past performance and individual differences. *Judgment and Decision Making*, *16*(2), 323–362.

Hirshleifer, D., Li, J., & Yu, J. (2015). Asset pricing in production economies with extrapolative expectations. *Journal of Monetary Economics*, *76*, 87–106.

Hoppe, E. I., & Kusterer, D. J. (2011). Behavioral biases and cognitive reflection. *Economics Letters*, *110*(2), 97–100.

Kim, G., & Binder, C. (2023). Learning-through-survey in inflation expectations. *American Economic Journal: Macroeconomics*, *15*(2), 254–278.

Lane, P. (2022). Monetary policy in the euro area: the next phase. Remarks for high-level panel high inflation and other challenges for monetary policy. In *Annual meeting 2022 of the central bank research association*. Barcelona.

Malmendier, U., & Nagel, S. (2016). Learning from inflation experiences. *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.

Manski, C. F. (2017). Collaboration, conflict, and disconnect between psychologists and economists. *Proceedings of the National Academy of Sciences*, *114*(13), 3286–3288.

Mehra, Y. P. (2002). Survey measures of expected inflation: revisiting the issues of predictive content and rationality. *FRB Richmond Economic Quarterly*, *88*(3), 17–36.

Mellers, B., Stone, E., Atanasov, P., Rohrbaugh, N., Metz, S. E., Ungar, L., ... Tetlock, P. (2015). The psychology of intelligence analysis: Drivers of prediction accuracy in world politics. *Journal of Experimental Psychology: Applied*, *21*(1), 1.

Meyer, B., Venkatu, G., & Zaman, S. (2013). Forecasting inflation? Target the middle. In *Economic commentary, (2013-05)*.

Niu, X., & Harvey, N. (2023). Are lay expectations of inflation based on recall of specific prices? if so, how and under what conditions? Available at SSRN 4326565.

Pfajfar, D., & Santoro, E. (2010). Heterogeneity, learning and information stickiness in inflation expectations. *Journal of Economic Behavior and Organization*, *75*(3), 426–444.

Sinayev, A., & Peters, E. (2015). Cognitive reflection vs. Calculation in decision making. *Frontiers in Psychology*, *6*(532).

Surowiecki, J. (2005). *The wisdom of crowds*. Anchor.

Tetlock, P. E., & Gardner, D. (2016). *Superforecasting: the art and science of prediction*. Random House.

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, *11*(1), 99–113.

Vartanian, O., Beatty, E. L., Smith, I., Blackler, K., Lam, Q., Forbes, S., & De Neys, W. (2018). The reflective mind: Examining individual differences in susceptibility to base rate neglect with fMRI. *Journal of Cognitive Neuroscience*, *30*(7), 1011–1022.

Verbrugge, R., & Zaman, S. (2024). Improving inflation forecasts using robust measures. *International Journal of Forecasting*, *40*(2), 735–745.

Yellen, J. (2018). *A fed duet: janet yellen in conversation with ben bernanke*. The Brookings Institution.