

**Judgmental forecasting: Factors affecting lay people's expectations of  
inflation**

by

**Xiaoxiao Niu**

A dissertation submitted for the degree of

**Doctor of Philosophy (PhD)**

of

**University College London**

Department of Experimental Psychology

Faculty of Brain Sciences

February 2023

## **Abstract**

In this thesis, laypeople's judgmental forecasting about inflation is reviewed and experimentally explored in six chapters. Inflation is defined as the Consumer Price Index (CPI) across the whole thesis. In Chapter 1, I review work on the formation of inflation expectations, drawing mainly from the economic literature. In Chapter 2, I review research on judgmental forecasting, drawing mainly from the literature in cognitive psychology and management science. In Chapter 3, three experiments are presented that were designed to determine how and when people employ internal information of experienced price changes to form inflation expectations. In Chapter 4, three experiments are used to investigate the effects of providing within-series and across-series historical information (inflation rates, interest rates and unemployment rates) on inflation expectations. In Chapter 5, two experiments are reported that examine how training using simple outcome feedback increases the accuracy of inflation judgments and improves the calibration of confidence in those judgments. Chapter 6 reports experiments designed to examine the effects of using different elicitation methods (point forecasts, interval forecasts and density forecasts) on the accuracy of inflation judgments. Chapter 7 is a concluding chapter that summarises findings from these experiments and suggests avenues for future work.

## **Impact statement**

After finding that lay inflation expectations are influenced by various factors, including inflation environment, the use of information, survey measurements, the provision of outcome feedback, and elicitation methods, this thesis benefits the understanding of the mental model of inflation forecasts of lay people. Also, it offers some approaches to help to improve the accuracy level of inflation judgments of lay people, which contributes to research areas of judgment and decision making, psychology and economics.

For laypeople themselves, this thesis demonstrates when and why their inflation judgments are inaccurate under certain circumstances. By knowing this, lay persons are encouraged to and are able to conscientiously de-bias their expectations and adjust their consequent decision making.

The research could inform national and international guidelines for governments and central banks on how to design an appropriate consumer survey in order to obtain more accurate inflation judgments. This could have profound advantages for the development of the economy. Specifically, the findings provide a framework for selecting or designing new consumer survey and policy options based on the changes of inflation environment, intentions to understand expectations and/or motivations to change behaviour.

By revealing the biases (over-estimation and over-confidence) existing in the inflation expectation formation process, it could also raise awareness among journalists and policy makers about the importance of their special roles. They are able to use media or other channels to improve how well lay people are informed to help them make better financial decisions. Some new government interventions stimulated by current findings could reduce judgment biases and educate lay people in a timely and efficient manner.

This is the first study to directly demonstrate how inflation expectation accuracy can be improved through methods like training. The conclusions of this thesis could provide benefits in the context of inflation expectations of lay people and also be generalized to other fields in which judgmental forecasting has a unique role, such as business forecasting (for planning) and weather forecasting (for warnings).

## List of Contents

Abstract.....	1
Impact statement .....	2
List of Contents.....	3
Author's Declaration .....	8
Acknowledgments .....	9
Publications arising directly from this submission .....	10
UCL Research Paper Declaration Form.....	11
List of Figures .....	16
List of Tables .....	18
List of Appendices.....	19
Part 1 Review .....	20
Introduction .....	20
Chapter 1 Review of inflation expectations .....	21
1.1. Background .....	21
1.2. Economic assumptions about inflation expectations.....	23
1.3. Characteristics of inflation expectations .....	25
1.3.1 Lay inflation expectations .....	25
1.3.2. Experts' inflation expectations .....	26
1.4. Explanations for overestimation and homogeneity of inflation expectations.....	27
1.4.1. Demographical factors.....	27
1.4.2. Question interpretation.....	28
1.4.3. Purchasing experience .....	29
1.4.4. Financial literacy (FL) .....	31
1.4.5. The role of information.....	32
1.4.6. The use of information .....	33
1.4.7. Expectation-consistent bias.....	35
1.4.8. Other factors.....	36
1.5. Confidence (uncertainty) in judgments .....	36
1.6. A model from Ranyard et al. (2017) .....	38
1.6.1. Direct experience.....	40
1.6.2. Official statistics .....	41
1.6.3. News media .....	42
1.6.4. Interactions between perceptions and expectations.....	44
1.7. Psychological mechanisms underlying inflation judgments.....	45

1.7.1. Cognitive processes .....	45
1.7.2. Mental representations of inflation .....	46
1.7.3. Perception, evaluation and updating .....	51
1.7.4. Retrieval .....	55
1.8. Decision making .....	59
Chapter 2 Review of judgmental forecasting .....	62
2.1. Background .....	62
2.2. Economic forecasting.....	63
2.3. Heuristics and Biases .....	64
2.3.1. Representativeness.....	65
2.3.2. Availability.....	66
2.3.3. Anchoring.....	66
2.4. Judgmental time series forecasting.....	67
2.4.1. Ecological validity.....	68
2.4.2. Trend.....	69
2.4.3. Other characteristics of data series.....	70
2.4.4. Data presentation format.....	71
2.5. Contextual information .....	74
2.5.1. Rule-based forecasting .....	74
2.5.2. Information advantage .....	75
2.5.3. Information overload.....	77
2.5.4. Pertinent information .....	77
2.5.5. Technical knowledge .....	78
2.5.6. Causal information.....	79
2.5.7. Domain knowledge .....	80
2.5.8. Experts .....	81
2.5.9. Simple Heuristics.....	82
2.5.10. Presentation format .....	83
2.6. Forecasting under uncertainty .....	84
2.7. Interval forecasting.....	85
2.7.1. Effects of interval width.....	85
2.7.2. Effects of providing interval information on accuracy .....	86
2.7.3. Factors influencing interval forecasts.....	87
2.7.4. Interval construction.....	91
2.7.5. Overconfidence in interval forecasts.....	95
2.7.6. Heuristics and biases .....	98
2.8. Probability density forecasts .....	103
2.8.1. Factors affecting density forecasts.....	106

2.8.2. Density forecast construction.....	108
2.9. Improving judgment .....	109
2.9.1. Learning and education .....	109
2.9.2. Judgmental adjustment .....	110
2.9.3. Combining forecasts .....	110
2.9.4. Group judgment.....	110
2.9.5. Feedback and advice.....	111
2.9.6. Decomposition .....	112
2.10. Techniques used to improve realism of confidence judgments .....	113
2.10.1. Feedback .....	113
2.10.2. Elicitation question .....	114
Part 2 Experimental chapters .....	116
Introduction .....	116
Overview of experiments .....	116
Chapter 3 How lay people form expectations about inflation .....	117
3.1. Introduction .....	117
3.1.1. Three models of the development of expectations of inflation .....	118
3.1.2. Review of previous work .....	118
3.2. Experiment 1.....	120
3.2.1 Method .....	121
3.2.2. Results.....	124
3.2.3. Discussion .....	126
3.3. Experiment 2a and 2b.....	127
3.3.1. Method .....	129
3.3.2. Results of Experiment 2a .....	131
3.3.3. Results of Experiment 2b.....	133
3.3.4. Discussion .....	135
3.4. General Discussion.....	137
3.4.1. Limitations .....	139
3.4.2. Implications.....	139
3.4.3. Conclusion.....	140
Chapter 4. How lay expectations are measured in surveys .....	141
4.1. Introduction .....	141
4.1.1. Effects of survey format .....	142
4.1.2. Information context: Differences between surveys of lay and expert forecasters .....	142
4.1.3. Judgment heuristics in forecasting: The nature of the information available .....	143
4.2. Experiment 3.....	145
4.2.1 Method .....	147

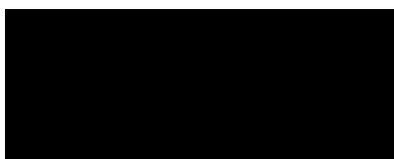
4.2.2. Results.....	149
4.2.3. Discussion .....	152
4.3. Experiment 4.....	152
4.3.1 Method .....	153
4.3.2 Results.....	155
4.3.3 Discussion .....	157
4.4 Experiment 5.....	158
4.4.1. Method .....	158
4.4.2. Results.....	160
4.4.3. Discussion .....	162
4.5. General discussion .....	163
4.5.1. Potential limitations.....	164
4.5.2. Implications.....	165
4.5.3. Conclusions .....	166
Chapter 5 How does feedback improve measurement of inflation expectations .....	167
5.1. Introduction .....	167
5.1.1. Effects of outcome feedback on forecast accuracy.....	167
5.1.2. Effects of outcome feedback on confidence in forecasts .....	169
5.1.3. Identifying the nature of outcome feedback effects.....	172
5.2. Experiment 6.....	173
5.2.1. Method .....	174
5.2.2. Results.....	175
5.2.3. Discussion .....	180
5.3. Experiment 7.....	181
5.3.1. Method .....	182
5.3.2. Results.....	183
5.3.3. Discussion .....	188
5.4. General discussion .....	189
5.4.1. Inflation forecasts .....	189
5.4.2. Confidence in forecasts .....	191
5.4.3. Implications.....	192
5.4.4. Limitations and suggestions for future work.....	193
5.4.5. Conclusions .....	194
Chapter 6 How do elicitation methods improve measurement of inflation expectations .....	195
6.1. Introduction .....	195
6.2. Experiment 8.....	196
6.2.1. Method .....	198
6.2.2. Results.....	201

6.2.3. Discussion .....	203
6.3. Experiment 9.....	204
6.3.1. Method .....	205
6.3.2. Results.....	207
6.3.3. Discussion .....	210
6.4. Experiment 10.....	211
6.4.1. Method .....	212
6.4.2. Results.....	214
6.4.3. Discussion .....	217
6.5. General Discussion.....	218
6.5.1. Limitations .....	219
6.5.2. Implications.....	220
6.5.3. Conclusions .....	220
Chapter 7 Conclusions .....	222
7.1. Summary of findings .....	222
7.2. Implications.....	223
7.2.1. Implications for improving judgmental forecasting accuracy .....	224
7.2.2. Implications for media guidance .....	225
7.2.3. Implications for governmental policymaking .....	226
7.3. Limitations and future directions .....	227
7.4. Conclusions .....	229
References .....	231
Appendices .....	290
Appendix 1 .....	290
Appendix 2 .....	293
Appendix 3 .....	305
Appendix 4 .....	306
Appendix 5 .....	309
Appendix 6 .....	310
Appendix 7 .....	311

## Author's Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text. Any contribution made to the research by others, with whom I have worked at UCL, is explicitly acknowledged in the thesis. This work has not been submitted for any other degree or professional qualification except as specified.

Signature:

A solid black rectangular box used to redact the author's signature.

Xiaoxiao Niu

December 2022

## Acknowledgments

This thesis would not have been produced without the invaluable opportunity, support and guidance given by my amazing supervisor, Professor Nigel Harvey. I greatly appreciate Nigel for guiding me into the world of research with great interest, wisdom, inspiration, patience, support and encouragement throughout my whole PhD journey in every aspect (research, teaching, working, public engagement, connections, mental health, etc.). His continuous pursuit of innovation, dedicated research attitude, excellence and teaching methodology inspires me deeply in my training as a psychology researcher and as a teacher. He has had a significant influence on my research, my future career and my life. I also obtained unconditional help and care from Nigel's lab throughout my PhD: Dr Shari de Beats and Dr Ariel Cecchi.

I would also thank my second supervisor, Professor David Lagnado. He gave me the opportunity to join his lab meetings and offer academic guidance and help together with Nigel, especially during the first year of my PhD, when I just stepped into the Experimental Psychology department. I also thank Professor Adam Harris for giving me full trust and understanding to balance my study and work.

Special thanks go to current and past researchers and staff in 26 Bedford Way and other UCL departments for their valuable academic teaching, intellectual advice, technical assistance and support and help from other aspects (in alphabetical order): Dr Alice Liefgreen, Miss Alicia Burton, Mrs Antonietta Esposito, Dr Christos Bechlivanidis, Dr Daphne Sobolev, Dr Emily Thomas, Dr Katie Fisher, and Dr Henrik Singmann.

I would not have enjoyed my time at UCL as much without all my dear friends giving me love, support, encouragement and power: Hui Zhong, Yaqi Hu, Kang Kang, Qiu Ting, Pro. Lin Lin, and my lovely tennis buddies.

Finally, I want to thank my parents, and my brothers for being my pillar of strength, and for walking this entire journey with me with unconditional love, patience and financial and mental support.

Everything was more difficult with the occurrence of Covid-19 compared to normal times, but all the people I thanked above made my work and life easier, happier and better. I would express my sincere gratitude to all people around me.

### Publications arising directly from this submission

The work presented in Chapter 4 is based on the following publications:

Niu, X., & Harvey, N. (2022). Context effects in inflation surveys: The influence of additional information and prior questions. *International Journal of Forecasting*, 38 (3), 988-1004. <https://doi.org/10.1016/j.ijforecast.2021.07.009>

The work presented in Chapter 5 is based on the following publication:

Niu, X., & Harvey, N. (2022). Outcome feedback reduces over-forecasting of inflation and overconfidence in forecasts. *Judgment and Decision Making*, 17(1), 124-163. <https://journal.sjdm.org/21/210609a/jdm210609a.pdf>

The work presented in Chapter 6 is based on the following publication:

Niu, X., & Harvey, N. (2022). Point, interval, and density forecasts: Differences in bias, judgment noise, and overall accuracy. *Futures & Foresight Science*, 1-18. <https://doi.org/10.1002/ffo2.124>

## UCL Research Paper Declaration Form

### referencing the doctoral candidate's own published works

Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:

- have been uploaded to a preprint server;
- are in submission to a peer-reviewed publication;
- have been published in a peer-reviewed publication, e.g. journal, textbook.

<b>1. For a research manuscript that has already been published</b> (if not yet published, please skip to section 2):		
<b>a) Where was the work published?</b> (e.g. journal name)	Click or tap here to enter text.	
<b>b) Who published the work?</b> (e.g. Elsevier/Oxford University Press):	Click or tap here to enter text.	
<b>c) When was the work published?</b>	Click or tap to enter a date.	
<b>d) Was the work subject to academic peer review?</b>	Please select.	
<b>e) Have you retained the copyright for the work?</b>	Please select.	
[If no, please seek permission from the relevant publisher and check the box next to the below statement]:		
<input type="checkbox"/> <i>I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.</i>		
<b>2. For a research manuscript prepared for publication but that has not yet been published</b> (if already published, please skip to section 3):		
<b>a) Has the manuscript been uploaded to a preprint server?</b> (e.g. medRxiv):	No	<b>If yes, which server?</b> Click or tap here to enter text.
<b>b) Where is the work intended to be published?</b> (e.g. names of journals that you are planning to submit to)	Journal of Economic Psychology	
<b>c) List the manuscript's authors in the intended authorship order:</b>	Xiaoxiao Niu, Nigel Harvey	
<b>d) Stage of publication</b>	Undergoing revision after peer review	
<b>3. For multi-authored work, please give a statement of contribution covering all authors</b> (if single-author, please skip to section 4):		

<b>Xiaoxiao Niu</b> has the following contributions: Idea formation, Experiment design, Data collection, Data analyses, Writing-up original draft; <b>Nigel Harvey</b> has the following contributions: Idea formation, Supervision, Writing review and editing.			
<b>4. In which chapter(s) of your thesis can this material be found?</b>			
Chapter 3			
<b>5. e-Signatures confirming that the information above is accurate</b> (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):			
<b>Candidate:</b>	Xiaoxiao Niu	<b>Date:</b>	05/12/2022
<b>Supervisor/ Senior Author</b> (where appropriate):	Nigel Harvey	<b>Date:</b>	05/12/2022

<b>1. For a research manuscript that has already been published</b> (if not yet published, please skip to section 2):	
<b>a) Where was the work published?</b> (e.g. journal name)	International Journal of Forecasting
<b>b) Who published the work?</b> (e.g. Elsevier/Oxford University Press):	Elsevier
<b>c) When was the work published?</b>	2022
<b>d) Was the work subject to academic peer review?</b>	Yes
<b>e) Have you retained the copyright for the work?</b>	Yes
[If no, please seek permission from the relevant publisher and check the box next to the below statement]: <input type="checkbox"/> <i>I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.</i>	
<b>2. For a research manuscript prepared for publication but that has not yet been published</b> (if already published, please skip to section 3):	
<b>e) Has the manuscript been uploaded to a preprint server?</b> (e.g. medRxiv):	Please select. <b>If yes, which server?</b> Click or tap here to enter text.
<b>f) Where is the work intended to be published?</b> (e.g. names of journals that you are planning to submit to)	Click or tap here to enter text.

<b>g) List the manuscript's authors in the intended authorship order:</b>		Click or tap here to enter text.	
<b>h) Stage of publication</b>		Please select.	
<b>3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4):</b>			
Xiaoxiao Niu has the following contributions: Idea formation, Experiment design, Data collection, Data analyses, Writing-up original draft; <b>Nigel Harvey</b> has the following contributions: Idea formation, Supervision, Writing review and editing.			
<b>4. In which chapter(s) of your thesis can this material be found?</b>			
Chapter 4			
<b>5. e-Signatures confirming that the information above is accurate</b> (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):			
<b>Candidate:</b>	Xiaoxiao Niu	<b>Date:</b>	05/12/2022
<b>Supervisor/ Senior Author</b> (where appropriate):	Nigel Harvey	<b>Date:</b>	05/12/2022

<b>1. For a research manuscript that has already been published</b> (if not yet published, please skip to section 2):	
<b>a) Where was the work published?</b> (e.g. journal name)	Judgment and Decision Making
<b>b) Who published the work?</b> (e.g. Elsevier/Oxford University Press):	Cambridge University Press
<b>c) When was the work published?</b>	2022
<b>d) Was the work subject to academic peer review?</b>	Yes
<b>e) Have you retained the copyright for the work?</b>	Yes
[If no, please seek permission from the relevant publisher and check the box next to the below statement]:	
<input type="checkbox"/> <i>I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.</i>	
<b>2. For a research manuscript prepared for publication but that has not yet been published</b> (if already published, please skip to section 3):	

<b>a) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv):</b>	Please select.	<b>If yes, which server? Click or tap here to enter text.</b>	
<b>b) Where is the work intended to be published? (e.g. names of journals that you are planning to submit to)</b>	Click or tap here to enter text.		
<b>c) List the manuscript's authors in the intended authorship order:</b>	Click or tap here to enter text.		
<b>d) Stage of publication</b>	Please select.		
<b>3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4):</b>			
Xiaoxiao Niu has the following contributions: Idea formation, Experiment design, Data collection, Data analyses, Writing-up original draft; <b>Nigel Harvey</b> has the following contributions: Idea formation, Supervision, Writing review and editing.			
<b>4. In which chapter(s) of your thesis can this material be found?</b>			
Chapter 5			
<b>5. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):</b>			
<b>Candidate:</b>	Xiaoxiao Niu	<b>Date:</b>	05/12/2022
<b>Supervisor/ Senior Author (where appropriate):</b>	Nigel Harvey	<b>Date:</b>	05/12/2022

<b>1. For a research manuscript that has already been published (if not yet published, please skip to section 2):</b>	
<b>a) Where was the work published? (e.g. journal name)</b>	Futures & Foresight Science
<b>b) Who published the work? (e.g. Elsevier/Oxford University Press):</b>	Wiley
<b>c) When was the work published?</b>	2022
<b>d) Was the work subject to academic peer review?</b>	Yes

<b>e) Have you retained the copyright for the work?</b>		Yes	
[If no, please seek permission from the relevant publisher and check the box next to the below statement]:			
<input type="checkbox"/> <i>I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.</i>			
<b>2. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3):</b>			
<b>a) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv):</b>		Please select.	<b>If yes, which server? Click or tap here to enter text.</b>
<b>b) Where is the work intended to be published? (e.g. names of journals that you are planning to submit to)</b>		Click or tap here to enter text.	
<b>c) List the manuscript's authors in the intended authorship order:</b>		Click or tap here to enter text.	
<b>d) Stage of publication</b>		Please select.	
<b>3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4):</b>			
Xiaoxiao Niu has the following contributions: Idea formation, Experiment design, Data collection, Data analyses, Writing-up original draft; <b>Nigel Harvey</b> has the following contributions: Idea formation, Supervision, Writing review and editing.			
<b>4. In which chapter(s) of your thesis can this material be found?</b>			
Chapter 6			
<b>5. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):</b>			
<b>Candidate:</b>	Xiaoxiao Niu	<b>Date:</b>	05/12/2022
<b>Supervisor/ Senior Author</b> (where appropriate):	Nigel Harvey	<b>Date:</b>	05/12/2022

## List of Figures

Figure 1.1	Chapter 1: Sources of information for inflation perceptions and expectations (extracted from Ranyard et al., 2017)	39
Figure 1.2	Chapter 1: Information contributing to laypeople's inflation expectations	40
Figure 1.3	Chapter 1: An intuitive model involved in inflation expectation formation	46
Figure 3.1	Chapter 3: Experiment 1: Effects of task type on constant error and absolute error in judgments of current and expected inflation	125
Figure 3.2	Chapter 3: Experiment 2a: Effects of task type on the constant errors and absolute errors of the direct and indirect inflation forecasts in each of the four conditions	132
Figure 3.3	Chapter 3: Experiment 2b: Effects of task type on the constant errors and absolute errors of the direct and indirect inflation forecasts in each of the four conditions	134
Figure 4.1	Chapter 4: Experiment 3: Summary task instructions followed by examples of tables ready for a) entry of the first inflation judgment in group 2 (upper panel) and b) entry of the fourth inflation judgment in group in group 3 (lower panel)	148
Figure 4.2	Chapter 4: Experiment 3: Interaction between Contextual Information and Judgment Number in the analysis of absolute error (together with standard error bars)	151
Figure 4.3	Chapter 4: Experiment 4: Response tables ready for a) entry of the first inflation judgment in group 1 (upper panel) and b) entry of the fourth inflation judgment in group in group 3 (lower panel)	154
Figure 4.4	Chapter 4: Experiment 4: Effects of Task and Judgment Number on absolute error (together with standard error bars)	157
Figure 4.5	Chapter 4: Experiment 5: Response tables for inflation judgments in a) group without information provided (upper panel) and b) group with the unemployment rate provided (lower panel)	159
Figure 4.6	Chapter 4: Experiment 5: Effects of Year and Information Type on absolute error (together with standard error bars)	161
Figure 5.1	Chapter 5: A table showing annual inflation figures between 2009 and 2018 and an empty cell for participants to enter their inflation estimate for 2019	175
Figure 5.2	Chapter 5: Experiment 6: Graph of mean confidence level (with standard error bars) in the two sessions for each group	179
Figure 5.3	Chapter 5: Experiment 6: Bias plotted against forecast difficulty in the no-feedback group (left panel) and the feedback group (right panel)	180

Figure 5.4	Chapter 5: Experiment 7: Mean Judgment (upper panel), Absolute Error (second panel), Constant Error (third panel), and Variable Error (lower panel) in each group, all shown with standard error bars	185
Figure 5.5	Chapter 5: Experiment 7: Mean Confidence level (upper panel), Bias (middle panel), and Calibration score (lower panel) in each group, all shown with standard error bars	187
Figure 5.6	Chapter 5: Experiment 7: Bias plotted against forecast difficulty in the no-feedback group (left panel) and the feedback groups (right panel)	188
Figure 6.1	Chapter 6: Experiment 8: Time series showing 20 years of annual inflation figures from Country 1	199
Figure 6.2	Chapter 6: Experiment 8: Examples of point forecasting (upper panel), interval forecasting (middle panel), and density forecasting (lower panel)	200
Figure 6.3	Chapter 6: Experiment 8: Bar chart showing ME, VE, and RMSE error scores (with standard error bars) in each forecasting task	202
Figure 6.4	Chapter 6: Experiment 9: Bar charts showing ME, VE, and RMSE scores (with standard error bars) in each condition	208
Figure 6.5	Chapter 6: Experiment 9: Bar charts showing mean confidence levels, bias levels and calibration scores (with standard error bars) in each condition	209
Figure 6.6	Chapter 6: Experiment 10. The 20 inflation series each showing 20 years of historical inflation data between 1998 and 2017 and ranging between -6% and 10%	213
Figure 6.7	Chapter 6: Experiment 10: ME, VE, and RMSE scores for inflation forecasts made from low and high volatility series	216

## List of Tables

Table 4.1	Chapter 4: Experiment 3: Means and standard deviations (in parentheses) of inflation judgments, their directional errors, and their absolute errors	150
Table 4.2	Chapter 4: Experiment 5: Means and standard deviations (in parentheses) of inflation judgments, their directional errors, and their absolute errors	156
Table 4.3	Chapter 4: Experiment 5: Means and standard deviations (in parentheses) of inflation judgments, their directional errors, and their absolute errors	160
Table 5.1	Chapter 5: Experiment 6: Means and standard deviations (in parentheses) of inflation judgments, their total errors, constant errors and variable errors	176
Table 5.2	Chapter 5: Experiment 7: Means and standard deviations (in parentheses) of confidence judgments, their bias scores, and calibration scores	178

## List of Appendices

Appendix 1	Chapter 3: Graphic illustrations for three experiments.	290
Appendix 2	Chapter 3: Additional results and discussion for three experiments.	293
Appendix 3	Chapter 4: Demographical statistics for participants whose data were analysed in three experiments (percentages or standard deviations in parentheses).	305
Appendix 4	Chapter 5: Means and standard deviations (in parentheses) of inflation judgments and confidence judgments of two experiments.	306
Appendix 5	Chapter 5: Comparisons of error scores in each experiment with two simple algorithmic models.	309
Appendix 6	Chapter 6: Uncertainty estimation in interval and density forecasting.	310
Appendix 7	Chapter 6: A summary table of the hypotheses and results in three experiments.	311

## **Part 1 Review**

### **Introduction**

This thesis has its main substance the examination of factors affecting lay people's expectations of inflation. In practice, consumer surveys are commonly employed in order to obtain lay people's inflation expectations across countries. For each country, it is of great importance of monitoring perceived and expected price changes and conducting monetary policies to control the actual inflation rate. In Chapter 1, the observed phenomena of their survey responses, potential explanations as well as existing psychological models and theories will be reviewed. It will highlight the significance of the understanding the psychological mechanism underlying inflation expectation of lay people.

Although economists usually acquire massive amounts of macro data and construct complex macroeconomic models to predict future inflation, judgmental forecasting, introduced in Chapter 2, is the fundamental forecasting process of future inflation rate among lay people. It is used here to ascertain the factors affecting lay people's formation of inflation judgments. Specifically, different types of judgmental forecasting such as time series extrapolation, contextual information utilization, judgments with uncertainty will be discussed. In addition, Chapter 2 will describe influential characteristics of and improvement techniques for both people's forecasts and their confidence judgments in their forecasts.

## Chapter 1 Review of inflation expectations

### 1.1. Background

Prices of goods and services change upwards or downwards over time. Movements are usually upwards resulting in inflation. Inflation rate (measured via the Consumer Price Index, CPI) is defined as the proportional increase in nominal value of a given set of goods and services over a year.

Judgments and expectations related to inflation rates are important for economic entities and individuals. This is because having accurate inflation judgments and inflation expectations benefits personal finance and the development of the country's economy. It is well-established that price perceptions and price expectations influence individuals' financial behaviours (D'Acunto et al., 2015; Nyamekye & Poku, 2017). Indeed, economists assume that financial decisions concerning such matters as savings, investments, purchasing durable goods, and wage negotiations are all made with implicit or explicit assumptions about future inflation. Moreover, accurate measurements of the public's beliefs of inflation expectations are important for scholars and policymakers because the behaviours of households in the aggregate are an important driver of economic activities.

As a result, macroeconomists increasingly use outside estimates of inflation expectations as inputs to their models, such as the economic theory of rational expectations (Maddock & Carter, 1982). In order to achieve a relatively low and steady inflation rate, central banks and policymakers need accurate measures of inflation expectation to calibrate monetary policy efficiently (Cunningham et al., 2010; Galí, 2015). They regularly compare the public's beliefs about inflation with their policy objectives to monitor the effectiveness of their communications. Thus the major role played by households in economics is now well recognized both in academic and in central banking circles (Bernanke, 2004, 2007).

In the real world, central banks rely on various approaches to measuring inflation expectations, and prototypical inflation forecasting models can be grouped into four families (Stock & Watson, 2009): (1) forecasts based solely on past inflation; (2) forecasts based on activity measures ("Phillips curve forecasts"), such as the unemployment rate, an output gap, or output growth; (3) forecasts based on explicit and implicit forecasts of others, such as implicit expectations derived from asset prices and the explicit forecasts obtained from the Survey of Professional Forecasters; and (4) forecasts based on other predictors.

Existing measures of inflation expectations can be divided into two broad categories in terms of whether they are obtained directly or indirectly. Direct measures are elicited by surveys where consumers, businesses, or professional forecasters are asked to self-report their subjective beliefs about future inflation (Bruine de Bruin, Van der Klaauw, et al., 2010; Curtin, 1996; Ranyard et al., 2008). Indirect measures of public inflation expectations are inferred from market-based measures, financial instruments (such as TIPS, the Treasury Inflation-

Protected Security), the term structure of interest rates, or past realizations of inflation rates. Each of these measures has potential weaknesses (Armantier et al., 2015). Reid and Siklos (2021) summarized the advantages and disadvantages of different surveys; for example, asset price-based measures of inflation are available at high frequencies and a range of horizons, while survey-based methods allow differentiation between groups.

Under the rational expectation assumption, standard macroeconomic models postulate that any economic agent is able to produce a unique inflation expectation for any horizon. Currently, surveys of inflation expectations are widely conducted around the world. They include consumer surveys, such as the Reuters/Michigan Survey of Consumers (MCS), the Federal Reserve Bank of New York's Survey of Consumer Expectations (SCE), and the Bank of England's Inflation Attitudes Survey (IAS). Surveys of professional forecasters include the Federal Reserve Bank of Philadelphia's Survey of Professional Forecasters (US-SPF), and the European Central Bank's Survey of Professional Forecasters (EU-SPF). Together with these, market-based measures and surveys of businesses and external forecasters are used by central banks to inform monetary policy decisions.

Examination of the questions used in surveys of inflation expectations in different countries clearly reveals a high degree of heterogeneity in survey design (Reid & Siklos, 2021). Different forms of questions designed to elicit expectations of inflation include price point forecasts<sup>1</sup> (a best guess), price range (interval) forecasts<sup>2</sup> (a best guess for a range) and probabilistic price forecasts<sup>3</sup>. A point forecast is a simple and clear forecast method for both forecasters and decision-makers. Interval forecasts are assumed to provide information that enables the users to better assess future uncertainties that are associated with inflation forecasts and so to enhance decision performance (Goodwin et al., 2010; Johnson, 1982). Using probability density forecasts to elicit individuals' subjective probability distributions across a range of

---

<sup>1</sup> For example, the Michigan Survey has been measuring inflation expectations in the United States for more than 50 years (Curtin, 2006) by asking, "During the next 12 months, do you think that prices in general will go up, or go down, or stay where they are now?" followed by the response options "go up," "stay the same," and "go down." Those who respond "go up" or "go down" are then asked to give a specific percentage for the change in inflation rate. Median responses to the Michigan question generally track actual inflation estimates such as the CPI and sometimes outperform professional and model-based forecasts in predicting actual inflation (Ang et al., 2007).

<sup>2</sup> As reported in Curtin (2009), some respondents in the telephone survey provide an answer, after which they are prodded for a best guess. For example, web-based instructions to respondents are as follows: 'Below, please give your best guess OR your best a range' followed by answer options 'My best guess is that prices will go [up/down] by \_\_\_percent' as well as 'My best guess for a range is that prices will go up between \_\_\_percent and \_\_\_percent.' Respondents who only fill out the lower bound or the higher bound of the range are prompted to fill out both. Those who only give a range are subsequently also asked for a best guess.

<sup>3</sup> The probabilistic question about expected price inflation follows a format similar to that employed, among others, in the Survey of Professional Forecasters and the Bank of Italy's Survey of Household Income and Wealth. These questions are presented with instructions adapted from those used previously in the Survey of Economic Expectations (Dominitz & Manski, 1997): "Now we would like you to think about the percent chance that different things may happen to prices in general during the next 12 months. The percent chance can be thought of as the number of chances out of 100. You can use any number between 0 and 100. For example, numbers like: 2 and 5 percent may be 'almost no chance,' 20 percent or so may mean 'not much chance,' a 45 or 55 percent chance may be a 'pretty even chance,' 80 percent or so may mean a 'very good chance,' and a 95 or 98 percent chance may be 'almost certain.'" Underneath the question, it states "Please note: The numbers need to add up to 100 percent." Respondents who nevertheless give answers that do not add up to 100 percent receive the notice "Your total adds up to [x%]. Please go back and change the numbers in the table so they add up to 100 percent or choose next to continue."

inflation outcomes has shown that measures of central tendency derived from density forecasts provide more reliable summary expressions of expectation levels than point forecasts (Bruine de Bruin, Manski, et al., 2011; Engelberg et al., 2009).

Probability density forecasting could provide a comprehensive picture of individual behaviours of benefit to central banks and policymakers. However, when people judge probabilities to assess associations or to estimate quantities in uncertain environments, they often make inappropriate inferences and use heuristics (Tversky & Kahneman, 1974) by, for example, applying the 'law of small numbers' (Tversky & Kahneman, 1971). Although some studies have documented correlations between the above three types of inflation expectation question format, relatively little effort has been put into making direct comparisons between them. As a result, I still cannot conclude which of them is the best option for use in surveys.

## **1.2. Economic assumptions about inflation expectations**

Regarding the formation of inflation expectations, different economic assumptions about how agents behave and interact in economic activities have been developed. The rational expectations hypothesis (Lucas, 1972; Muth, 1961) claims that people can learn to use the available economic information to form an unbiased forecast of future inflation and that agents update their expectations continuously. However, since individuals have limited information processing capacity, keeping a track of changing prices for specific goods and of general trends in inflation is a very complex activity and seems unrealistic. Consequently, the principles of bounded rationality are likely to apply (Gigerenzer & Selten, 2001; Kahneman, 2003; Simon, 1957; Weber & Johnson, 2009).

A variety of models have been formulated by taking this constraint into account. For example, the sticky information theory assumes that agents have sticky or noisy information about economic variables so that new information is slow to diffuse through the population (Mankiw & Reis, 2002). Agents only probabilistically pay attention to experts or to news (Carroll, 2003) because of their limited attention and the cost of updating information (Coibion & Gorodnichenko, 2012, 2015; Maćkowiak & Wiederholt, 2009; Mankiw & Reis, 2002; Sims, 2003; Woodford, 2003). In terms of adaptive learning models (Evans et al., 2001), people are considered to act as econometricians when they make forecasts, re-estimating their models whenever new data become available. Various papers have provided empirical evidence for this learning model at the macroeconomic level. For example, models with learning have typically outperformed models with rational expectations in terms of their ability to fit macroeconomic time series (Milani, 2007; Slobodyan & Wouters, 2012a). Survey data also indicate that inflation forecasts are better fitted by learning models (Branch & Evans, 2006; Malmendier & Nagel, 2016; Sorić et al., 2020). The rational expectations hypothesis has generally been rejected in favour of models that explain the formation of expectations of

future prices in terms of adaptive expectations (e.g., Camerer, 1992; Hey, 1994; Schmalensee, 1976).

Some economists have proposed that people utilize simple heuristics that deviate from assumptions underlying the traditional macroeconomic paradigm (Hommes, 2011, 2021; Pfajfar & Žakelj, 2014). In these approaches, people switch between different prediction rules (Branch, 2004; Assenza et al., 2013; Anufriev et al., 2019). Because the economy is too complex to fully understand, especially for laypeople, use of simplified learning rules has been proposed (Hommes & Makarewicz, 2021; Slobodyan & Wouters, 2012a, b). Potential heuristics that have been suggested include adaptive expectations (the dynamics will monotonically converge to the fundamental price, independently of initial conditions), anchoring and adjustment (forecasting using an anchor of a sample average of past prices), and trend extrapolation (forecasting using the last price observation and adjusting in the direction of the last price change).

More specifically, Gennaioli et al. (2016) found that expectations are more extrapolative than rational in corporate investment decisions. Fuster et al. (2010) confirmed that agents employed a simple prediction model that involved extrapolative expectations that were overlaid by biases that resulted in forecasts showing optimism or pessimism relative to those provided by rational expectations. Most agent-based models in macroeconomics build a completely new system from the “bottom-up” by modelling agents’ use of simple micro-decision rules (heuristics) (Dawid & Gatti, 2018). For example, De Grauwe (2012) proposed that agents use boundedly rational heterogeneous expectations and switch between simple forecasting heuristics based upon their relative performance. This approach received support from Hommes (2021) who found that heuristic-switching models fitted observed micro and macro behaviour surprisingly well.

In summary, economists have developed various assumptions and proposed diverse models to predict inflation expectations. Models broadly based on bounded rationality assumptions have been validated in empirical and experimental settings (for a review, see Hommes, 2021). The importance of expectation formation (forecasting) has been acknowledged by psychologists who have investigated people’s underlying mental processes. As Hommes (2021, p. 151) pointed out: “In the last two decades a rich behavioural theory of expectations that fits empirical time-series observations, laboratory experiments, and survey data has emerged that should become part of the standard toolbox for policy analysis”. Hence, the psychological mechanisms underlying the formation of inflation expectation are worth investigating if we are to explain the observed phenomena and research findings.

### 1.3. Characteristics of inflation expectations

#### 1.3.1 Lay inflation expectations

Individual inflation expectations are typically over-estimates: inflation expectations are much higher than actual inflation rates (e.g., Curtin, 2006; Georganas et al., 2014). Bryan and Venkatu (2001a, b) reported that, in the Michigan survey, mean values of household inflation expectations persistently exceeded the inflation rates measured by the Consumer Price Index (CPI). D'Acunto, Malmendier, et al. (2022) plotted the distribution of US households' 12-month-ahead numerical inflation expectations from June 2013 to November 2021 and showed skewness: the means of inflation expectations were systematically one percentage higher than the medians. This indicates that some consumers had over-forecasted inflation.

Inflation perceptions are measured together with inflation expectations sometimes, since it is plausible that consumers use perceived inflation for the development of their inflation expectations (Ranyard et al., 2017). Consistent with the overestimation bias in inflation expectations, inflation perceptions are found to be higher than actual inflation rates (Arioli, et al., 2017; Abildgren & Kuchler, 2021; Ahn & Tsuchiya, 2022). Jonung and Laidler (1988) found that inflation perceptions were, in general, not rational: average values for them were too high, and individual perceptions of inflation were strongly correlated with demographic characteristics and socioeconomic status (Jonung, 1981). For example, the fact that perceptions obtained from survey data tend to be higher for young people, women, unmarried individuals, minorities (non-whites), and lower-income individuals has been documented in the US (Bryan & Venkatu, 2001a). Similar differences by gender and socioeconomic group have been reported in Australia (Brischetto & de Brouwer, 1999), Ireland (Duffy & Lunn, 2008), Denmark (Abildgren & Kuchler, 2021) and New Zealand (Leung, 2009).

Excessively high perceptions of inflation after the Euro cash changeover were found in different samples of the population, not only in the general public in Germany (Traut-Mattausch et al., 2004, Experiments 1–3) but also in students (Greitemeyer et al., 2005; Traut-Mattausch et al. 2004, Experiments 4) and acknowledged financial experts such as bankers and controllers (Schulz-Hardt et al., 2007, cited in Traut-Mattausch et al., 2007). Similar illusorily high perceived price rises were found in samples from other countries, as “some citizens of the European Monetary Union believed that prices had increased with the introduction of the Euro” (Hofmann et al., 2007, p. 368). Moreover, inflation perceptions in Italy were higher for less well-off and less educated consumers, housewives and pensioners than for full-time or part-time workers (Del Giovane & Sabbatini, 2008).

Hommes (2011) has argued that explaining heterogeneity should be a crucial aspect of the theory of expectations. He supported his view by reviewing theoretical, empirical, and experimental research into the modelling of heterogeneous expectations. Findings of dispersed expectations have been documented widely in different contexts, such as inflation expectations (Armantier et al., 2016; Cavallo et al., 2017; Lahiri & Sheng, 2010b), exchange

rate expectations (Menkhoff et al., 2009), and asset prices and equity returns (Chiarella et al., 2014).

Given that laypeople have limited knowledge, lack of understanding of relevant economic concepts and variables, and have restricted mental processing ability, inaccuracies and incoherencies are to be expected. Indeed, they are found in research: individuals' expectations are naturally more diverse than those of professionals (Mankiw et al., 2003). Measures of perceived inflation and expected inflation obtained from national household surveys have been found to be biased. Close study of survey responses derived from the MCS (Bruine de Bruin, Potter, et al., 2010; Bruine de Bruin, Van der Klaauw, et al., 2011; Bryan & Venkatu, 2001a, b; Curtin, 2006; Mankiw et al., 2003), harmonized European Union consumer survey (Lein & Maag, 2011), the Booth Expectations and Attitudes Survey (CBEAS) (D'Acunto, Malmendier, Ospina et al., 2021) and from annual Eurobarometer (EB) surveys (Drakos et al., 2020) have revealed considerable disagreement between respondents, with some reporting relatively extreme inflation expectations and positively skewed distributions. Significant heterogeneity of expectations is also a robust feature of data derived from experimental research (Pfajfar & Žakelj, 2014; Burke & Manz, 2014; Hommes, 2011; Anufriev & Hommes, 2012).

To deal with this, economists have emphasised the importance of modelling heterogeneous expectations (Branch & McGough, 2011; Berardi, 2009). For example, Branch (2004), Pfajfar and Santoro (2010), and Cole and Milani (2021) investigated heterogeneity at an individual level through the lens of the adaptive learning model. However, in addition to simulating the disagreement over inflation expectations, it is important to recognize critical factors that affect expectation formation by individuals in order to acquire a better understanding and to facilitate the efficiency of government policy and its communication.

### *1.3.2. Experts' inflation expectations*

Overestimation and diversity in inflation expectations are not specific to laypeople. Mankiw et al. (2003) found disagreement in inflation expectations among professional forecasters, economists and the general population. Professionals' inflation judgments are not accurate. Instead, disagreement is well-documented in various professional forecaster surveys, such as Blue-Chip Financial Forecasts (BCFF) survey (Andrade et al., 2016), Surveys of Professional Forecasters (SPF) (Coibion & Gorodnichenko, 2012; Mankiw et al., 2003; Lahiri & Sheng, 2010a) and Consensus Economics Survey (Lahiri & Sheng, 2010b; Patton & Timmermann, 2010).

However, it is also found that inflation estimates made by professional forecasters are more accurate and show less disagreement than those made by lay forecasters (Armantier et al., 2013; Coibion, Gorodnichenko, Kumar, et al., 2020; Mankiw et al., 2003). Cornand and Hubert (2022) compared disagreement in expectations and the frequency of forecasts revisions across different categories of individuals and showed similar findings: policymakers and

professionals showed a low disagreement while firms and households exhibited a stronger heterogeneity. Revision frequency was the highest for policymakers followed by firms and professionals. Households revised their forecasts least frequently. Palardy and Ovaska (2015) went further and found the underlying disagreement in expectations was influenced by the visibility of prices of a select few goods and less so by long-term trends. Specifically, consumer expectations were uniquely sensitive to changes in the prices of necessity goods (energy and food) relative to the professional and market expectations.

In terms of overestimation bias, although overestimation was a ubiquitous phenomenon in inflation judgments, there are still some controversial arguments about the performance of experts. Some research indicates that financial experts are biased towards higher expectations (Traut-Mattausch et al., 2007; Chen et al., 2016) whereas other studies have found the reverse effect. For example, Das et al. (2016) used the newly released Inflation Expectations Survey of Households (IESH) and Survey of Professional Forecasters (SPF) from India to compare data from households and experts. They found that these groups differed greatly and that experts consistently underestimated true inflation. Similar patterns were found in Brazil (Gaglianone, 2017). In another study (Dovern & Weisser, 2011), the overall performance of professionals' inflation forecasts was unbiased and there were very few biased forecasters for Canada, the UK, the US, and France. It is possible that professionals tend to anchor their inflation expectations to published inflation targets (Coibion, Gorodnichenko, Kumar, et al., 2020).

In this thesis, I primarily investigate the formation of inflation expectations by lay people.

#### **1.4. Explanations for overestimation and homogeneity of inflation expectations**

A number of factors have been explored to in attempts to understand the nature and reasons for biased perceptions and expectations of inflation, including demographical factors, questions used, personal experience, financial literacy and so on.

##### *1.4.1. Demographical factors*

A large number of studies have documented overestimation and heterogeneity of inflation expectations with respect to demographical characteristics, particularly, related to factors of socio-economic status, education level, personal income levels, age and gender. Specifically, this work has shown that female, lower-income, low-education, and low-financial-literacy respondents report higher inflation expectations in various countries (Armantier et al., 2016; Bruine de Bruin, Van der Klaauw, et al., 2010; Easaw et al., 2013; Johannsen, 2014; Jonung, 1981; Stanisławska et al., 2019). However, reported relationships between age and inflation expectations have been inconsistent, with some studies finding higher inflation expectations among younger participants (Bryan & Venkatu, 2001a; Jonung, 1981; Easaw et al. 2013), and others supporting higher expectations among older participants (Blanchflower & Coille, 2009;

Lombardelli & Saleheen 2003; Malmendier & Nagel, 2016). Also, not all studies have significantly higher inflation expectations in women (Bruine de Bruin, van der Klaauw, et al., 2010).

Because effects of many demographic factors are inter-correlated, some studies have aimed to identify those that are most critical demographical after controlling for others. For example, in a random sample of households, men tend to have higher education levels and earn a higher income compared with women (Bryan & Vankatu, 2001b). Even after adjusting for the respondents' age, race, education, and income, it has been found that women tend to report perceived inflation that is 1.9 percentage points higher than men. Analysis for respondents' predictions of future inflation yielded the same result: after allowing for the effects of other major demographic factors, on average, women expected prices to rise 2.1 percentage points more than men. D'Acunto, Malmendier and Weber (2021) came to a similar conclusion: on average, women reported higher inflation expectations than men. Khor et al. (2020) and Fisher and Yao (2017) found that women had a higher sensitivity to price increases (risk aversion) than men and that, as a result, their judgments were more often categorised as indicating rising inflation. However, in a dynamic modelling study, Corduas (2022) found that the gender difference was smaller during periods in which extreme economic events, such as Euro changeover, occurred.

Other studies using regression models that controlled for interrelationships between demographic variables (Blanchflower & Coille, 2009; Lombardelli & Saleheen, 2003) have indicated that (higher) inflation expectations are mostly driven by (older) age and (lower) socioeconomic status (i.e., less reported education, less income and living in council houses or rented accommodations). Additionally, Hobijn et al. (2009) found that the differences in responses across demographic groups far exceeded the variations in inflation experienced by the groups; this suggests that the differences arise from factors other than the differences in purchasing behaviour between groups. However, previous research has done little to address the psychological underpinnings that account for these demographic differences.

#### *1.4.2. Question interpretation*

It is possible that different interpretations of survey questions lead to differences in the accuracy of different respondents when answering inflation expectation questions. Bruine de Bruin, Potter et al. (2010) manipulated question wording to vary people's interpretations of "inflation". They found participants who answered questions worded as "price in general" gave higher prices for gas and food than those who answered questions worded as "rate of inflation". The larger interquartile range of the distribution of responses across participants for both inflation perception and inflation expectation was also larger in the former case. Similar findings were reported by Van der Klaauw et al. (2008). This work shows that average consumer inflation expectations are heterogeneous and systematically higher than realized inflation, especially among respondents who interpret inflation questions as asking about

their personal expenses or the prices they pay rather than the official US inflation rate (Bruine de Bruin et al., 2012). Hence, it is likely that different question wordings evoke different understandings and interpretations of inflation expectations.

Even for answering questions with the same wording, varied interpretations can occur and affect later responses. Bruine de Bruin, Van der Klaauw, et al. (2011, Experiment 2) found that half of their respondents thought of specific prices when forming their inflation expectations to the question of “year-ahead expectations of inflation”, and this group reported more extreme and more heterogeneous inflation expectations.

The tendency to think more about prominent price changes in one’s own experience and, as a result, report higher inflation expectations is stronger among certain demographical groups. Bruine de Bruin, Van der Klaauw, et al. (2010) showed that “especially for respondents with lower levels of income and education, questions about inflation may trigger relatively stronger concerns about their personal financial experiences, relative to the US inflation rate (p. 398)”. For these people, large price changes tend to be more salient than smaller ones, and increasing prices tend to be more salient than decreasing ones (Brachinger, 2008; Jungermann et al., 2007; Fluch & Stix 2005). As a result, their inflation expectations are biased more in an upward direction.

#### 1.4.3. Purchasing experience

One salient stream of research emphasises that the extremity and heterogeneity of inflation expectations are affected by differences in people’s personal purchasing experiences. On the one hand, it is reasonable to assume that the more experience people have with higher price changes, the more they will tend to produce higher inflation perceptions and expectations. By investigating the effect of items purchased, their purchase frequency and the price paid for them, D’Acunto, Malmendier, Ospina, et al. (2021) discovered that frequency of purchase and positive price changes are significant factors that should be taken into account when modelling how individuals form their inflation expectations. Households that bought items with the highest levels of inflation over the previous 12 months produced higher inflation expectations than those experienced the lowest levels of inflation. Price changes that vary across product categories could partially explain the observed variations in experienced inflation and subsequent inflation expectations (Bryan & Venkatu, 2001a; McGranahan & Paulson, 2006; Ranyard et al., 2008; Das et al., 2019).

It is plausible to assume that the particular price changes that come to mind depend on the actual inflation conditions in the country: Bruine de Bruin, Van der Klaauw, et al. (2011) found more overestimation when the actual inflation was relatively high (“*in periods of moderate to high inflation*”, p.835). During periods of relatively high actual inflation, actual price increases are more common whereas decreases in specific prices are much less frequent. Consequently, people are more likely to produce expectations of extreme inflation than extreme deflation.

To test this explanation, it would be reasonable to assume that more extreme deflation expectations would be elicited during periods or countries showing deflation and that the extent of the effect varies across countries with different levels of inflation. With this in mind, Das et al. (2016, as cited in Das et al., 2019, p. 986) explored the asymmetry and heterogeneity in these quantitative responses in India. They found that a large number of the respondents (who tended to be older, poorer, and generally of lower socioeconomic status) who gave extreme responses were from cities experiencing higher levels of inflation locally, often due to fluctuations in food and energy prices. Findings by Das et al. (2019) lead to similar conclusions. By developing a new and indirect measure of inflation expectation (asking consumers how their incomes would have to change to make them equally well-off relative to their current situation such that they could buy the same amount of goods and services as they can today), Hajdini, Knotek, Pedemonte, et al. (2022) showed that higher inflation experiences were correlated with higher inflation expectations at a city level and at a country level.

The influence of personal purchasing experience has been directly shown to shape inflation expectations. Malmendier and Nagel (2016) found it hard to reconcile empirical heterogeneity in the survey data with the existing models. Instead, discrepancies in experiences strongly predicted discrepancies in expectations, which confirmed the primary assumption that people learn from experience when generating inflation expectations. Using data from Argentina, Cavallo et al. (2017) showed that households use inflation statistics in a sophisticated way: when their own experiences differ from the official statistics, households process the information in a way that is consistent with their own experiences. Additionally, Cavallo et al.'s (2017) findings imply that rationally inattentive economic agents put considerable weight on their own price memories to form household inflation expectations.

According to psychological theories of perceived inflation, extreme price changes are especially salient for categories that are purchased frequently, such as food and gas (Bates & Gabor, 1986; Brachinger, 2008; Jungermann et al., 2007; Ranyard et al., 2008). If a person bought items with increasing prices more frequently, they were more likely to generate higher inflation perceptions and expectations. Nevertheless, the salience of specific price changes can vary across individuals, depending on their personal experiences with price changes in specific product categories. Bruine de Bruin, Van der Klaauw, et al. (2011) showed that, even when individuals focus on the same product categories, they may remember different instances in which prices showed extreme changes and thus produce different judgments of inflation.

Elicitation methods that force people to consider broader ranges of purchasing experience affect individuals' perceived inflation. Hobijn et al. (2009) developed a comprehensive approach to measuring household inflation experience by combining an interview survey and a diary survey. This combined method produced a lower and more accurate description of expenditure than one only using the interview survey. This was because the diary survey

captured a much broader set of expenditure categories whereas people responding only to the interview survey over-weighted items in those categories showing particularly high price changes. Also, the addition of diary information resulted in less dispersion in inflation experiences across households because people put more weight on inflation rates across the categories with uniformly low inflation rates (e.g., food and personal care) when the diary surveys were added.

#### *1.4.4. Financial literacy (FL)*

Over time, the definition of FL has expanded to include more and more assessments of different types of knowledge and skill. For example, it has been conceptualized as “the knowledge of basic financial concepts and the ability to do simple calculations” (Lusardi et al., 2010). Huston (2010) went further and associated FL with two dimensions: understanding knowledge of personal finance and application to business finance. Research has also resulted in a broader interpretation of FL as a measure of the degree a person understands key financial concepts and possesses the ability and confidence to manage finances to make proper decision-making and sound plans in the short term and in the long term while taking into account economic events and changes (Remund, 2010). Furthermore, FL has also been defined to include financial awareness, attitude and knowledge (Tuffour et al., 2020). Moore (2003) devised a proxy measure of FL by calculating scores for knowledge level, experience levels, and positive and negative financial behaviours.

Measurement of FL in academic settings uses self-reported surveys comprising True/False and multiple-choice questions that can be scored as correct or incorrect. For example, Lusardi and Mitchell (2007) used five basic financial literacy questions and eight advanced financial literacy questions with only one correct answer for each question. FL has been measured in different countries and organizations using a variety of question items, including De Nederlandsche Bank’s Household Survey (DHS, Tilburg), the National Financial Capability Study (NFCS, America), the Rand American Life Panel (ALP, America), the CogEcon Survey (America), and the China Household Finance Survey (CHFS, China).

In addition to these objective measures of FL, some subjective assessments of FL have also been employed (Lusardi & Mitchell, 2007; Liao et al., 2022), including, for example, the question: “On a scale from 1 to 7, where 1 means very low and 7 means very high, how would you assess your understanding of economics?” (Lusardi & Mitchell, 2007). This measure is used to quantify the perceived FL rather than the actual FL levels. Because it is related to the confidence level, individuals tend to answer according to their own perception of themselves rather than on the basis of knowledge about their actual FL levels. To achieve a similar goal, Bruine de Bruin, Van der Klaauw, et al. (2010) directly asked for confidence judgments: after answering each true/false statement on the financial literacy measure, respondents indicated their confidence in their answers to objective FL questions, on a scale anchored at 50% (just guessing) and 100% (absolutely sure).

Regarding the importance of FL, it has been confirmed that FL affects certain financial behaviours such as mortgage transactions (Moore, 2003), retirement planning (Lusardi & Mitchell, 2007), stock market participation (Almenberg & Dreber, 2012), personal savings (Bayar et al., 2017) and stock investment returns (Liao et al., 2022). As for inflation judgments, direct evidence that groups with lower FL levels reported higher inflation expectations have been confirmed by Bruine de Bruin, Van der Klaauw, et al. (2010). In their study, 16 questions were selected and extracted from different studies comprising the direct measure of the understanding of inflation, basic numeracy, and advanced numeracy because those items were judged to be more associated with inflation. However, those questions are hardly representative of the FL levels of individuals.

Given its definition and measurement, it is unlikely that FL is directly acquired from pure education. However, FL does appear to be related to general educational level (Bruine de Bruin, Van der Klaauw, et al., 2010; Moore, 2003) and some other demagogical factors. Specifically, among different demographic groups, women, lower incomes individuals, and lower educated individuals are more likely to show lower financial literacy scores compare to their counterparts (Potrich et al., 2015). This is consistent with other findings showing that FL is particularly low among older women, Blacks and Hispanic groups, and the least educated people (Lusardi & Mitchell, 2011a).

#### *1.4.5. The role of information*

Burke and Manz (2014) studied how economic literacy affects formation of inflation expectations and argued that it was influenced through two specific channels: the selection of information and the use of given information. These two channels have been studied to explain the inaccuracy and disagreement of inflation expectations.

Regarding the information used in forming inflation expectations, some researchers have argued that inaccurate expectations from laypeople are due to information constraints or information inattention (Binder & Rodrigue, 2018; Cavallo et al., 2017; Carroll, 2003; Mankiw & Reis, 2002). Saakshi and Sahu (2019) analysed heterogeneity in household inflation expectations across cities in India. They suggested that information friction is a main source of heterogeneity. In addition, El-Shagi et al. (2012) compared different professional forecasters and showed that the Fed's staff were much more accurate in projecting inflation trends than private forecasters, primarily because they had the advantage of additional information. Romer and Romer (2000) noted that this information advantage leads to commercial forecasters discarding their own judgments in favour of those produced by the Fed.

Once the information was provided, laypeople were able to respond to it actively. Roos and Schmidt (2012) and Armantier et al. (2016) experimentally demonstrated that individuals reacted to information about German and US inflation statistics by adjusting their reported

inflation perceptions. When people were reminded about products with extreme price changes, they tended to report high inflation expectations (Bruine de Bruin, van der Klaauw, et al., 2011). Galashin et al. (2020) confirmed the effect of information using a natural field experiment administrated as a phone survey in order to elicit the respondents' expectations of exchange rate and inflation in Malaysia. Their findings showed that some individuals updated their macroeconomic expectations in response to information about expert forecasts of inflation and exchange rate expectations. Carrillo and Emran (2012) showed that public signals about prices play an important role in households' price expectations, and that changes in price expectations affected their savings choices. Furthermore, the effect was heterogenous: the impact of the public information (published inflation rate) on household inflation expectation formation was stronger among more educated and older individuals.

Some studies have examined the importance of media coverage of inflation statistics. They have found that, although increased intensity of news coverage of inflation statistics improves the consumer forecasts (supporting theoretical concepts regarding information rigidities), the media content (tone) in media reports reduces consumers' forecast accuracy and increases disagreement (Lamla & Lein, 2015; Badarinza & Buchmann, 2009; Lamla & Maag, 2012; Dräger, 2015).

#### *1.4.6. The use of information*

A large number of studies have shown that mental models of the formation of inflation expectation differ between forecasters and have suggested that mental models have an importance beyond the information that they represent. In Lahiri and Sheng's (2010b) economic forecasting models of GDP and inflation, disagreement among forecasting professionals arises from two primary sources: differences in their initial prior beliefs and differences in their interpretation of public information. These authors argued that the expert disagreement was due to different use of the models, methods and philosophies of the data rather than due to differences in the availability of data. This notion is consistent with the findings of Döpke and Fritsche (2006) indicating that forecasters do not share a common model of the economy. This is also true for professionals and laypeople.

People incorporate diverse information into their judgments. For example, Lamla and Maag (2012) have shown that the media play a role in producing disagreement between consumers but not between professional forecasters: the latter tend to restrict themselves to considering the most recent and pertinent news information. Palardy and Ovaska (2015) have shown that the underlying disagreement found in consumer expectations is uniquely sensitive to changes in the prices of necessity goods (energy and food) relative to the professional and market expectations. In contrast, professionals' inflation expectations are well-anchored to short-term monetary policy or inflation targets or to long-term average targets (Coibion, Gorodnichenko, Kumar, et al., 2020; Gábrriel et al., 2014).

People use information inefficiently or even mistakenly. Carroll (2003) showed that consumers in the United States updated their information about once a year and Döpke et al. (2008) found consumers in Europe did so about once every 18 months. Research has also revealed that information friction may arise from inattention. Cavallo et al. (2017) found that individuals in low inflation contexts had weaker priors about the inflation rate compared to those in high inflation contexts, which is consistent with a rational inattention model. For example, when provided with information about inflation statistics or prices of specific supermarket products, individuals in a low inflation context (the USA) assigned a weight of just 15 percent to their prior beliefs, whereas individuals in a high inflation context (Argentina) assigned a weight of roughly 50 percent to their prior beliefs. This suggests that in a low inflation country, such as the USA, individuals are less informed. Conclusions consistent with these findings have been reported by Carroll (2003), Armantier et al. (2016) and Coibion and Gorodnichenko (2015).

In addition, Cavallo et al., (2017) provided information from different sources, including inflation statistics and tables with historical prices of specific supermarket products, using a series of online and offline surveys. Their results indicated effects of cognitive limitations: individuals placed significant weight on less accurate sources of information retrieved from their own memories even though statistics were available to them. A study by Galashin et al. (2020) showed that provision of inflation forecasts from experts significantly influenced updating of inflation expectations but not of exchange rate expectations. Likewise, the provision of experts' forecasts of exchange rates significantly influenced updating of exchange rate expectations but not of inflation expectations. Therefore, people do not appreciate that information about the nominal exchange rates is correlated with indicators of inflation. Similarly, Armantier et al. (2016) showed that information about food and beverage price changes caused limited pass-through to consumer' inflation expectations since it had no meaningful impact on expectation revisions, implying that "providing information to respondents does not guarantee more accurate expectations" (p. 522).

Different types of people use information to update their knowledge in different ways. By comparing different samples, Blomqvist (1983) showed that lay people tend to use an adaptive expectations rule whereas more informed people (economists) more often use an extrapolative rule to predict future levels of inflation. Furthermore, Pfajfar and Žakelj (2014) performed laboratory experiments with 216 professionals (predominantly economics and business majors) and found that rationality held for about 40% of them but that more than 20% of them used adaptive learning models. Interestingly, their behaviour was better described by switching between models than by using a single forecasting model.

Learning effects after being provided with information have been investigated in various studies (e.g., Galashin et al., 2020; Cavallo et al., 2017; Bottan & Perez-Truglia, 2020). Cole and Milani (2021) used data from the Survey of Professional Forecasters (SPF) to conclude that professionals' expectations were heterogeneous and attributed this to different gain

coefficients being adopted by different forecasters. Cavallo et al. (2017) also found heterogeneity in learning rates in a manner consistent with existing literature: i.e., female, less educated, and younger individuals in the United States tended to be less informed about inflation and have more biased inflation expectations (Armantier et al., 2016; Bruine de Bruin, van der Klaauw, et al., 2011; Malmendier & Nagel, 2016). Galashin et al. (2020) summarized three reasons why the degree to which people incorporate the information can be lower than that reported in other studies: a) educational level and confidence in their prior beliefs, b) type of survey method (a computer screen vs a phone survey), and c) provision of financial incentives or not.

#### *1.4.7. Expectation-consistent bias*

In addition to inflation-related information, internal beliefs have an important role in inflation expectation formation. Their effect produces “expectation-consistent bias”, a term derived from social psychology (Asch, 1946). In a simulated macroeconomic model, Adam (2007) studied forecasts of future inflation rates and found people adopted a ‘restricted perception equilibrium’, in which they used simple forecast functions, outcomes and beliefs to reinforce each other. Thus, when the higher price increases were expected by participants, higher overestimation occurred when they estimated trends in prices (Traut-Mattausch et al., 2004, Experiments 2 and 4). This expectancy-consistent judgment bias is a robust effect that has been found in a series of experiments in which judgments are made in the presence of ambiguous information (Asch, 1946; Darley & Gross, 1983; Lord et al., 1979) and in the presence of unequivocal counter information (Traut-Mattausch et al. 2004; Traut-Mattausch et al., 2007).

People typically evaluate new ambiguous evidence in such a way that it supports their own initial beliefs (Edwards & Smith, 1996; Greitemeyer & Schulz-Hardt, 2003; Lord et al., 1979). For example, they more frequently seek erroneous confirmations of the expectations than erroneous disconfirmations of them, and they overweight evidence that confirms their expectations, phenomenon known as the “prior belief effect” (Edwards & Smith, 1996; Koehler, 1993; Lord et al., 1979; Traut-Mattausch et al., 2004).

Expectations or beliefs influenced inflation judgments even when clear disconfirming evidence was available. In several studies during the Euro changeover, it was demonstrated that inflated expectations for price trends were related to biased price trend estimations: following the introduction of the Euro notes and coins in Germany, German people ‘perceived’ dramatic price increases. Specifically, after introduction of the Euro, there was a big difference between their inflation judgments (usually above 10%) and the actual consumer price index reported by the Federal Statistical Office in Germany of 1.2% (Traut-Mattausch et al., 2007).

More direct evidence was obtained in a series of experiments reported by Traut-Mattausch et al. (2004). In four studies, participants estimated the trend in restaurant prices by examining two menus: an old menu with prices in German Marks (DM) and a new menu with prices in Euros. Price trends in menus were systematically manipulated: They were either +15%, 0%, or -15%. Participants' judgments showed a consistent bias towards rising prices (in spite of disconfirming evidence) and this was related to their expectations concerning price increases. Greitemeyer et al. (2005) also provided evidence supporting the causal influence of consumers' expectations on their price trend judgments. They found that an expectation of rising prices led participants to perceive increased prices when those prices were actually stable. Conversely, an expectation of stable prices led to the underestimation of price increases. This excludes the possibility that people have a general tendency to expect price increment as suggested by Traut-Mattausch et al. (2004). It is also worth noting that effects of expectation on perception imply inefficient use of available information, as shown by evidence of "anchoring effects" (Lahiri & Sheng, 2010b).

#### *1.4.8. Other factors*

Personality traits have been recognized to be affect inflation judgments. Abildgren & Kuchler (2021) analysed data obtained from Harmonised EU Programme of Business and Consumer Surveys covering the period from August 2007 to December 2016. They found that households showing "over-pessimism" overestimated of inflation level more than those showing "justified pessimism". These findings persisted in their subsequent interview rounds of the survey. This is consistent with the findings of Fluch & Stix (2005): Austrian individuals with a negative attitude towards to the Euro cash changeover were more likely to perceive price rises than those with a positive attitude. As they also found a positive correlation between inflation perception and inflation expectation, it is reasonable to assume that "over-pessimism" would also affect inflation expectations. Del Missier et al. (2016) also confirmed that attitudes towards inflation influence perceived inflation together with the factor of product accessibility.

Another factor that has been recently considered is cognitive ability. D'Acunto et al. (2019) showed that individuals with lower cognitive abilities (median-to-low IQ levels) form higher inflation expectations and have larger absolute errors. They also investigated which types of cognitive abilities matter in the formation of inflation expectations. Results showed that arithmetic, verbal, and visuospatial cognitive abilities were all important.

### **1.5. Confidence (uncertainty) in judgments**

Faced with uncertainty, people may express confidence levels in their judgments and these may determine whether decisions implied by those judgments are translated into action and decision-making (Fitzgerald et al., 2017; Gill et al., 1998; Navajas et al., 2017; Sniazhko, 2019). In the economic world, the level of certainty associated with inflation expectations can influence the actual financial behaviours of laypeople, such as their allocation of wealth or

assets. As a result, distorted perceptions and expectations are, in aggregate, more likely to have real consequences for the economy when people are more certain about their judgments. Hence, confidence judgments should be investigated along with the inflation perceptions and expectations to determine how likely it is that judgments will influence actual behaviours. It is important to find appropriate and accurate measures of the uncertainty in inflation expectations.

Uncertainty can be measured directly by eliciting range forecasts and density forecasts (Armantier et al., 2013; Blomqvist, 1988; Bruine de Bruin, Potter, et al., 2010; Van der Klaauw, et al., 2008). Bruine de Bruin, Manski, et al. (2011) explored the uncertainty of inflation forecasts using surveys and they found internal consistency of interval forecasts and density forecasts measures. Bruine de Bruin, Manski et al. (2011, p.475) also established the relationship between uncertainty and judgment revision: they showed a Bayesian updating process at the individual level over time, such that 'a more diffuse prior at one point in time is associated with larger revisions in point forecasts in such subsequent periods.'

Besides revision behaviour, some indirect measures of uncertainty were investigated. It has been found that the rounding behaviour (reporting figures in multiples of five or ten) is associated with uncertainty level (Binder, 2017; Krifka, 2009; Reiche & Meyler, 2022). People who are more uncertain are more likely to round their inflation perceptions and expectations. This, in turn can cause overestimation and increases in heterogeneity (Reiche & Meyler, 2022). In addition, Lahiri and Sheng (2010a) showed that aggregate forecast uncertainty can be expressed as the disagreement among the forecasters plus the perceived variability of future aggregate shocks.

Bruine de Bruin, Potter et al. (2010) investigated factors that influence laypeople's reported uncertainty. They found that different question wordings can affect the interpretation of questions as well as confidence in their answers. Using the "rate of inflation" wording revealed a slightly higher non-response rate than using the normal Michigan Survey wording of "prices in general". However, it induced less disagreement and caused a lower overall uncertainty for both near- and long-term inflation expectations and for perceptions of past inflation. This suggests that people interpret the "rate of inflation" wording more consistently and with more certainty.

Confidence judgments usually increase with level of knowledge or education (Harvey et al., 1987; Harvey, 1994). Bruine de Bruin, Manski et al. (2011) found that people with higher education levels were more reluctant to update their inflation judgments when they were provided with additional information; this implies that they possessed a higher level of certainty and more confidence in their initial judgments than less educated people. Reiche and Meyler (2022) found that lower formal education levels contributed to higher uncertainty levels (classified by rounding behaviours) and higher inflation expectations. However, higher confidence levels were not always associated with more accurate judgments. As Stephensen

et al. (2021, p. 6) point out “although experience is often intuitively associated with expertise, the relationship between experience, confidence, and the degree to which a judgment corresponds with reality is not straightforward.”

As with over-estimated inflation expectations, inflation uncertainty varies systematically across socio-demographic groups other than education. Reiche and Meyler (2022) found that people who are lower in age, female, belong to a lower income group, have worse economic sentiments, and have a lower ability to save are among those who typically show higher uncertainty and greater heterogeneity in their inflation expectations. Reiche and Meyler (2022, p. 29) point out that: “Socio-demographic characteristics which tend to increase uncertainty about inflation (for instance younger, female, low education and low income) are also those that are found to cause overestimation. Therefore, a significant portion of the heterogeneity across consumers may be driven by heterogeneity in certainty”.

In a summary, various factors that influence uncertainty in inflation expectations have been identified and it appears that uncertainty contributes to overestimation and heterogeneity in inflation expectations. However, experts, who have economic knowledge, information advantages and relatively high confidence levels, still show inflation expectations that are heterogeneous and too high – though these effects are not as large as those found in laypeople. The mental processes underlying inflation expectation remain unclear. In the following sections of Part 1, psychological mechanisms of inflation expectation formation will be reviewed in an attempt to provide explanations for the formation of laypeople’s inflation judgments.

### **1.6. A model from Ranyard et al. (2017)**

The rational expectations hypothesis is violated because laypeople and professionals produce biased and heterogeneous inflation judgments. Even eliminating information rigidity does not guarantee the formation of accurate inflation judgments because of the well-documented limitations in human cognitive processes.

Before introducing a mental model of inflation expectation, it is useful to discuss two types of information that people receive every day. It is commonly recognized that people can use two kinds of information when making judgments and decisions: a) the information obtained in the form of descriptions provided by others and b) the information drawn from their own experience. Most decision-making research has been based on the former (Fantino & Navarro, 2012): people are given information about the potential outcomes of their choices and associated probabilities by reading descriptions of available options (Kahneman & Tversky, 2013a, 2013b; Tversky & Kahneman, 1992). However, in everyday life, individuals often make decisions based on their direct experience: they obtain information by observing samples of outcomes over time (Knox et al., 2012; Yechiam & Rakow, 2012). Of course, in some cases, people integrate their experience with descriptions of choice outcomes, for example, Weiss-

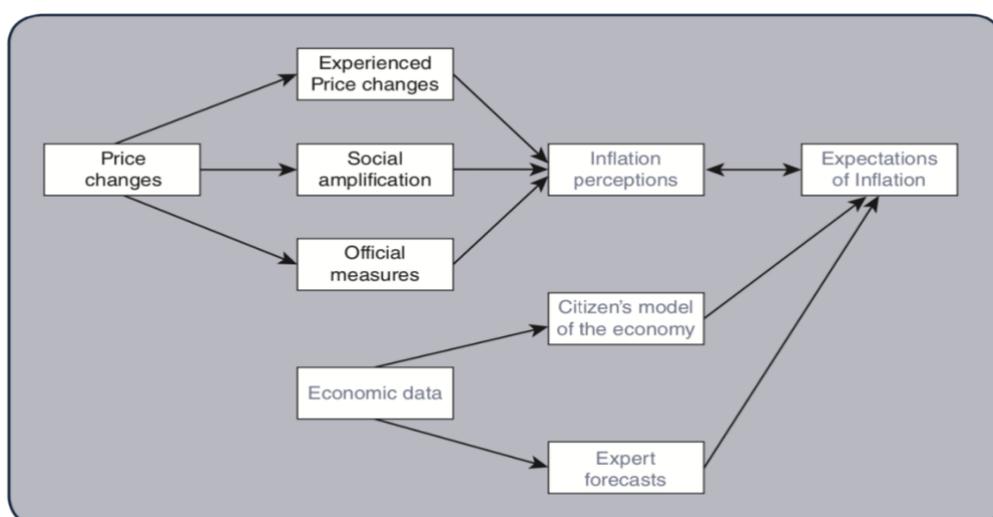
Cohen et al. (2016) found that, in these circumstances, experience is the dominant source of information, though descriptions are also taken into consideration.

Similarly, people mentally process economic information from a diversity of sources to form their economic judgments. Regarding inflation-relevant information, laypeople can learn from direct exposure to prices (and price changes) as well as from indirect information obtained from the media or from other people (Svenson & Nilsson, 1986; Wärneryd, 1986; Ranyard et al., 2008; Leiser & Drori, 2005). Thus, people can use their direct experience of prices, descriptive information (when available), or combine these two types of information to form their inflation expectations.

The most comprehensive work in this area is that of Ranyard et al. (2017). They summarized a conceptual framework for understanding perceived and expected inflation (Figure 1.1). When prices change in the market, evaluation of present (past) inflation is derived from various sources of information: personal experience of price changes (which are biased by frequency of purchase and salience), media reports (which amplify and bias perceptions), official statistics and prior inflation expectations (Ranyard et al., 2008). The formation of expectations of future inflation involves the resources that are used to form inflation perceptions, along with expert forecasts (when available) and lay mental models of the economy (Ranyard et al., 2008; Ranyard et al., 2017; Leiser & Shemesh, 2018). In other words, information sources that are responsible for consumers' inflation expectations include a) directly experienced price changes, b) descriptive information of official measures and c) descriptive information from social media as shown in Figure 1.2. Experts' forecasts are often classified as official measures because they are published by authorities.

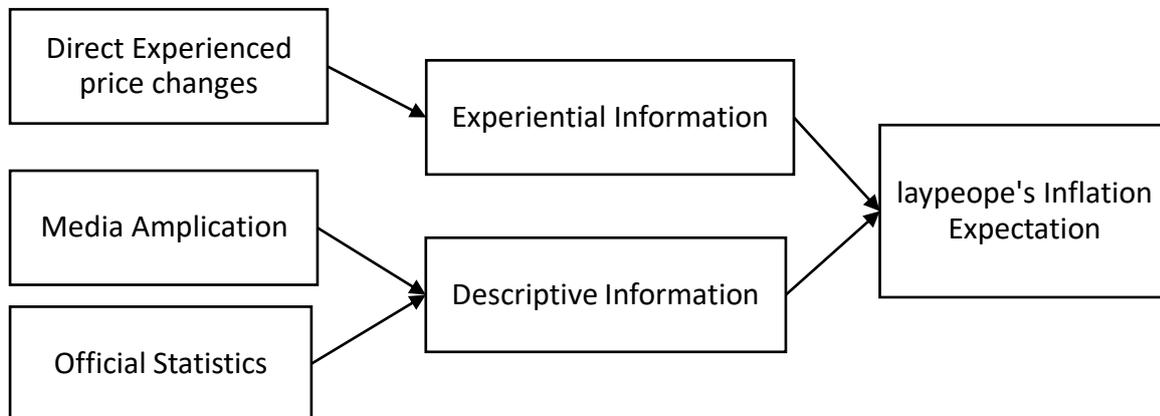
**Figure 1.1**

*Sources of information for inflation perceptions and expectations (extracted from Ranyard et al., 2017)*



**Figure 1.2**

*Information contributing to laypeople's inflation expectations*



### *1.6.1. Direct experience*

Bastounis et al. (2004) ran an extensive survey of economic beliefs in several countries and found that lay economic thinking is based on circumscribed economic phenomena (individual perceptions) rather than on integrative theories. It provided straightforward support for the significant role of direct experience in economic settings. Specifically, in the inflation context, individuals' inflation expectations are constructed and influenced directly by their perception of past and current price changes (Carlson & Parkin, 1975; Gärling & Gamble, 2008; Jonung, 1981; Bruine de Bruin, van der Klaauw, et al., 2011).

It is well documented that consumers pay attention to personal purchasing experience in everyday life when forming their inflation expectations (Bruine de Bruin, Potter, et al., 2010; Bruine de Bruin, Van der Klaauw, et al., 2011; Bruine de Bruin, et al., 2012). Based on micro-level panel data from six Euro-area countries participating in the survey, Stanisławska & Paloviita (2021) provided robust evidence showing that consumers' short-term inflation expectations were affected by evaluations of current price changes. Additionally, consumers revised their medium-term inflation expectations in response to their changing views about current inflation rather than in response to actual inflation. The survey confirmed the significant role of price changes on inflation expectation formation.

Additionally, Madeira and Zafar (2015) developed a learning model of inflation expectations using the panel data of the Michigan Survey of Consumers. It included a significant role for life experience in the formation of one-year-ahead inflation expectations. Consistent with adaptive learning models, Malmendier and Nagel (2016) proposed a learning-from-experience model in which inflation expectations rely on individuals' accumulated personal experiences. It distinguished differences in learning between older and the younger people. Differences in lifetime inflation experiences explain the variation in expectations across age groups shown in cross-sectional survey data.

Biased inflation judgments can be partially explained by the price information. Inflation expectations that are too high have been found when people have been explicitly asked to think of specific prices (Bruine de Bruin, Van der Klaauw, et al., 2011). In the surveys of Bruine de Bruin, Potter, et al. (2010), questions were asked about past (realized) changes in prices and it was found that some dispersion of future price changes was associated with the dispersion of past price changes. D'Acunto, Malmendier, Ospina, et al. (2021) used micro data collected from a representative US sample to further support the notion that it is the frequency and the size of price changes that matters for inflation judgments. This is likely to be because more frequent stimuli have a greater influence on people's perceptions because they are more representative of people's purchases or because they are easier to bring to mind (availability).

### *1.6.2. Official statistics*

Official statistics are an important input for inflation judgments. They include past inflation data, price data and professionals' forecasts. Armantier et al. (2016) showed that people use past inflation statistics to update their inflation expectations by providing them with data from the Bureau of Labor Statistics for the average price of food and beverages in the USA. Roos and Schmidt (2012) conducted an in-class survey in Germany and showed the historical time-series information had a crucial impact on both inflation forecasts and GDP forecasts. Participants were randomly allocated to one of three groups to receive different types of information: a) two full time-series charts of German monthly inflation rate and the quarterly rate of GDP growth over the last 60 periods, b) the same charts but without labelled variables, and c) no information. Results showed that historical information was a dominant factor no matter whether it was labelled or not.

However, people do not always evaluate the inflation statistics correctly. In an online experiment, Cavallo et al. (2017) manipulated four sources of information concerning the previous 12 months' inflation rate: supermarket purchase price changes (containing six products), the official statistics of inflation, both of these pieces of information, and a hypothetical product price change (to test for spurious learning). Participants estimated past inflation first. Then they received either one of the four types of information or no information. Finally, they forecast the inflation rate for one year ahead. Results showed that, although information about inflation statistics and supermarket prices affected inflation expectations, individuals irrationally assigned significantly higher weights to information given to them about supermarket price changes that were closer to their own experiences. (Ideally, people should have put all weight on the official statistics of inflation when both kinds of information were available).

When people access reported experts' forecasts, they are prone to incorporate them and adjust their inflation expectations. Carroll (2003) showed that households updated their expectations probabilistically to take account of news coverage of information concerning the

Survey of Professional Forecasters. Expert inflation forecasts obtained from professional surveys caused consumers to update their expectations considerably, showing more significant changes than those produced by information about recent prices for food and beverages (Armantier et al., 2016).

Although these findings showed the effects of providing official statistics, the manner in which people comprehended and appreciated the available information was not clear. For instance, Armantier et al. (2016) found that the provision of food and beverage *price information* had limited value on consumers' *inflation* expectations. It is possible that participants did not fully understand the concept of inflation expectation and that, as a result, it was difficult for them to use price information to modify their inflation expectations. In contrast, experienced price changes were over-weighted. When individuals own experiences differed from the accurate official statistics, households processed the information in a way that was consistent with their own experiences of prices by putting more weight on information sourced from their *own memories* (Cavallo et al., 2017).

Therefore, much work remains to be done to investigate how people evaluate and use statistical information received from official sources.

### 1.6.3. News media

Media news is another source of input into expectation formation. On the one hand, it provides plenty of information; consumers' inflation expectations are more accurate when they have received more news information (Dräger, 2015; Cavallo et al., 2017). For example, Badarinza and Buchmann (2009) found that the distortion of consumers' inflation expectations was negatively related to news intensity in the Euro area after the introduction of the common currency. On the other hand, the tone of the news reports matters as much as its amount and contents. Indeed, Soroka (2006) showed that the media reported negative news more extensively than positive news, resulting in asymmetric news coverage and Van Raaij (1989) found that such selective reporting led readers to overestimate the importance of causal mechanisms and tended to make them over-react.

Lamla and Lein (2014) suggested that these two aspects of media news (amount and tone) correspond to two different ways ('channels') in which the media influence individuals' inflation expectations. The volume channel refers to the higher intensity of media reporting that makes consumers more likely to pick up news about inflation and update their expectations. The other channel is the tone channel. This is often labelled as a media bias, a tendency to report positive or negative news that causes biased judgments. They showed that more news on rising inflation improves forecast accuracy only if it is more neutrally phrased. Lamla and Maag (2012) developed this framework and showed that disagreement between households in their inflation forecasts depends on the content of news (tone) but not on reporting intensity (volume) or on the heterogeneity of story content (information entropy).

They also reported that disagreement between professionals did not depend on media coverage at all, a finding that suggests media information plays no part in the formation of inflation expectations by professionals.

Asymmetry in reactions to news have also been found (Dräger, 2015; Lamla, & Lein, 2014): people overreact to bad news relative to good news, possibly due to an asymmetric loss function or loss aversion (Soroka, 2006; Dräger et al., 2014; Fluch & Stix 2005). Respondents' focus on large price increases (bad news) is exacerbated after receiving media news (media amplification), thereby affecting public perceptions even after controlling for actual economic conditions (Goidel & Langley, 1995). Consequently, biased inflation perceptions (because of biased reporting by media) lead to biased expectations of inflation (Ranyard et al., 2008). Soroka (2006) also demonstrated that reports of negative changes in other economic indicators (e.g., the unemployment rate) had a significant impact on their expectations, a finding that further supports the generality of the media amplification hypothesis.

Del Giovane and Sabbatini (2008) reported that measured and perceived changes in inflation after the Euro changeover in Italy could be partially explained by the interplay of perception and media coverage. They chose two leading newspapers to investigate the relationship between inflation perception and the number of newspaper articles covering the changeover. They performed Granger-causality tests and revealed a bidirectional nature of this link, such that "a sharp deterioration in inflation perceptions is newsworthy, and extensive media coverage may, in turn, validate and reinforce individual perceptions" (pp.47). A large amount of coverage about price changes and the contrast between people's perceptions and the statistics reported in the media led to a sudden increase of inflation perception. This confirmed the effect of media amplification and further fuelled the perception of a general price acceleration. Boeri (2004, as cited in Del Giovane & Sabbatini, 2008) also emphasized the exceptional media coverage of the Euro changeover contributed to the difference between perceived and official inflation in Italy. He also mentioned the issue of information inaccuracy because many newspapers reported non-official estimates of inflation without checking whether they were correct.

Recent work has identified the impact of different media channels on households' inflation expectations. Conrad et al. (2022) collected data from the Bundesbank Online Pilot Survey on Consumer Expectations (SCE) in June 2019. The survey used a representative sample of the German population. The authors provided preliminary evidence for a link between information channels and inflation expectations: households who relied on traditional media (e.g., newspaper, television, radio) reported lower and more accurate expectations and lower uncertainty compared with social media users. They suggested that central banks should put more emphasis on disseminating accurate information through social media channels other than traditional media. Because they did not report the accuracy analysis of the social media information, it is hard to assess whether the overestimation associated with exposure to

social media was caused by inaccurate information. The difference between traditional media and social media could be explored more to identify how their influences differ.

Media information exhibits different effects on inflation expectations with respect to short-term and long-term forecast horizons. Long-term inflation expectations are usually anchored on the long-term inflation targets which are set by central banks. In contrast, short-term inflation expectations react actively to news and surprising information. Armantier, Goldman et al. (2022) investigated how consumers responded to inflation surprises differently before and after the pandemic using data from the New York Fed's Survey of Consumer Expectations (SCE) and from the Michigan Survey of Consumers (MSC). One-year-ahead inflation expectations demonstrated that people were very responsive to inflation news before the pandemic (0.69 revision) as well as after the pandemic (0.75 revision). But three-year-ahead inflation expectations after pandemic (0.19 revision) were less responsive to inflation surprises than before pandemic (0.45 revision). This suggests that, after the pandemic, people did not take news into consideration as much when updating their well-anchored long-term inflation expectations.

These studies have shed a light on the effect of media on the formation of inflation expectations in terms of biased attention and biased responses. However, there are still many questions that remain, such as the extent to which the media influences expectations of inflation in comparison with direct experience of price changes, how individuals evaluate and the trust of the different sources of reported information, and how consumers perceive news and integrate that information into their short-term and long-term inflation expectations.

#### *1.6.4. Interactions between perceptions and expectations*

The final aspect of Ranyard et al.'s (2017) model that I need to discuss is the interactive effect between inflation perception and inflation expectation. There is a bi-directional linkage between them (Figure 1.1). The effect of expectation on perception is consistent with Bartlett's (1932) hypothesis that memory is reconstructive, and that people store and retrieve information according to their expectations.

This feature of the model is closely related to the expectation-consistent bias discussed earlier. Traut-Mattausch et al. (2004) used four experiments to demonstrate that a major cause of misperception underlying distorted inflation expectations arises from expectation confirmation. Greitemeyer et al. (2005) showed that the expectancy-consistent judgment bias happens not only with expectations of increasing trends but also with expectations of decreasing trends. By using conventional materials of rent prices in different cities, Schulz-Hardt et al. (2007, as cited in Traut-Mattausch et al., 2007) showed that the bias occurs in an ecologically valid context and that it is not limited to the case of the Euro introduction.

## 1.7. Psychological mechanisms underlying inflation judgments

### 1.7.1. Cognitive processes

Inflation, as a macroeconomic factor, deals with the economic functioning of a country as a whole, what is often called the behaviour of the aggregate economy. Additionally, the markets for goods and services, labour, finance and currency are all interrelated. However, the mental model of economy of laypersons is relatively simplified. Given the lack of knowledge and causal models of economics, laypeople usually employ an association-based system. In other words, people compute similarity and statistical structure between the current stimulus and previously associated stimuli rather than using a rule-based system that would allow them to reason on the basis of an underlying causal or mechanical structure (Sloman, 1996). Thus, lay thinking about the economy focuses on direct links and uses simple one-way linear causal chains (Leiser & Shemesh, 2018), which means that people ignore indirect links, feedback loops and aggregate effects (which are the main causal factors in economics).

Given that citizens' mental models of economics include different information sources and factors (Ranyard et al., 2017), it is reasonable to assume that heterogeneity of inflation expectations derives partly from the use of different types of information (e.g., purchasing experiences, media news and official statistics). Although Ranyard et al.'s (2017) framework covers information sources that feed into inflation perceptions and expectations, it is still unclear how this information is processed and organized in lay people's minds. Information obtained from the outside world is not the only factor that contributes to the accuracy of judgments; appropriate utilization of that information is also important. Hammond and Summers (1972) suggested that judgment accuracy is influenced not only by the appropriate weighing of the cues but also by the way these weights are used by the person.

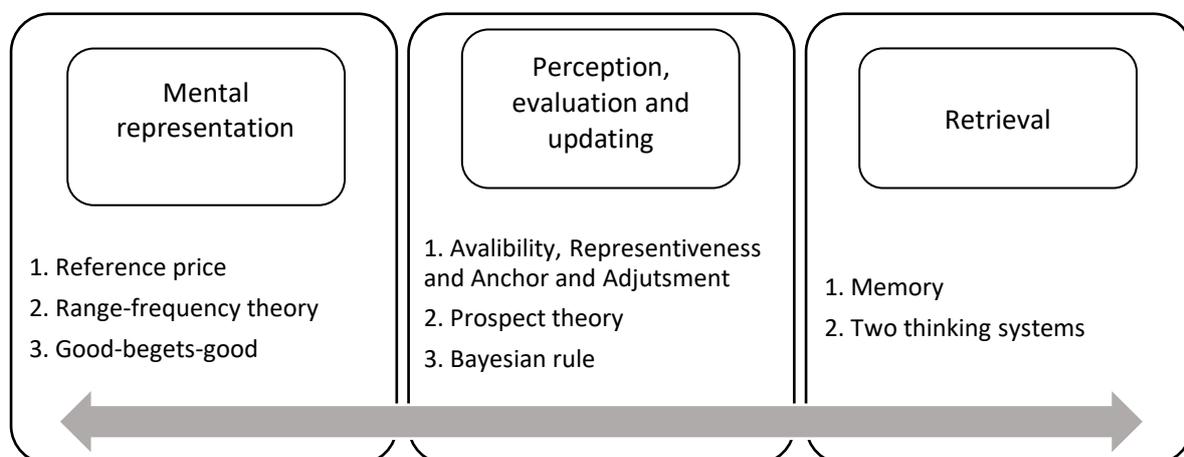
Regarding the mental processes underlying inflation judgments, Huber (2011) proposed a model of perceived inflation, which posits a two-phase process. The two distinct phases comprise a) a price experience phase, which is a continuous process of gathering experience about price changes during purchasing of goods and b) an integration phase, in which all recent experiences are combined into a single estimate of perceived inflation. However, Huber (2011) though included experience of prices, he ignored other external information sources, such as statistics, which were included in Ranyard et al. (2017)'s framework.

Also, his two-step simple mental procedure for producing inflation judgments assumed that people are rather rational and neglected the possibility that both phases could be affected by human biases, including those arising from use of the anchoring-and-adjustment, representativeness and availability heuristics (Tversky & Kahneman, 1974) and from use of simple cue-based inferences (Chaiken, 1980). All these factors could be responsible for the observed biases in inflation judgments.

In order to explain the large discrepancies observed between the German consumer price index and perceived inflation in the years after the introduction of Euro notes and coins, Brachinger (2008) developed a novel instrument: the index of perceived inflation (IPI). This has a psychological foundation by including the encoding of price changes relative to reference prices, asymmetries in evaluations related to loss aversion, and the weighting price changes based on use of the availability heuristic (i.e., related to purchasing frequency). As Brachinger (2008) commented, it “provides a psychologically, economically and statistically well-founded instrument that gives an explanation for this phenomenon [i.e., the discrepancy mentioned above]”. Jungermann et al. (2007) also experimentally demonstrated that factors including purchasing frequency, loss aversion, product segmentation, and price levels of items predict IPI scores. These two studies assumed that the well-known Weber–Fechner psychophysical law (Fechner, 1860) holds for the perception of price changes; in other words, equal relative changes in a stimulus bring about equal absolute changes in perception and absolute changes in perception are a linear function of relative changes in the stimulus. This implies that perception of price ratios is independent of the price change level and is a linear function of relative price changes. However, more empirical evidence is needed to support this.

**Figure 1.3**

*An intuitive model involved in inflation expectation formation*



### 1.7.2. Mental representations of inflation

Figure 1.3 shows three psychological components of a model of people’s processing of economic information. In this section, I will focus on the first of these, mental representation. Lay processing of inflation-related information is simplified compared with that of experts. Although its parsimony can be regarded as adaptive, it does cause errors in inflation judgments. Bruine de Bruin, Van der Klaauw, et al. (2011) found that 52.2% of people who generated expectations for inflation with “inflation” worded questions had thought about specific prices. However, the rest who did not think about specific prices reported less

extreme and less dispersed inflation expectations. Previous research by Bruine de Bruin, Van der Klaauw, et al. (2010) had suggested that some individuals probably think of general indicators for overall inflation and personal finances rather than examples of specific prices in certain situations. Thus it seems that different people use different approaches to forecasting inflation rates and that, as a result, they have different accuracy levels.

Research on cognitive representations of inflation can be partitioned into two broad categories: (1) cognitive representations of prices and price changes; (2) cognitive representation about inflation-related variables.

Inflation is calculated through the annual price changes across a basket of goods and services while laypeople personally experience those price changes in daily life. Representations of prices and price changes can be used to form representations of overall price changes (inflation) (Ranyard et al., 2008). Comparison processes are assumed to be psychologically crucial for making sense of prices since people react to their psychological perception of prices rather than the nominal values of those prices (Niedrich et al., 2001).

Thus the concept of reference price is crucial for how people formulate inflation judgments. Reference price, as defined by Monroe (1973, 1990), is still a crucial concept because laypeople usually use internal information of prices they pay held in memory. According to reference price theories, consumers evaluate the price of a product by comparing it with other prices in their memory (internal reference prices) or available in the environment (external reference prices) (Jacobson & Obermiller, 1990). Specifically, in terms of the adaptation-level theory Helson (1947, 1964), price is mentally presented as a prototypical reference price for a category, which is calculated as a weighted mean of prices for this category. Exemplars form different types of prototypical representation. They represent specific locations in a price distribution and exemplar-based models have received some support (Ashby & Maddox, 1998; Medin et al., 1984; Janiszewski & Lichtenstein, 1999). The following discussion applies to both reference price and exemplar-based models.

Early research showed that some consumers do not always remember the actual prices last paid for specific goods (Behrend, 1977; Gabor & Granger, 1961). More than half of shoppers could not even remember the price of items they had just put into a shopping cart (Dickson & Sawyer, 1990). Furthermore, though people perceived the current inflation rate through their everyday experiences, they learn only limited numbers of prices and do not have precise mental representations of prices and price changes (Svenson & Nilsson, 1986). Hence, it appears that prices could be mentally represented as a range in memory that relies on learning just a very few prices, as postulated by range theory (Volkman, 1951) and range-frequency theory (Parducci, 1965).

In Parducci's (1965, 1995) range–frequency theory, an attribute value is a weighted sum of its ordinal rank within the immediate context and its interval scale position within the range is

set by the immediate context. In a study of price evaluation, Niedrich et al. (2001) manipulated the mean, range and distribution of prices in their experiments and found range-frequency theory fitted better than range theory since customers not only relied on the range but also the frequencies of prices they encountered. In their other experiment, they further showed that judgments were moderated by how these prices were experienced (processing environment), suggesting that ‘consumers place greater weight on extreme prices anchoring the range for internal reference prices than for external reference prices (p.339)’. Specifically, the range values are more heavily weighted when data are presented in a sequential presentation condition than in a simultaneous presentation condition. In contrast, intermediate values are more heavily weighted when data are presented in a simultaneous presentation condition.

Thus, it appears that individuals remember experienced price information as exemplars and that these memories are tagged with characteristics of their range and frequency. Hence, in order to estimate the general price change (inflation), people need to judge differences in prices (i.e., price changes) that are stored and represented in their memory. By considering variability in the stimuli and noise of the environment, Batchelor (1986) used survey data on eight countries and showed that this process (the perception of inflation) is consistent with signal detection theory (Bank, 1970; Tanner & Swets, 1954; Green & Swets, 1966). Specifically, he showed “the costs of missing inflation rise as the rate of inflation increases (or departs further from moderate or past values)” and “all of the models show a strong positive relationship between the thresholds and the noise variance”, in conformity with the theory.

Once the mental representation of prices is agreed with a degree of consensus, one issue that still needs to be examined is the question format use to elicit inflation judgments. Ranyard et al. (2008) pointed out that the accuracy of inflation judgments may be impaired if prices are encoded differently from how they need to be recalled. This is because conversions between different formats are likely to require cognitive effort. This may explain why certain kinds of question format elicit more accurate information from memory than other question formats.

Suppose that price changes are stored as a range of prices rather than a number and that inflation expectations are produced by sampling from different ranges of prices. The results would then be more accurate if people are asked to report a range or distribution for future levels of inflation than if they were asked to report just a single value. The finding that density forecasting results perform better than point forecasting levels is consistent with this assumption (Bruine de Bruin, Manski et al., 2011). Other types of conversion could be needed in other situations. For example, people may have to transform their memories of experienced general price changes into the unfamiliar concept of “inflation”: when inflation expectations are required, participants have found that question wording using the term ‘inflation’ was somewhat more difficult to respond to than the question wording using the phrase ‘prices in general’ (Bruine de Bruin et al., 2012; Bruine de Bruin et al., 2017).

The other important feature of the cognitive representation of inflation concerns how the concept of price change is integrated into an overall schema that mentally represents the relations between macroeconomic variables. Although lay people participate in economic activities every day, they have a limited understanding of economics and tend to judge it in an over-simplified manner. Though economic systems involve complex chains of cause and effect, and their behaviours may be cyclic, people tend to construct one-way linear chains when explaining them. Leiser and Shemesh (2018) found that laypeople used short-range reasoning, making single-step inferences from the consequences of inflation to 'price increases' or 'devaluation' while ignoring consequent effects on unemployment, wage and salaries and so on.

To better understand how people comprehend inflation, Svenson and Nilsson (1986) examined lay cognitive representations of inflation among psychology students. In three studies, they found that cognitive representations in this sample were not always congruent with those of economic students. More specifically, they investigated a) the variables that were considered to be related to inflation (obtained via questionnaire responses), b) the directions and strengths of relationships between different economic variables related to inflation (obtained from ratio estimates of importance), and c) the causal chains of variables linking variables to one another (obtained by causal mapping). All participants observed the inflation rate mainly through increases in food prices and rents for housing. The most important factors that were perceived as affecting the inflation rate were 'increase of international inflation rate', 'wage increases' and 'increased prices of raw material and oil'. Some differences in responses between the economic students (experts) and psychology students (non-experts) were reported; for example, 'increased expected inflation rates' was perceived as one of the most important factors only by economics students.

Leiser and Drori (2005) showed that there is indeed a socially shared concept of inflation, consistent with social representation theory (Moscovici, 1981, 1984), but this representation is significantly different from the normative understanding of professionals. A multi-faceted questionnaire was presented to four groups in an effort to identify commonalities and differences in their concept of inflation. Groups were classified on the basis of occupying different positions in the economic world: students in psychology, technical high school students, grocers, and schoolteachers. When participants were asked to generate at least four terms (or concepts) related to inflation, Leiser and Drori (2005) found that, across all participants, the main associations with expected inflation were money (losing its value), price increases, the cost of living (COL) index, and devaluation. However, depth of understanding was found to be widely different across groups. For example, university students and the schoolteachers gave significantly better quality and more detailed explanations than high school children and shopkeepers. The model of inflation held by laypeople was influenced by differences in their social knowledge, and this depended on their age, economic circumstances, cultural background, religion, position, and so forth. This may

partially explain the differences in expected inflation rates that have been found between different demographic groups.

Lay understanding of inflation is regarded as a negative economic phenomenon. That is to say that lay views on inflation are typically seen as a problem (Leiser & Aroch, 2009) with various undesirable consequences, such as reductions in the value of the currency (Leiser & Drori, 2005). People tend to be averse to inflation because they think that it makes them poorer (Shiller, 1997). As a result of the Good-Begets-Good heuristic, the inflation rate is positively related to other bad economic indicators, such as unemployment. In a long-term experiment with a large data set, Dixon et al. (2014) found that, in contrast to professional economists, laypeople believe that there is a positive relation between inflation rate and unemployment rate. This finding was confirmed in a study by Dräger et al. (2016) that used responses from the Michigan Survey of Consumers.

Other studies have also found that links between inflation and economic activities by consumers are misperceived. Ehrmann et al. (2017) reported that consumer inflation expectations are higher during recessions. Savadori et al. (2001) investigated the mental representation of economic crises in Italy and found that inflation was considered to be the prime symptom of economic crises. Bruine de Bruin, Manski et al.'s (2011) survey found that less financially literate individuals reported a lower level of uncertainty concerning price inflation during the 2008 financial crisis compared with those who had higher financial literacy. Hajdini, Knotek, Leer, et al. (2022) studied the influence of expected inflation on expected income growth. Their experimental results showed it varies systematically with demographical and socio-economic variables, such as income levels. Their study also addressed heterogeneity of inflation expectation formation to some extent and provided some explanation for why people dislike inflation.

Laypeople seem to have various understandings of inflation, so it is hard to find a clear way to predict how people respond to the same event. This was especially true regarding the occurrence of the pandemic from 2019, which had a large influence on economies around world. Inflation expectations are likely to have been affected by this shock. However, there is no consistent evidence about the changes in inflation expectations; several papers have been recently published in this field showing conflicting results. Stanisławska and Paloviita (2021) provided survey-based evidence for how the Covid-19 pandemic affected inflation expectations and revealed that consumers' inflation expectations increased in response to the Covid-19 pandemic. Binder (2020) also reported that the pandemic had contributed to higher inflation expectations, as did Dietrich, Kuester, et al. (2022). However, results in the opposite direction were obtained by Coibion, Gorodnichenko and Weber (2020); people expected the pandemic to lower future inflation. According to Armantier et al. (2020), there was no clear trend in aggregated inflation expectations after the outbreak of the pandemic, and long-term inflation expectations appeared to be as well anchored as at the start of the pandemic (Armantier et al., 2021). Stanisławska & Paloviita's (2021) suggested that the

responses to the pandemic that they reported could be related to previous findings showing that people misunderstand the correlation between inflation and economic activities. Armantier, Boumahdi et al.'s (2022) analysis of the New York Fed's Survey of Consumer Expectations showed that consumers showed more uncertainty about near-horizon inflation expectations during the pandemic than they did in the pre-pandemic period. The raised uncertainty about future inflation was more pronounced for consumers than for professionals (Dietrich, Kuester, et al., 2022). These results show that it is hard to predict people's judgment and behaviours without knowing how they (mis)understand economics and react to substantial changes in the economic environment.

### *1.7.3. Perception, evaluation and updating*

In this section, I deal with the second of the three components shown in Figure 1.3. Mental processes concerned with the perception of price exemplars and other descriptive information are important because people perceive price information every day and absorb other types of inflation information occasionally. Laypeople use new information to update their inflation expectations. Perception is the first stage of this process and will involve evaluation of whether information is relevant to updating.

For laypeople, price-relevant information obtained from daily life is the most direct information used in estimating inflation rates. The mental mechanisms used in processing new price exemplars and detecting price changes can be regarded as fundamental to the process of generating estimates of current and future inflation.

According to Nerlove's (1958) adaptive expectation model, people adjust their expectations by taking account of differences between new information and prior expectations. From a more psychological perspective, Sherif and Hovland's (1961) assimilation-contrast theory implies that new price information is assimilated and integrated into internal reference prices (IRP) once it is judged as belonging to the distribution of IPR. The IPR distribution would be updated using a Bayesian updating rule.

According to Mazumdar et al. (2005, p. 89) and consistent with the above models, consumers update their reference prices "(a) by weighting their existing reference price and the observed prices and (b) by factoring in a price trend observed from prior prices". They also suggest that prior purchasing prices are the strongest determinant of a consumer's IRP and that more recent prices have a greater effect on IRP (a reasonable claim given the recency effects that have been identified by psychologists).

Given the expectation-consistency bias that I discussed above, expected price trends are likely to distort the evaluation of new evidence. This was confirmed by Greitemeyer et al.'s (2005) findings: they showed a causal effect of consumers' expectations on their judgements of price trends. Providing participants evidence that should have disconfirmed their inflation

expectations did not always do so because they may have given more weight to the established trend than to the new evidence because the series was very noisy. In addition, individuals who were presented with new information that was inconsistent with their prior beliefs were sometimes *less* likely to revise their beliefs and, instead, developed more polarized beliefs (Lord et al., 1979; Gentzkow & Shapiro, 2006). Evidence may not change our prior views, but our prior views may change our perception of the current evidence.

Evaluation occurs differently in different contexts (Kahneman and Tversky, 2013a; Tversky & Kahneman, 1992). In terms of prospect theory, the same absolute differences are evaluated differently at the lower nominal values (£10 to £20) and higher nominal values (£500 to £510). More importantly, prospect theory treats subjective gains and losses relative to a reference point. Tversky & Kahneman (1991, p. 1039) point out that “losses and disadvantages have greater impact on preferences than gains and advantages” even though they are of the same objective size. In other words, consumers perceive a loss more strongly than the same amount of gain. As a result, price increases are evaluated as more impactful than the equivalent price reductions (Brachinger, 2008; Jungermann et al., 2007). Similarly, inflation expectations are biased by the negative tone of bad news of price increases (e.g., Dräger et al., 2014).

Assimilation-contrast theory was developed by combining range theory and range-frequency theory for price judgments and emphasises that the assimilation effect is determined by the endpoints of the range (Janiszewski & Lichtenstein, 1999) and the frequency distribution of prices (Niedrich et al., 2001). Decision-by-sampling theory (Stewart et al., 2006) provides plausible psychological mechanisms (sampling and rank-based evaluation) that produce these effects. Price evaluation occurs by taking into account both the immediate context and long-term memory. It produces effects consistent with Parducci’s (1965, 1995) range-frequency theory. Decision-by-sampling theory assumes that a sample of instances is derived from long-term memory and/or from the external environment in relation to the item to be evaluated. The item is evaluated based on the range of prices in that sample and reflects the relative ranked position of the target item within that range. A stochastic sample is drawn from memory or the immediate context because it seems implausible that consumers would recall the entire distribution of experienced prices in their memory when evaluating a target price. Decision-by-sampling theory emphasises that sampling from memory depends on an individual’s experiences. It therefore provides an explanation for the heterogeneity of the inflation expectations (Bruine de Bruin, Manski, et al., 2011; Bruine de Bruin, Van der Klaauw, et al., 2010; Bryan & Venkatu, 2001a).

Janiszewski and Lichtenstein (1999) suggested that individuals may be reluctant to adjust the high point of their price range upwards. In support of this view, Ackerman and Perner (2004) found an asymmetry of reference price change: consumers were less willing to raise than to lower the endpoints of their acceptable range; a single low price can cause range adjustment whereas multiple exposures to higher prices are needed before high endpoints of the acceptable range are increased. This is in line with the loss aversion effect. It is also possible

that accumulated higher prices in memory amplify the impression of prices rising and cause the observed over-estimation of inflation judgments.

Previous studies have documented the role of official statistics and experts' forecasts when people form their inflation perceptions and inflation expectations. However, most research has explored how people perceive and utilize that descriptive information to modify their original inflation judgments based on their personal experience. Besides the updating of purchased prices I mentioned before, this is a main research direction on updating of inflation expectation. For example, Armantier et al. (2016) used online experiments to investigate how consumers' inflation expectations for the next 12 months are affected by new information. They found that respondents, on average, sensibly update their expectations in response to information about SPF forecasts and that they do so in a manner consistent with Bayesian updating. These results support the assumption of *rational inattention* in economics (Sims, 2003; Reis, 2006; Coibion & Gorodnichenko, 2012, 2015). The idea of rational inattention refers that "economic decision makers cannot absorb all available information but can choose which pieces of information to process" (Maćkowiak et al., 2021, p. 2).

In contrast with the rational inattention model, some experiments have produced substantial evidence consistent with *irrational inattention* (Bruine de Bruin, van der Klaauw, et al., 2011; Malmendier & Nagel, 2016). For example, the sticky information model suggested that information is slow to diffuse through the population because agents only probabilistically pay attention to new information (Mankiw & Reis, 2002). Cognitive limitations (irrational inattention) were also identified by Cavallo et al. (2017) who found that individuals irrationally place significant weight on inaccurate sources of information. Economic professionals are supposed to assign all weight to information derived from inflation statistics and ignore the supermarket price information when given both pieces of information. However, when lay people were given these two types of information simultaneously, they implicitly assign as much weight to supermarket prices as they do to inflation statistics. In other words, even when information about inflation statistics was readily available, individuals did not use it efficiently to update their initial beliefs. Additionally, researchers have found biases in thinking about experienced inflation. Bruine de Bruin, van der Klaauw, et al. (2011) showed that participants gave significant weight to specific price changes that they had thought of. Malmendier and Nagel (2016) also showed that participants tend to overweight inflation experienced during their lifetimes. Also, more recent experiences were weighted more by younger people and, as a result, they updated their expectations more than older people.

Findings about the importance of different types of descriptive information in forming inflation judgments formation are not consistent. Armantier et al. (2016) failed to find a significant influence of new information on the updating expectations of food price inflation expectation of the next 12 months after people had been given information about the *previous 12 months'* average food price inflation. But in the SPF condition, where information of SPF *forecasts* of the next 12 months was offered, they did update their inflation

expectations. In contrast, Cavallo et al. (2017) did find that provision of information about previous inflation based on either supermarket prices or published statistics significantly influenced inflation expectations.

Closer inspection suggests that these conflicting results may have arisen because of an effect of the format in which price change information was presented. In their study, Armantier et al. (2016) provided statistical information as a single number as follows: "According to the most recent data available from the Bureau of Labor Statistics, the average prices of food and beverages in the US INCREASED by 1.39% over the last twelve months". This contrasts Cavallo et al. (2017) who provided a table with the percent price changes of six products and the average change. It seems that, in the latter case, people's expectations were influenced by revising the self-experienced price changes; in contrast, provision of only an official statistical figure in the former case had no effect. The findings from the two studies can be reconciled if we assume that people put more weight on the information derived from self-experienced price changes and professional forecasts than on official inflation rate statistics. Hence, even though the previous inflation rate was provided in the food condition of Armantier et al. (2016), it was not enough to override the effect of self-remembered price changes.

The difference in updating inflation perceptions when receiving descriptive information can be partly attributed to demographical factors. Armantier et al. (2013) found that respondents who were female, low-income, low-education, and low-financial literacy were more responsive to the information treatments relative to others. Armantier et al. (2016) found that women were more responsive to SPF information than men. Madeira and Zafar (2015) modelled expectation updating using MSC data and found significant demographic differences in updating speed of inflation expectations in their lifetime but that were inconsistent with Armantier et al.'s (2013, 2016) results: women, blacks, Hispanics, lower income, and less educated agents were slower to respond to recent movements in inflation to update their expectations. This may explain why higher inaccuracy and high heterogeneity in inflation expectations were observed in these groups and why they demonstrated more responsiveness to new information in experimental conditions.

In general, providing additional information (either price exemplars or statistics) to laypeople does not guarantee that their expectations will be more accurate. The type of information that they are given matters: people weight information from different sources differently. The reason that they do this may depend on affective factors. As Katona noted, 'rarely do many people mention both favourable and unfavourable business news at the same time. According to whether they feel that business conditions are improving or deteriorating, only good news or only bad news is salient to them' (1975, p. 200). Also, individuals infer a higher quality (credibility) of a news source if its content conforms to the consumer's prior expectations (Gentzkow & Shapiro, 2006).

#### 1.7.4. Retrieval

This section refers to the third components in Figure 1.3, which is the retrieval process. The retrieval process is related to two cognitive components: the intuitive system involving long term memory (LTM) and the reasoning system demanding resources from working memory (WM). Their limitations have been summarized by Leiser and Shemesh (2018): they suggest that the narrow scope of LTM and the short range of WM cause unreliable retrieval of information as well as some difficulty in handling a large number of items. In addition, these two limitations interact with each other. In other words, pieces of information stored in LTM cannot be related or made compatible because LTM is only minimally able to specify the relevance and usefulness of cues. Once people engage in conscious cognitive processing, information recalled from LTM is transferred to WM. However, because of the limited capacity of WM, only small amounts of information can be processed and harmonized. Limitations in both types of memory process have a significant impact on final judgments.

The two components are congruent with the two modes of cognitive functioning (Sloman, 1996): intuition and deliberation, respectively. They are also in line with the two thinking systems proposed by Kahneman (2011). System 1 is characterized as associative, effortless and intuitive, whereas System 2 is defined as deliberative and analytic. Research has revealed that laypeople typically use System 1 to answer inference questions in the economic domain (Ziv & Leiser, 2013). Although System 1 seems lazy, fast and prone to error, it is efficient and useful in certain circumstances (Gilovich et al., 2002). In addition, heuristic processes are not suitable for dealing with complicated questions but can cope with easy questions about likelihood, frequency and prediction. System 2 is activated when casual relationships are framed clearly in laypeople's minds and causal information is given or successfully recalled. In conclusion, when judgments are required, people do not recall all information they possess. The process is simplified so that judgment will be made once just enough information comes to mind.

When people rely only on information about their previous price experiences, they can only make inflation judgments based on what they remember about those prices. Monroe and Lee (1999) suggested that recalled price information provides a basis for recognising whether a newly experienced price is of the expected magnitude. In addition, recalling price information included two types of memory: explicit memory of episodes involving price exposure (characterized as 'remembering') and implicit memory of non-conscious retrieval of previously encountered stimuli (characterized as 'knowing'). The implicit memory processing further includes conceptually driven implicit memory (semantically related) and perceptually driven implicit memory (physical resemblance). In situations involving implicit memory, a customer cannot remember specific prices/scenarios to judge whether a new product is expensive or cheap but they can rely on their underlying knowledge, which is the basis of their internal reference price(s) for the category.

As for inflation judgments, the remembering process should be the mental comparison of two or more episodes of the same product over time whereas the knowing process would be the comparisons of IRPs. Thus, some individuals recall specific prices while others employ their underlying knowledge of the prices ranges for different categories. Monroe and Lee (1999) also argued that some price information processing is more automatic than others. For example, lower-price range comparisons (typical for daily supermarket purchases) may be more likely to be processed automatically since lower prices involve fewer digits. These conclusions are consistent with the results of Bruine de Bruin, Van der Klaauw, et al. (2011) that half of their participants thought of food prices more than other prices when making inflation judgments.

What kind of price information will be recalled by people who have a huge number of alternatives in their long-term memory? Memory-based heuristics are likely to be important. The availability heuristic relies on memory retrieval processes; judgments depend on how easy it is to retrieve information or exemplars (Tversky & Kahneman, 1973). Ranyard et al. (2008) identified the important role of the availability heuristic in inflation judgements. Specifically, they drew attention to four features that influence the availability of price changes: recency of purchase, size of price change, direction of change, and the frequency of purchase.

It is reasonable to assume that the more recently an item has been bought, the easier it will be recalled due to familiarity. Inflation judgments would then rely heavily on that information. As Mazumdar et al. (2005) noted, when judging the overall price level charged by a store, the more familiar product categories of the store will be sampled from memory and given greater weight in judgments of the overall price level. However, recency of purchase as a determinant of inflation judgments has not been experimentally investigated so far.

In a priming experiment, Del Missier et al. (2008) demonstrated that product availability influences perceived inflation. Participants were asked to name products that either increased in price or decreased in price over the last year. Then they reported the perceived inflation. Results showed an assimilation effect: focusing on the increased prices led to higher inflation perception while priming with the decreased prices led to reports of lower inflation judgments. This result was confirmed by Del Missier et al. (2016): product priming consistently affected inflation judgments in the direction of an assimilation effect.

Judgments of the size of price changes use the availability heuristic because of the salience of the information. People have the tendency to pay attention to dramatic and unexpected changes. Price rises are more strongly perceived than price reductions (Fluch & Stix, 2005). In addition, it has been shown that, when people make predictions, they tend to incorporate their past experiences with the more extreme experiences being more likely to come to mind; as a result, extreme forecasts are produced (Morewedge et al., 2005).

In the same vein, in another study (Bruine de Bruin, Van der Klaauw, et al., 2011), it was found that participants who were instructed to recall “any” price changes reported the extreme price changes which were identical to answers obtained from participants who were asked to recall “the largest” price changes. In addition, these two groups of forecasters forecasted more extreme one-year-ahead inflation compared to another group who were required to recall “average” price changes. This implies that people are prone to recall extreme price changes even though they are not instructed to do so, suggesting the importance large price changes in forming inflation judgments. This is consistent with the results showing that some respondents of inflation surveys generate relatively extreme inflation expectations (Bruine de Bruin, Van der Klaauw, et al., 2011; Bryan & Venkatu, 2001a; Curtin, 2006).

Research has shown that different directions of price change have different roles on inflation judgments. For example, participants who were asked to recall any price change focused more on extreme increases in prices rather than on extreme decreases in prices and so reported much higher price changes in the previous year than group who were required to recall “average” price changes (Bruine de Bruin, Van der Klaauw, et al., 2011). If people pay equal attention to increases and decreases in price, both types of information should be recalled and be equally influential. The asymmetry response to news discussed before also confirms the effect of direction of price changes.

The frequency effect can also be regarded as a phenomenon based on use of the availability heuristic (Brachinger, 2008; Georganas et al., 2014; Huber, 2011; Jungermann et al., 2007; Del Giovane & Sabbatini, 2008). Jungermann et al. (2007) showed experimentally that perceived inflation is related to frequency of product purchase, with more frequently bought products contributing more to the judgments. Huber (2011) confirmed this by direct manipulation of the frequency of products with different prices. He found inflation judgments were accurate in a condition using expensive products with low-frequency presentation, but were over-estimates in a condition using cheap products with high-frequency presentation. The total expenditure changes were identical in two conditions. In addition, Del Giovane and Sabbatini (2008) attributed excessively high perceptions of inflation after the Euro changeover to the fact that frequently purchased goods happened to show higher levels of inflation during that period.

Georganas et al. (2014) measured perceived inflation in a simulated economy by manipulating price changes of goods so that they were either flat, large and positive, or large and negative. To mimic key aspects of actual consumer purchases for the different goods with different frequencies in a simulated shopping experience over the 96 days of trials, participants were told which type of goods to purchase (labelled abstractly as goods A, B, C, and D) and were asked to select the cheapest price for that good each day. Then they were asked to give estimates for the inflation rate for a basket of the goods and also for each type of good separately. They demonstrated that consumers' perceptions of aggregate inflation were

systematically biased toward the perceived inflation rates of the most frequently purchased items, such as food and gasoline.

To overcome the effects of the availability heuristic, more samples could be retrieved from memory so that consumers are more likely to generate more accurate estimates of current and future inflation. Given the 'wisdom of the inner crowd' effect (Herzog & Hertwig, 2009; Herzog & Hertwig, 2014), it is likely that, if a question is asked multiple times, more accurate judgments could be obtained by averaging the answers. The 'wisdom of the inner crowd' effect has been demonstrated not only for judgments about future events, but also for responses to spatial knowledge questions (Montgomery & Lee, 2022).

Why does thinking about specific prices lead to more extreme inflation expectations? One explanation is that people realize that salient price changes are more extreme than overall inflation but fail to make sufficient adjustments downwards when assessing overall inflation (Ranyard et al., 2008). In other words, they make use of the anchor-and-adjustment heuristic (Tversky & Kahneman, 1974). This means that, when forming overall inflation expectations, focusing on specific price changes would lead to more extreme inflation expectations and more disagreement between respondents.

Use of heuristics often produces biases because of the way in which those heuristics make selective use of information in memory. However, memory itself could be impaired in ways that damage inflation judgments. Recall of specific prices may contaminate the accuracy of estimates of past prices and future expectations. Kemp (1991) and Kemp and Willets (1996) found prices recalled via long-term memory from a remote time (e.g., 15 years ago) were overestimates and thus resulted in the underestimation of inflation. In another experiment, Kemp (1987) showed that people do not really 'remember' the past prices of items, instead, they employed a general model of how prices change over time to estimate the past costs of specific items as well as the cost in general. This could be demonstrated by the positive correlations across participants between the size of the price increase on one item and the size of the price increase on the average cost of all items. Hence, the price changes for disparate items were reported as *more* similar than their actual price changes across items. This suggests that memories of past prices for different items might be stored in an amalgamated form, and thus the prices of one item can influence the recall of other items' past prices. Kemp (1987) further found that increased familiarity with specific prices resulted in less accurate estimates of past prices for the item in question. Kemp reconciled his findings using the principle of (retroactive) interference so that recent prices of a particular item interfere with (or even replace) memory for more distant prices of the same item. This effect was strongest for individuals with the greatest experience with recent prices. Thus, memory was affected by the recall process and by the integration of information within memory.

In most studies, descriptive information is provided, so that people do not need to recall it from memory. One exception is provided by Roos and Schmidt (2012). They examined how

people forecast inflation by utilizing information provided as a historical time-series and by utilizing knowledge acquired outside the experimental setting (knowledge of expert forecasts). In the full information group (FI), participants received historical inflation data charts whereas, in the no information group (NI), they received no information. All participants were asked to provide estimates of expert expectations. Self-reported use of information showed that about one-quarter of participants reported that they had used expert information acquired outside the experimental setting, and there was no significant difference in this percentage between the FI group (19%) and NI group (25%). As expected, the predictions in the NI group correlated with the recalled expert predictions. However, there was no significant correlation between estimated expert predictions and participants' own forecasts in the FI group that had the time-series chart available. This was so even though those in the FI group had reported that they had used expert predictions to form their forecasts. For the NI group, people's inflation expectations seem to be anchored on recalled expert predictions. However, the inflation expectations made by the FI group were de-anchored from their recalled experts' forecasts by the provision of the historical inflation data.

Other research that may be relevant to retrieval processes indirectly investigated relationships between inflation news coverage and inflation forecasts. Carroll (2003) found that people adopted the expectations of experts published in the news media to update their own expectations and Lamla and Lein (2014) found that the volume and tone of new media biased inflation perceptions and expectations. Loss aversion has also been documented with respect to inflation perceptions and expectations (Dräger et al., 2011; Soroka, 2006).

In summary, people are able to recall information from their memory when generating inflation expectations although they make some mistakes and suffer from some biases. However, these effects were diminished when official statistical information concerning historical data was provided in the immediate context.

## **1.8. Decision making**

In modern macroeconomics, inflation expectations are considered to be a crucial determinant of agents' decisions. In equilibrium, the public's behaviour determines actual inflation, thus establishing a direct link between expected and realized inflation. The importance of accurate measurements of agents' subjective expectations has been increasingly recognized in economics because it is important for economic decision making (Armantier, et al., 2013; Armantier et al., 2015; Manski, 2004). However, research has revealed a contradictory picture about whether the inflation judgments actually influence their financial behaviours.

Despite imperfect price perceptions and limited cognitive capacities, consumers' inflation expectations inform consumers reported financial behaviours (Bruine de Bruin, Manski, et al., 2011) and actual financial behaviours (Armantier, et al., 2013; Armantier et al., 2015; Coibion et al., 2019). This provides direct empirical evidence for the basic assumptions underlying

most macroeconomic models. For instance, Amantier et al. (2015) compared the reported inflation expectations of 502 consumers in a survey and their financial behaviours in a financially incentivized investment experiment. Results confirmed that inflation expectations were a key determinant of their financial behaviours. Indeed, stated beliefs and experimental decisions were found to be highly correlated and consistent, and respondents who changed their inflation expectations from one survey to the next also tended to adjust their decisions in a consistent way (both in direction and magnitude). Bruine de Bruin, Manski, et al. (2011) found that people's reported uncertainty about future price inflation was significantly positively correlated with their reporting shorter planning horizons for household financial decisions and perceiving less responsibility for household investment decisions. D'Acunto et al. (2018) and D'Acunto, Hoang, et al. (2022) also found strong evidence that by implementing unconventional policies (such as announcements of future increases in consumption taxes), consumers who generated higher inflation expectations demonstrated a higher propensity to purchase durable goods.

However, Galashin et al. (2020) found that changes in macroeconomic expectations were not translated into changes in actual consumption behaviour. Although information provision effectively shifted expectations of inflation and exchange rates, planned and actual spending behaviour did not show a consistent change in accordance with their expectations (as measured in survey data and administrative data, respectively). This implies that individuals failed to react to their macroeconomic expectations.

The discrepancy between inflation expectations and actual financial behaviours could be explained by lack of knowledge (Beshears et al., 2018). People who do not fully understand economy failed to factor their macroeconomic expectations into their spending decisions. Similarly, Amantier et al. (2015) showed that consumers with low numeracy and low financial literacy were less likely to act on their reported beliefs in accordance with expected utility theory. This work indicates the importance of lay people acquiring economic knowledge if they are to make wise financial decisions.

In summary, there are inconsistent findings about the reason why lay people form biased and heterogeneous inflation expectations and how they affect their following decisions. In addition, there is limited evidence addressing the underlying psychological mechanism. Therefore, in my first two experimental chapters, I investigated potential factors from the perspective of psychology and mainly focused on: (1) how and when lay people use experiential information (price changes), and (2) how lay people utilize descriptive information (historical data of inflation rate, unemployment rate and interest rate) if they are provided in the survey.

Most of the studies relating to inflation expectations used the survey-based measure of inflation expectation. A relatively novel approach is the use of laboratory-based experiments. In a recently experimental work, Rötheli (2020) experimentally developed a "pattern

extrapolation” model to illustrate the formation of inflation expectations. He brings elements of ‘similarity matching’ and ‘scaling relationship” into the computation of pattern-based expectations and validated his models by comparing expectations based on extrapolation from patterns with expectations measured by the MSC. From this work, he concluded that consumers form inflation expectations based on pattern extrapolation. However, the model is not complete in terms of the different information resources that people can draw on each day. In addition, it is not convincing because consumers had no access to these patterns of past inflation rates recorded in the MSC. So, one critical question concerns consumers’ ability to remember the past inflation data accurately and then generate inflation expectations relying on pattern extrapolation.

Even though Rötheli’s (2020) model does not capture the expectation formation process completely (e.g., the role of media is not included), his work is of interest here because it highlights the relevance of empirical and theoretical research into the broader area of judgmental forecasting. This work deals with factors that influence the accuracy of judgmental forecasts in any domain, the reasons both organizational and psychological forecast errors, and approaches that can be used to improve their quality. Hence, in the following chapter, I will discuss experimental studies into the processes underlying judgmental forecasting more generally (i.e., not just concerned with the forecasting of economic variables).

## Chapter 2 Review of judgmental forecasting

First I provide a brief outline of the role that judgment plays in the forecasting process and its involvement in economic forecasting. I then outline the three judgment heuristics that have been implicated in the forecasting process before going on to review the literature in more detail.

### 2.1. Background

Forecasting under uncertainty is essential in daily life as changes in forecasts will change subsequent behaviours. As an important aspect of human behaviour, forecasting can be accomplished by statistics and/or judgment. Judgment can be used unaided to make forecasts, combined with statistical forecasts, or employed to adjust statistical forecasts to take account of contextual factors. Even when a primarily statistical approach is chosen, judgment is still involved in the forecasting process. For example, forecasters need to judge which set of data to include, decide which variables to consider or incorporate, and choose which statistical methods to utilize in order to conduct subsequent decisions.

Judgmental forecasting is still widely used in practice even though the statistical approaches are becoming more popular (Webby & O'Connor, 1996; Fildes & Goodwin, 2007a; Fildes & Petropoulos, 2015). In a survey study (Fildes & Petropoulos, 2015), only about 29% of companies indicated that they exclusively used statistical methods for forecasting; the remaining companies employed judgment to adjust statistical forecasts, combined judgment with statistical forecasts, or relied purely on judgmental forecasts. Statistical forecasting has the advantage of using a large amount of data simultaneously. Its objectivity also benefits forecasting accuracy because the same algorithm used on the same data always produces the same result. However, statistical forecasts are not guaranteed to be perfectly accurate. Patterns in past data that statistical techniques can exploit may not be clear: the quality of statistic forecasts heavily relies on the quality of data. Also the future may not continue with the same data patterns as the past. Though more recent data are more important than less recent data for short-term forecasting (e.g., Lau & Bearden, 2013; Rötheli, 2020), the reality is that the future does not always follow past patterns, a factor that is especially important in a medium- and long- term forecasting (such as in the evolution of technology). Therefore, to make accurate forecasts, human judgment is necessary to monitor contingencies and systemic changes in the environment and to deal with them.

Experienced practitioners can utilize up-to-date information and employ both their explicit and implicit knowledge to make their forecasts. Thus human judgment can incorporate some information that cannot be identified by statistical methods. It is, however, undeniable that judgmental forecasts can be biased so as to damage forecast accuracy (Armstrong, 1985; Hogarth, 1987). Limitations of judgmental forecasting have been documented for a long time,

termed as “conventional wisdom” (Makridakis et al., 1998). These biases include optimism, wishful thinking, lack of consistency, confirmation bias, hindsight bias and so on.

## **2.2. Economic forecasting**

Economic forecasting is of fundamental concern to countries, businesses, organizations and individuals. There have been debates about whether economic forecasts generated by models or by pure judgments are better. Recently, research on macroeconomic forecasting has focussed on the role of judgment in forecast adjustment and forecast combination (Fildes & Stekler, 2002; McNees, 1990; Turner, 1990; Davydenko & Fildes, 2013). This is because econometricians employ quantitative techniques and then use their judgment to modify their models or final forecasts. As Den Butter (1992) emphasized, macroeconomic forecasts are a mixture of scientific knowledge (models) and judgment.

Models can provide a basis for macroeconomic forecasts, but judgment may still improve accuracy. Goodwin, Önköl, et al. (2011) outlined the crucial role of (expert) judgment in economic forecasting beyond the advantages of quantitative models. They pointed out that judgment is involved in economic forecasting in three ways: (1) formulation of models (e.g., Bolger & Harvey, 1998), (2) revision of model components (e.g., Donihue, 1993; McNees, 1990; Bunn, 1992) and (3) formation of final forecasts (through judgmental adjustment) (Lamont, 2002).

It has generally been accepted that the use of judgmental adjustment improves the quality of model forecasts (Fildes & Stekler, 2002). For example, Clements (1995) found that judgmental adjustment reduced the variation over time in macroeconomic forecasts. Franses et al. (2011) showed that all forecasts created by the Netherlands Bureau of Economic Policy Analysis (CPB) were judgmentally adjusted model-based forecasts and those were, in general, more accurate than the model-based forecasts. It has also been found that adjusted forecasts for short-term interest rates, consumer prices, and residential investment typically outperform mechanical forecasts. However, they are less helpful for longer horizon forecasts and may even impair forecasts for business inventories and real imports (McNees, 1990).

Inadequate judgmental performance has been found in some other studies. Batchelor and Dua (1990), found persistent biases in individual economic forecasters, in which those forecasters showed consistent optimism or pessimism in their forecasts of macroeconomic variables, including growth in real gross national product (RGNP), inflation in the GNP deflator, unemployment rate, and levels of short-term interest rates (BILL) and long-term interest rates (BOND). They suggested the accuracy of macroeconomic forecasts could be improved by pooling forecasts from a number of sources to cancel out biases to optimism and pessimism. This would also cancel out random noise.

Fildes and Stekler (2002) reviewed the predictions of short-run real GDP and inflation forecasts for the US and the UK. Professional forecasters made systematic errors and their forecast accuracy did not improve over time. For example, there was under-prediction of inflation when it was accelerating and over-prediction when it was decelerating (Zarnowitz & Braun, 1993). Evidence from the UK and other European countries was generally in accord with US evidence. Additionally, accuracy of predictions for GNP growth and inflation improved as the length of the prediction horizon decreased.

In fact, in the context of economic forecasts, there is still a large group of people (including lay people) who make predictions purely by judgment without input from quantitative models.

### **2.3. Heuristics and Biases**

Judgment plays an important role in forecasting. However, it is inevitable that judgment can be subject to biases and inefficiencies. Biases in judgmental forecasting have been found since the 1930s. For example, Ogburn (1934) showed that sportswriters exhibited a conservatism bias and underestimated football scores. McGregor (1938) found an optimism bias, such that individuals forecast an outcome consistent with their preferences. Other studies have also shown that wishful thinking produces errors in judgment. For example, people assign probabilities to the occurrence of desired events that are too high (Zakay, 1983; Babad, 1987; Harvey, 1992). Bolger and Harvey (1998) outlined how heuristics and biases influence different forecasting stages: problem structuring (selective attention, framing effects, anchoring and so on), information collection (effort/accuracy trade-offs), choice of forecasting strategy (non-compensatory strategies and non-consequentialism), strategy application (representativeness, availability, anchoring-and-adjustment, framing effects and overconfidence) and evaluation of forecasts (hindsight bias, confirmation bias and the sunk-cost effect).

Biases can be motivational or strategic: Batchelor (2007) points out that they may be 'rational' in nature. For example, forecasters tend to differentiate their forecasts from those of other forecasters to gain publicity, a feature termed contrarianism (Ashiya, 2009; Cipriani & Guarino, 2008). In contrast, forecasters may have a propensity to follow the "market forecast" so as to avoid suffering significant forecast errors and consequent losses. The resulting herding of forecasts (Pierdzioch et al., 2016) is more common in high volatility environment and among inexperienced forecasters (Hong et al., 2000), which was consistent with herding theories that are based on people's motivations being focussed on their concern for their careers. Furthermore, forecasters may be reluctant to make revisions because of a concern that doing so will reduce their credibility (Ehrbeck & Waldman, 1996; Kirchgässner & Müller, 2006). Optimism represents another bias of this type. It is found in analysts' earnings forecasts (Schipper, 1991; Easterwood & Nutt, 1999), possibly because of a need to maintain good relationships with management (Francis & Philbrick, 1993). Mello (2009) described how forecasters use game-playing strategies when sales forecasting in attempts to reach their own

personal objectives. As Goodwin (2017) has argued, behaviours like these do not aim to improve forecast accuracy and can hardly be considered to be proper forecasting; instead, they are aimed at attaining personal targets.

Other biases arise from the heuristics that forecasters use. Heuristics are employed to offset the limited capacity characteristic of human information processing (Harvey, 1998; Önkâl-Atay et al., 2002; DellaVigna, 2009). Bolger and Harvey (1998) pointed out that people rely on simplification strategies (heuristics) to reduce cognitive effort. This has been characterised as cold (i.e., non-emotional) cognitive error (Abelson, 1963). The use of heuristics usually produces satisfactory but rarely optimal judgments and decisions (Simon, 1956).

Tversky and Kahneman (1974) identified various heuristics affecting judgment and decision making. They argued that these mental processes represent cognitive shortcuts (simple rules) that enable people to respond quickly and automatically and that their use challenges assumptions that humans are rational agents. Their original work focusses on three types of heuristic that cause systematic and predictable errors and biases: representativeness, availability and anchoring- and-adjustment.

### *2.3.1. Representativeness*

The representativeness heuristics may be used when people are asked probability questions. Specific events are regarded as dominant and representative of all possible events of a given type. The heuristic ignores base rate and sample size, and so violates the normative statistical theory, Bayes' theorem and regression toward to mean. Use of this heuristic produces outcomes (judgments and predictions) that are always most representative of the input information and, typically, they are made with unwarranted confidence (the illusion of validity) (Tversky & Kahneman, 1974).

Kahneman and Tversky (1973) suggested that redundant input information can trigger use of representativeness and so impair forecast accuracy while boosting confidence. Johnson (1983) asked students to make judgmental predictions of corporate bankruptcy based on two types of information: (1) base rate probabilities of the bankruptcies, and (2) financial profiles of each company. Use of the representativeness heuristic was triggered by the higher salience of the financial profiles; these were considered representative of company performance and, as a result, bankruptcy base rates were ignored. If a financial profile was considered not representative enough, then the base rates were taken into account but were underweighted.

Use of representativeness can also lead to biases in time-series forecasting. People may have the illusion of seeing patterns in noisy series and represent those 'patterns' in their forecast sequences (Eggleton, 1982; Goodwin & Fildes, 1999), a tendency that can be exaggerated by provision of narrative explanations (Fildes & Goodwin, 2007b, 2021). Of course, the 'patterns' that people perceive do not exist – people are representing noise in data series in their

forecast sequences. As a result, the noise in a sequence of forecasts is proportional to the noise in the time series (Harvey, 1995; Harvey, Ewart et al., 2007).

### *2.3.2. Availability*

Use of the availability heuristic means that people judge the probability of an event according to the ease with which that event comes to mind, including the ease of retrieval, the ease of construction and the ease of association (Tversky & Kahneman, 1974). Availability is usually an effective way of assessing frequency and probability because high-frequency instances are more easily recalled compared to low-frequency instances. This is because they have been encountered more often and encountered more recently, two factors that facilitate recall.

Salience may also contribute to the retrievability of information about events. Hence, people are likely to overestimate the probability of events that they find highly salient. For example, Krause (2006) found that this occurred when US federal agencies made macroeconomic forecasts for highly desired or highly undesired events (e.g., extreme economic booms or extreme slumps). Forecasts were overestimates in the former case but underestimates in the latter one. News can influence operation of the availability heuristic because unusual or important events receive more exposure in the media and therefore result in biased judgment (Barber & Odean, 2008).

Forecasters may also employ easily recalled co-occurring information as indicators for forecasting. Illusory correlations may arise as a consequence of using the availability heuristics in this situation; people may falsely believe that two variables are associated because of this biased recall of supportive evidence (Tversky & Kahneman 1974; Goodwin, Önkal, et al., 2011).

### *2.3.3. Anchoring*

Use of anchoring-and-adjustment heuristics occurs when people select a mental anchor and then adjust away from it to make their judgment (Slovic & Lichtenstein, 1971; Tversky & Kahneman, 1974). It results in biased judgments because adjustments are typically insufficient. This may be because people stop adjusting soon after reaching a satisfactory or adequate value. Anchoring and adjustment occurs when people make judgmental extrapolations from time series. Forecasters anchor their judgments on the most recent value (Bolger & Harvey, 1993), the long-term mean (Lawrence & O'Connor, 1992) or the trend line (Bolger & Harvey, 1998) and then make an insufficient adjustment.

Anchoring also provides a partial explanation for poor calibration in interval probability judgments because people start from their best guesses and adjust them to the ends of possible range (Tversky & Kahneman, 1974; Juslin et al., 1999). The tendency to use the anchoring and adjustment heuristic applies to both lay forecasts and professional forecasters (Campbell & Sharpe, 2009).

Factors that influence the anchoring and adjustment heuristic have been identified (for a review, see Furnham & Boo, 2011), including cultural differences (Czerwonka, 2017), thinking style (Cheek & Norem, 2017), cognitive ability (Bergman et al., 2010), and personality (Eroglu & Croxton, 2010). Characteristics of data series may also affect use of this heuristic. Harvey et al. (1997) manipulated four different noise levels of trended series and suggested that noise “renders the anchor-and-adjust heuristics underlying them less effective” (p. 111).

Use of particular heuristics depends on specific situations. The heuristic that is automatically applied relies on how information concerning the predictor and outcome is presented (Czaczkes & Ganzach, 1996; Amir & Ganzach, 1998). For example, a salient anchor can increase the use of anchoring and adjustment heuristics whereas compatibility between predictor and outcome may induce representativeness. Within the forecasting context, Harvey (2007) proposed that selection of heuristics depends on the information available: representativeness is used when the value of the variable to be forecast is given with explicit information about the value of another variable; availability is used when forecasts are based on information held in memory; and anchoring-and-adjustment is employed when the variable is forecasted based on its historical values.

In the following sections, I discuss literature associated with judgmental forecasting under various task conditions. First, forecasting when only time-series data is available. Second, time-series forecasting with various types of additional information: contextual information, domain knowledge (casual information, cross-sectional), technical knowledge and so on. Third, forecasting that takes uncertainty into account: interval forecasting and probability density forecasting. Lastly, forecasting under conditions that allow for improvements in performance.

#### **2.4. Judgmental time series forecasting**

Forecasting can be categorized into two types. Pure time-series forecasting employs data solely from the variable to be forecast. This can be contrasted with forecasting using multiple variables in addition to the variable to be forecast (e.g., econometric forecasting).

When only historical time series is included for judgmental predictions, judgmental forecasting can be modelled as exponential smoothing, or alternatively, as anchor and adjustment, where the anchor point is the average of recent time series values and the adjustment is the proportion of deviation of the most recent value from this average (Andreassen & Kraus, 1990; Harvey, 1988; Lawrence & O'Connor, 1992). This is still true for forecasts from cyclic data series (Harvey et al., 1994): forecasts for the next data point were made using a simple extrapolative heuristic employing the last data point as an anchor and the adjusting from it towards the trend line. People appear generally good at judgmental extrapolative forecasting: judgmental time series forecasting by novices has been found to be

not significantly worse than that of experts with some knowledge of forecasting (Lawrence et al., 1985; Sanders & Ritzman, 1992).

However, judgmental forecasting from time series may go beyond mere extrapolation; people may use additional components to produce their predictions. For example, according to “the major error of intuitive prediction” (Kahneman & Tversky, 1982, p. 416), people have the tendency to overweight specific circumstances and place too little weight on models. Thus, when forecasters ignore the quality of the anchor and make excessive adjustments to their forecasts on the basis of extra-series information, forecast accuracy is typically worse than when fewer or smaller adjustments are made (Lawrence & O'Connor, 1995). Similar results were reported by McNeis (1990) who showed that making large adjustments to statistical models impaired accuracy more than too timid adjustments. People may also mistakenly identify random fluctuations in time series as signals and respond to them (O'Connor et al., 1993); such overreaction to recent randomness in historical time series produces biased and inaccurate forecasts (Lawrence et al., 2000). All these studies demonstrate systematic deviation of human judgment from simple extrapolation.

In the following sections, I will discuss some factors that affects the accuracy of judgmental time series forecasting.

#### *2.4.1. Ecological validity*

In general, people are good at making judgmental forecasts for tasks that have high ecological validity, but they perform worse than statistical models when forecasting for artificial time series that do not reflect their natural ecology. Thus, in artificial and well-structured time-series tasks, judgmental forecasts are typically less accurate than statistical models (e.g., Lawrence & O'Connor, 1992). For example, O'Connor et al., (1993) examined how well people forecast time series that contained changes (major discontinuities and randomness). Even though we might expect people to be good at perceiving changes in their environment, judgmental forecasts were worse overall than statistical models based on Single Exponential Smoothing (SES) and the Adaptive Response Rate Method (ARR). The authors argued that people were less accurate than the statistical models because of forecasters' overreaction to each new value. However, it could also have been because the artificial series used in the study included discontinuities not typical of real environmental ones.

In complex real business and economic forecasting series, judgments may be more accurate or at least equal to the model's performance (Lawrence, 1983; Makridakis & Hibon, 1979). Using a sample of 111 real time series, Lawrence et al. (1985) found that naive forecasters with no knowledge of the nature of the time series were able to forecast as accurately as the best statistical models, and displayed a lower standard deviation of forecast errors. It may be that they were able to incorporate certain types of information from outside the time-series into their forecasts (Edmundson et al., 1988).

There are exceptions to this broad generalization. Using a sample of 10 series from the M1 competition, Carbone and Gorr (1985) reported that initial judgmental forecasts were less accurate than their later statistical forecasts using different methods in a within-subject design. Sanders and Ritzman (1992) found that judgmental extrapolation outperformed or equalled statistical projections for more unstable or more volatile real series while statistical forecasts did better than judges for time series that had less variability.

#### *2.4.2. Trend*

Trends are a common feature of time series. Regarding the different types of trends, there is a consistent finding of trend damping (Eggleton, 1982; Lawrence & Makridakis, 1989; O'Connor et al., 1997), which means people underestimate future values for upward trends time series and overestimate them for downward ones. Although some researchers initially attribute it to anchoring and adjustment heuristic, Harvey and Reimers (2013) provided evidence in support of the adaption to features of the natural environment. The finding that greater damping is observed with noisier series in judgmental forecasting experiments (Eggleton, 1982; Harvey & Bolger, 1996) is in line with this ecological framework.

In addition, researchers have found that people have more difficulties in forecasting downwardly trended series than upwardly trended series (asymmetric damping), that is, trend damping for downward trended series was greater than for the upward trended series. Mean absolute percentage error was higher in down trended series than up trended series (O'Connor et al., 1993; O'Connor et al., 1997). However, it may depend on the context and overall performance was better for negative trends (Thomson et al., 2003). One explanation for this is the artificially generated series were labelled as sales and they could not fall all the time without taking any actions being taken by managers (Lawrence & Makridakis, 1989). O'Connor et al. (1997) further documented this framing effect and suggested that it could be explained using prospect theory (Kahneman & Tversky, 2013a): people tend to be risk-seeking with losses but risk-avoiding for gains. However, the asymmetric effect has also been found when no context is specified (Tumen & Boduroglu, 2022), further supporting the argument of Harvey and Reimers (2013). Theocharis and Harvey (2016) developed one approach to reduce this bias; when several forecast horizons were required when making predictions from time series, eliciting the forecast for the most distant horizon first (end-anchoring) improved forecast accuracy by reducing trend damping.

However, when the trend is more complex, judgmental extrapolations become significantly biased, such as forecasting for series with the component of exponential growth (Timmers & Wagenaar, 1977; Wagenaar & Sagaria, 1975), cyclical features (Harvey et al., 1994) and seasonality (Sanders, 1992).

The exponential growth bias was originated from the well-known fact that humans consistently underestimate exponentials (Andreassen & Kraus, 1990; Wagenaar & Timmers,

1978), even when the data was presented in a graphical format (Wagenaar & Sagaria, 1975; Ciccione et al., 2022), with more data points (Wagenaar & Timmers, 1978) or as a computerized representation of growing plants (Wagenaar & Timmers, 1979). The magnitude of this bias did not correlate with participants' education level (Levy & Tasoff, 2016) but it was slightly reduced after a short lecture on it (Wagenaar & Sagaria, 1975), and could be mitigated by acquiring mathematical knowledge, by using a logarithmic scale, and by presenting a noiseless exponential curve rather than a noisy data plot (Ciccione et al., 2022).

Nevertheless, expertise matters in dealing with complicated series. Experts' extrapolation judgment is superior to that of lay people's and even long-horizon forecasts of statistical models when performing complex forecasting. This is because experts (e.g., currency analysts) have the ability to recognize randomness or to detect trends from noisy data while models tend to identify signals in the series when they are non-existent (Thomson et al., 1997).

#### 2.4.3. Other characteristics of data series

Other characteristics of data series also affect forecasting accuracy. Intuitively, *meaningful labels* should lead to cognitive control that reduces the random error and should benefit the performance (Koele, 1980; Sniezek, 1986). However, they can be detrimental so that forecasts are biased on some occasions. As Bolger and Harvey (1998) suggested, the lack of meaningful labels appears merely to increase the random error in forecasts, whereas there is reason to suspect that meaningful labels might actually bias forecasts. For example, over-forecasting produced by wishful thinking occurs when data were labelled as "sales" or "profits" (Harvey & Reimers, 2013).

Many judgment studies have shown that people have difficulties in appreciating *randomness* (Tversky & Kahneman, 1974; Lawrence & Makridakis, 1989; Sanders, 1992). Harvey (1995) found participants simulated both pattern and noise in data series when forming their forecasts. In a string detection task, people had difficulties in distinguishing random and non-random strings (Lopes & Oden, 1987). It appears that people have a tendency to seek patterns in pure noise in untrended series (Harvey, 1988; O'Connor et al., 1997; Moritz et al., 2014). As O'Connor et al. (1993, p. 163) argued, "people seemed to change their forecasts in response to random fluctuations in the time series, identifying a signal where it did not exist. This was especially true for series with high variability".

As for instability (discontinuity or regime change), O'Connor et al. (1993) found that people were insensitive to different types of change. They investigated the impact of series instability by presenting participants with a simple series exhibiting a major discontinuity at a certain point but there was no significant effect of type of discontinuity (step or ramp) on absolute percentage error in their judgmental forecasts. It has also been found that people tend to focus on the strength or extremeness of the available evidence while insufficiently consider its weight or credence, yielding the over-confidence and under-confidence documented in

studies (Griffin & Tversky, 1992). Kremer et al. (2011) used a controlled laboratory experiment to determine how individuals detect changes in demand forecasting using artificial series by asking participants to separate random noise from persistent change. They showed that decision-makers overreacted to stable time series and under-reacted to highly unstable time series (by comparison with the actual error-response level adjustment behaviours obtained via optimal single exponential smoothing). As Kremer et al. (2011) pointed out, the bias that people overreact to noise appears to be better adapted to detecting changes in volatile environments than in stable environments. This is consistent with the system-neglect hypothesis (Massey & Wu, 2005), according to which people overweight signals and neglect the system which generates those signals.

#### *2.4.4. Data presentation format*

The presentation format of time series has an effect on judgments. Jones (1979) proposed that predictions are extrapolative, and that perceived trends are affected by a factor ( $\alpha$ ) that depends on format used for data presentation (e.g., tabular versus graphical). Graphical presentation benefits forecast accuracy compared to numerical presentation in a table (Angus-Leppan & Fatseas, 1986; Dickson et al., 1986; Levy et al., 1996). Indeed, it had been found that people are fairly skilled at identifying trends by observing time series presented in graphs (Lawrence et al., 1985; Mosteller et al., 1981; Ciccione & Dehaene, 2021).

Conditions for utilizing different presentation approaches can be specified in terms of forecast horizon. For example, Lawrence et al. (1985) provided evidence indicating that tabular presentation is more accurate than graphical presentation for long-term forecasts while graphical methods were beneficial for short-term forecasts. This was because graphical presentation led people to anchor their judgments on recent values which was relatively beneficial for short-term forecasts. In contrast, the tabular approach focussed their attention on long term-trends or average values. However, it is still possible that the last values in a table might also anchor people's judgments to some extent. Furthermore, as I discussed above, presence of trends can influence the judgmental heuristics that people use.

Presentation formats can influence how well people are able to identify randomness in trended series. Graphical presentation helps people to ignore randomness in trended time series. When the trend is confounded with randomness, no impact of randomness was found for graphical presentation (Lawrence & Makridakis, 1989; Mosteller et al., 1981), while with a tabular display, people were distracted from the detection of trends by the presence of randomness (Andreassen & Kraus, 1990).

In addition, Harvey and Bolger (1996) pointed out that the choice of presentation mode should depend on characteristics of data. Specifically, the advantage of graphical presentation is due to facilitation of trend estimation (thus reducing error for trend series) but this type of presentation increased over-forecasting and inconsistency with untrended series (thereby

resulting in slightly more errors relative to tabular presentation). For trended data, trend damping was found to be greater with a tabular format because it obscured the trend. Thus it is better to use graphs for trended data and tables for un-trended data.

Graphical presentation includes various graphical types, such as bar charts, line charts and scatterplots and so research has examined the effects of using different types of graphical presentation. Handzic et al. (2002) showed that, in an artificially generated sales forecasting task, provision of both bar charts and scatterplots produced improvements relative to the naïve forecast; however, scatterplots were significantly more effective than bar charts in enhancing forecasters' knowledge and performance.

Newman and Scholl (2012) reviewed research on the within-the-bar bias (i.e., people judge the likelihood of point falling within the bar to be greater when the point is within the bar than when it is the same distance from the top of the bar but outside the bar). This bias persists whether or not error bars are shown, whether or not bars originate from the upper or lower axis, whether bars are directly perceived or recalled from memory, and in different population samples. Hence, forecasts are systematically lower (and errors correspondingly higher) when series are presented as bars than when they are presented in other formats (Reimers & Harvey, 2022). The effect acts to offset trend damping for downward trends but to strengthen trend damping for upward ones. The effect is reversed when data are presented as hanging bars. Tumen and Boduroglu (2022) also found that forecasts in bar graphs were significantly less accurate than point graphs and line graphs and that was because these graphs induced more trend damping than point graphs and line graphs in a context-free setting.

Bar graphs might give a biased impression of central tendency: systematically underestimate the average (Godau et al., 2016; Kang et al., 2021) thus influencing forecasts from them. However, in an average estimation task, it has been shown that using cumulative bars with different colours or instructing participants to estimate the average height of the bars based on the *edges* of bars can reduce within-the-bar bias because they serve as reference points for determining the averages (Kang et al., 2021). In addition, this within-the-bar bias in medical treatment decisions could be effectively reduced by using a point presentation (Okan et al., 2018).

Other types of graphs result in further biases. Glaser et al. (2019) found that showing return charts as opposed to price charts resulted in lower expectations and concluded that, when forecasting future prices and returns, use of typical bar charts for returns draws attention to the entire available history of the series (leading to the calculation of average return/ trend damping) but showing line charts for prices induces extrapolation from only the most recent past data (leading to a recency effect with forecasts based just on the last quarter's data). Hence, showing return charts as opposed to price chart results in lower expectations. Theocharis et al. (2019) used real-life hurricane time series in three forecasting tasks (point

forecasting, probability density forecasting and interval forecasting). The forecast made based on a line graphic format was strongly anchored to the last data point and caused a larger error than using a discrete point graphical format, a finding in line with that of Zacks and Tversky (1999). Across the three tasks, the continuous line graph consistently led to people erroneously seeing relations and patterns between data points that were not present. This suggests that point presentation is to be preferred for non-correlated data series.

Scale is also an important factor in graphical presentation. Lawrence and Makridakis (1989) evaluated judgmental forecasts for artificial series by comparing them with a simple regression model. This revealed that judgmental forecasts were significantly less biased when series were presented nearer to the top of the graph. The top of the graph or computer screen may act as a visual anchor point that influences the forecast when the presentation is graphical (Harvey & Bolger, 1996). Later, Lawrence and O'Connor (1992) showed that forecast error declined as the vertical scale increased (although insignificantly) since people are prone to be biased towards to over-forecast with a small scale. This finding confirmed the consensus that the most appropriate data rectangle ratio is 1 (long side) to 0.7-0.75 (short side) (Cleveland et al., 1988). Also, Cleveland et al. (1988) showed that the line segment of a graph that brings its orientation closer to  $\pm 45^\circ$  will enhance slope judgments.

Inconsistent findings have been found with respect to effects of series length on forecasting. Lawrence and O'Connor (1992) used simulated non-trended monthly or quarterly time series to study the influence of series length. They revealed that the longer series with 40 points lead to a decrement in attention so that performance was worse than with shorter series (20 points). However, Andersson et al. (2012) found that forecast error decreased when length of data series increased from five to 10 or 15 items. This finding is not consistent with Lawrence and O'Connor's (1992) information overload assumption. On the basis of two experiments, Theocharis and Harvey (2019) argued that the apparent contradiction can be explained in terms of a shift from heuristic to systematic processing as the number of items in the data series increases. As a result, there is an inverted U-shaped relation between the length of series and forecast errors. In a similar vein, Moritz et al. (2014) pointed out that the time taken to make a forecast affects forecast quality: a moderate amount of time used by forecasters led to a significantly higher accuracy level compared with those who made forecasts too quickly or too slowly. It is highly possible that the time taken by forecasters implies different thinking styles (neglect, over-thinking, reflection, etc.) (Goodwin et al., 2018).

Recently, Kusev et al. (2018) proposed a new perspective for the presentation modes by taking into account the type of experience. They found that judgments in the dynamic mode were different from those in static mode. The static mode means one simultaneous presentation of all values of a series whereas the dynamic mode refers to the events in a time series are experienced sequentially by clicking button. Specifically, participants in the dynamic mode were anchored on more recent events for both forecasting and estimation judgments. Thus the dynamic mode has different effects on different judgmental tasks: it improves

forecasting accuracy (because of trend damping) but worsens average performance in estimation tasks (because of recency effects). Conversely, the static model impairs forecasting performance but benefits estimation accuracy because it draws attention to the average of the most recent values. Advantages of dynamic experience have also been confirmed for other tasks. Information obtained via direct personal experience has a stronger influence in economic game experiments than provision of information by observation (Simonsohn et al., 2008). Similarly, probabilistic judgments are better made via trial-by-trial experience than verbally stated information because the latter judgments followed the representativeness heuristic (Gigerenzer et al., 1988; Spellman, 1996).

## **2.5. Contextual information**

### *2.5.1. Rule-based forecasting*

Although econometric approaches incorporate causal variables within large datasets, they do not handle qualitative information and one-off factors. However, people can use their judgment to incorporate their domain knowledge and available contextual information into their forecasts.

We have seen that heuristic use is automatic when people extrapolate from pure time series. In practice, when people are asked to make forecasts, they often possess additional information. Webby and O'Connor (1996) reviewed judgmental and statistical time series forecasting and, in their summary, said that “contextual information appears to be the prime determinant of judgemental superiority over statistical models” (p. 98). Therefore, besides examination of pure extrapolation, study of the effectiveness of utilizing and processing contextual information is a major strand in judgmental forecasting research. One interpretation of the relation between these two strands of forecasting research is that it corresponds to dual-system theory of decision making in psychological research (Kahneman, 2011; Sloman, 1996; Larrick & Lawson, 2021). The first system uses the associative system and reflects the use of similarity and temporary structure in order to draw references and make predictions: it is likely to underlie extrapolation from time series. The second system is rule-based and operates using causal and logical relations: it is likely to be responsible for incorporating contextual information into forecasts. While in practice, judgments often involve using both systems and the dual-process theory is considered as a continuum of processing styles, not discrete types (Evans & Stanovich, 2013).

The “slow” process requires relevant knowledge (Evans & Stanovich, 2013; Sloman, 1996; Kahneman, 2000) and better performance is obtained using appropriate rules. For example, Lawson et al. (2020) examined the knowledge of a basic rule by asking a conjunction question and discovered that understanding of the conjunction rule was more likely in those who had a higher score on the Cognitive Reflection test (a slower thinking process). As expected, the importance of knowledge has also been found in forecasting contexts. Brown (1996) concluded that earnings per share (EPS) forecasts produced by management analysts attained

substantially higher accuracy than the best statistical models (no matter simple or sophisticated those models were). As Asquith et al. (2005) argued, there is substantial information content in managements and analysts' forecasts when they are released. This implies that judgment possesses information that is distinct from that used in statistical forecasts; it must incorporate information that has not been included in the model. Thus, contextual information plays an important role in forming judgmental forecasts.

### *2.5.2. Information advantage*

Being aware of task context is important for task performance. For example, many studies have shown that labelling of variables influences performance in multiple cue probability learning tasks (Miller, 1971; Sniezek, 1986) and time series forecasting tasks (Armstrong, 1985). Sniezek (1986) argued that meaningful labels assist encoding and retrieval of information in cue probability learning tasks. Armstrong (1985) showed that people behave differently when making forecasts for a set of time series labelled as "US Production of automobiles" and those labelled as "Production of X in Transylvania". Similar findings were found in the study of Glaser, Langer, Reynders, et al. (2007). They documented a framing effect: the differences between asking for prices and asking for returns for upward and downward sloping time series. Glaser et al. (2019) also showed that, in the task condition (directly asking their expectations), forecasts of returns were higher than forecasts of prices but in the stimuli condition (presenting price line charts and return bar charts), forecasts showed an opposed result. Rötheli (2020) identified various patterns of four-data-point series of inflation (price level) data and real GDP data which were retrieved from both USA and Germany. They presented them in graphs and asked 45 participants to make forecasts (pattern extrapolations). He found dissimilarities in pattern extrapolations for different macroeconomic variables for a long-term forecast horizon, for example, the five-year ahead inflation forecast differed from the five-year ahead GDP forecast. In contrast, the formation of expectations was similar for a short-term forecast of inflation and real GDP. These results show that context does matter for generation of expectations.

The efficient use of contextual information increases accuracy of judgmental forecasts and it has been widely investigated in financial forecasting settings. Much research has shown that, with available contextual information, judgmental forecasters perform better than statistical extrapolation. For example, Armstrong (1983) examined 15 studies that compared the earnings forecasts made by security analysts and managers against statistical extrapolation methods. Only three studies reported better results for statistical methods, one reported equal accuracy, and the remaining 11 reported that analyst forecasts were better than the statistical methods. In another example, Affleck-Graves et al. (1990) compared the accuracy of earnings forecasts of analysts, students and time-series models. Compared with both students and models (which possessed only historical data), analysts performed better and did so primarily by accessing a broader information set. Usually, their additional information implied a salient or slight pattern change in the time series. Subsequent studies conducted by

Brown et al. (1987) and Hopwood and McKeown (1990) further confirmed the better performance of security analysts' judgments. Recently, some studies further investigated the superiority of analysts' forecasts compared to other models and showed that analyst forecasts only were superior for some cases and depended on some associated factors such as forecasting horizons (Bradshaw et al., 2012; Gatsios et al., 2021). The superior performance of short-term analysts' forecasts is still consistent with the point that, with available contextual information, judgmental approaches appear consistently better than statistical models that ignore such knowledge.

More specifically, the advantage of analysts is partially attributable to their private information acquisition activities that enable them to identify different types of shocks (Brown, 1993), their timing advantage of making forecasts after receiving information released after the end of the fiscal year (Fried & Givoly, 1982), their possession of management forecast information and other firms' earning reports (Brown et al., 1987), and their information about ongoing strikes (Brown & Zmijewski, 1987).

Access to judgmental forecasts that are based on contextual information is particularly useful when the inherent variability in the time series being forecast is relatively high. In demand forecasting tasks, Sanders and Ritzman (1992, 1995) found that judgmental forecasts reinforced by use of contextual information gained more advantage as the variability of series increased. They further suggested a linear relationship between data variability and the amount of contextual knowledge needed.

The information advantage disappears in certain situations, such as forecasts of interest rates. Angus-Leppan and Fatseas (1986) noted that the knowledge of the nature of series (knowing they were interest rate series) did not improve forecast accuracy of inexperienced management accounting undergraduates' short-term interest rate forecasts. Kolb and Stekler (1996) studied prominent analysts' forecasts of short-term and long-term interest rates published in *The Wall Street Journal*. They found few analysts outperformed naïve no-change random walk type of forecasts. In addition, forecasters did not differ in their ability to predict short term 90-day T-bill bonds but did do so when predicting the yield on 30-year Treasury bonds. Kolb and Stekler (1996) suggested that these differences between forecasting for long-term interest rates and short-term interest rates arose because forecasting long-term rates requires more external factors to be considered than the forecasting of short-term rates.

Spiwojs et al. (2008) scrutinized 10-year US government bond yield forecasts and three-month US Treasury bill rate forecasts within a total of 136 time-series obtained from 24 organizations. All forecasts were biased and only a few outperformed model forecasts. Specifically, long periods of underestimation occur when interest rates are rising while long periods of overestimation exist when rates are falling. They argued that, in the majority of cases (70.6%), this was because analysts inefficiently integrated the most recent information into their forecasts. The contrast between these results and those of Kolb and Stekler (1996)

may be related to the specific interest rate tasks used in their study but it is also possible that other factors are at work. For example, Spiwoks et al. (2008) examined the entire distribution of forecasts whereas Kolb and Stekler (1996) focussed on the median values of the forecasts.

### *2.5.3. Information overload*

As the number of pieces of information or cues increases, the complexity of the task rises. Though information may allow forecasters to consider broader aspects their forecasting tasks, forecast accuracy does not positively correlate to the amount of information available (Armstrong, 1985). Information exceeding human processing capacity may add to the cognitive burden. Faust (1986) found that judges were not able to make proper use of large numbers of cues; instead, they used only a subset of the available information (Brehmer & Brehmer, 1988).

In their judgmental adjustment study, Lim and O'Connor (1996a) made a clear distinction between how people appreciate different types of information: time-series, statistical forecasts and causal information. Pure judgmental extrapolation from time series could benefit from any type of additional information. However, there were no differences in forecast accuracy between groups receiving causal information and those receiving statistical forecasts or receiving both causal information and statistical forecasts. This suggests that people were able to properly use only one piece of additional information in this task. Similarly, Handzic (1997, as cited in O'Connor & Lawrence, 1998) examined the ability of people to utilise up to three cues (casual variables) in addition to the time series. Results indicated that people were, at best, able to utilise two cues and the second cue was used under-optimally. The view that these findings indicate effects of information overload was supported by Lawrence et al. (2000). They showed that data overload may lead contextual information to be either ignored or given an inappropriate weighting (e.g. overemphasised). Webby (1994) manipulated the event information load (no event information, 4 pieces and 8 pieces) in a time series forecasting task. It was clear that the high information load condition impaired forecast accuracy. However, assistance from a prototype forecast support system improved accuracy level.

These studies suggest that, when forecasting from time series in the presence of a number of explicit quantitative causal variables, forecasters should be aware of their limited cognitive capacity and the consequent potential for information overload. They should consider using just the most valuable contextual information to reduce that load or, alternatively, to consider use of a forecasting support system.

### *2.5.4. Pertinent information*

The reliability of causal cue information plays a significant role in forecast accuracy. Armstrong (1983) found that management forecasts for earnings outperformed analysts'

forecasts. One possible explanation for this was that managers have more internal, pertinent and recent information leading to superior performance. In Lim and O'Connor's (1996a) study, the causal information condition (presented in bar charts) was divided into three levels (groups): no causal information, low-validity causal information and high-validity causal information. As expected, the effectiveness of causal adjustment was contingent upon the reliability of the causal information: adjustment of forecasts using causal information of low reliability did not lead to significant improvement, but adjustments based on highly reliable causal information produced more accurate forecasts than the best statistical models. In addition, Lim and O'Connor (1996a) found that the accuracy of the high-causal cue group did not deteriorate over time, but performance of the no-causal information group deteriorated over the three blocks. This could indicate an additional advantage of provision of causal information: it makes the forecasting task more interesting and helps to maintain motivation.

Remus et al. (1995) provided judgmental forecasters with either no information about regime changes in time series, with imperfect information about those changes, or with perfect information about them. Although information was underused, the perfect (accurate) information significantly improved judgmental forecasting relative to provision of imperfect information or no information. In a follow-up study, Remus et al. (1998) went on to show that correct information generally increased forecast accuracy more than incorrect information (and more than no information). Hence, pertinent information is recognised as such and this benefits forecasting.

#### *2.5.5. Technical knowledge*

Technical knowledge refers to the knowledge about data analysis and forecast procedures and is usually distinguished from contextual information. There is little evidence showing that it improves forecast accuracy. Carbone et al. (1983) found that participants with limited training in technical knowledge generated forecasts as accurately as those with extensive technical knowledge. Similarly, in studies of demand forecasting, Sanders and Ritzman (1992, 1995) emphasized the importance of contextual information (demand environment and customer purchasing patterns) for the combination of judgmental forecasts and statistical forecasts over that of technical information (a thorough background on both statistical and judgmental forecasting procedures). In support of their argument, they pointed out that warehouse managers with experience of the demand environment and with technical knowledge produced better forecasts than undergraduate students who had only technical knowledge.

There is one exception. Lawrence et al. (1985) found that technical knowledge improved accuracy when data were presented in tabular format rather than in graphical format. Thus, technical knowledge may facilitate people's numerical processing but it could be replaced by other approaches, such as use of a graphical presentation format.

### 2.5.6. Causal information

Causal information is a type of contextual information (Sanders & Ritzman, 1992) and it is of great importance in improving judgment accuracy. In a laboratory study, Lim and O'Connor (1996a) required participants to predict sales of soft drinks on a beach. Factors varied in the time series forecasting task were the statistical forecast (present or absent) and levels of causal information (no causal cue, high causal cue of temperature, and low causal cue of the number of tourists). They concluded that people were able to use causal information and that it improved forecast accuracy more than provision of statistical forecasts (which were only time series based). This result confirmed earlier conclusions of Sanders and Ritzman (1992).

Product knowledge and industry knowledge, which are important types of information in business forecasting, have shown varied effects on forecast performance. Edmundson et al. (1988) examined the judgmental forecasting process at a large consumer products corporation. They found that judgment-based forecasts with specific *product* knowledge were better for forecasts of short-term key products and non-key products than three other forecasting conditions: forecasters with just industry forecasting experience; a quantitative forecasting method using only deseasonalized single exponential smoothing of the sales history; and forecasters making judgmental forecasts from graphs of historical data. Thus, it appeared that *industry* knowledge had no influence on accuracy.

The importance of one-off information or broken-leg cues (Meehl, 1957), such as the occurrence of a specialized advertising campaign or the development of new technology has also been documented (Kurke & Aldrich, 1983; Kleinmuntz, 1990). This is of vital importance due to its high relevance to the future; it can provide an advanced indication that a discontinuity may have occurred (or will occur) irrespective of patterns in the historical data. Gorr (1986) found that special event data provided insight into the behaviour of the time series in government information systems and were beneficial to the forecasting process. Given this information, people could potentially make dramatic changes to their forecasting. But, without it, people hardly identified changes as they occurred and performed poorly when forecasting the resulting discontinuous series (O'Connor et al., 1993, 1997).

It is not easy for novices to perceive the underlying dependence of the series to be forecast on soft information (qualitative information relating to the future) such as rumours, opinions or product shortages and, hence, difficult to make a proper response. Johnson (1988) showed that use of unusual soft information distinguished an expert from a novice in the task of predicting student performance at a university. Also, as we have seen, information load influences the forecasting ability. As a result, increasing the amount of soft information impairs forecasting accuracy. Webby (1994) devised an experimental study and allocated participants to one of three conditions: no information about events, information about four events and information about eight events. In addition, he supported the half of participants with a prototype decision support system (GRIFFIN) that enabled forecasters to decompose

the task by concentrating on the effects of one event at a time. Results revealed that, although the computer-supported group showed reduced error compared with the manual forecast group, accuracy quality decreased with the introduction of more events.

Though soft information is of great value for judgmental forecasting, people appear to perceive it in a biased manner. De Bondt and Thaler (1985) characterised a price reversal pattern (prior 'loser' stocks outperformed prior 'winner' stocks) and proposed that investors overreact to unexpected financial market news events, thereby causing them to underweight older information in constructing their forecasts. Andreassen (1990) suggested an alternative psychological model based on intuitive time-series extrapolation (the mean reversion effect): the effect of news varies due to two factors that influence the *relative salience* of (1) change information about older versus recent series, (2) the trend component in time series (increasing or decreasing). Thus people should make an effort to balance one-off information and base rate information, so as to avoid aggressive reactions to events and underestimation of the effects of base rate information (Denrell & Fang, 2010).

#### 2.5.7. Domain knowledge

Forecasters' domain knowledge affects their understanding and interpretation of contextual information and, therefore, can facilitate time series judgmental forecasting (Webby et al., 2001). Lawrence et al. (2000) demonstrated that discussion in sales forecasting meetings is heavily focused on the meaning and implications of the contextual information. As one might expect, knowledge of the forecasting context improves forecast accuracy. In their review, Webby and O'Connor (1996) illustrated the contribution of this type of knowledge to forecast accuracy in facilitating the integration of non-time-series information into the forecasting process.

In fact, the prominent role of domain knowledge in facilitating judgment forecasting has been observed in various tasks. For example, Edmundson et al. (1988) indicated that specific product knowledge contributed significantly to sales forecasting accuracy in a large multinational company with a one-third reduction in the forecast error compared to purely extrapolative approaches. Furthermore, the importance of domain knowledge for improving forecast accuracy across different periods has been supported in a series of judgmental adjustment studies (Mathews & Diamantopoulos, 1989; Sanders & Ritzman, 1992, 1995). Nikolopoulos et al. (2005) undertook a product demand forecasting study in a UK-based household consumer products company. Half the statistical forecasts were adjusted with managerial judgments which were backed up by domain knowledge. Overall, these forecasts were better, especially when substantial adjustments were made (but not when those were overshoots). Similar findings in retail contexts have been reported (Fildes et al., 2009; Li et al., 2018, as cited in Fahimnia et al., 2019).

### 2.5.8. Experts

Experts possess abundant domain knowledge and have a comprehensive understanding of information but, compared to novices, they could benefit more from accessing better sources and more up-to-date information and from better utilization of information. In addition, people have a better understanding of underlying processes in more familiar scenarios (questions/ tasks) and this, in turn, leads to better performance (Sanderson, 1989). In addition to the timing advantage of acquiring more information, Brown et al. (1987) demonstrated the contemporaneous advantage: in other words, firms' quarterly earnings forecasts are superior to those derived from univariate time series models because their analysts were able to better utilize existing information than the models.

Although performance between experts and laypeople has not been found to differ much in some studies, their information processing may still differ (Johnson, 1988; Picart, 1990). Shanteau's (1992) investigation showed that experts use an automatic mental process to recognize patterns in noise. They are able to quickly simplify complex issues by discriminating relevant and extraneous information (Goodwin, 2017). This argument holds for the forecasting process. For example, De Bondt (1993) showed that lay people predict prices by extrapolating from past trends, while experts do not. Additionally, he maintained that such differences in knowledge structure or cognitive processing originate from experience.

Nevertheless, domain knowledge does not always benefit experts. For example, Makridakis et al. (1993) showed additional expertise had a limited beneficial effect on the accuracy of statistical models in the M2 competition (a small-scale competition with five participants that took place in real-time between 1987 and 1989). Lawrence et al. (2006) pointed out the inconsistencies in the findings concerning how well experts' forecasts perform relative to those of laypeople and attributed them to the task characteristics.

Motivation to improve accuracy is important in attaining accurate forecasts (Fildes & Hastings, 1994; Webby & O'Connor, 1996; Fildes et al., 2009) because professional forecasters who add their internal biases to their forecasts produce disappointing judgmental forecasts. Lawrence et al. (2000) conducted a field study of 13 large Australian and international manufacturing organizations and recorded the companies' sales forecasts of key products over 6 months. The forecast accuracy of 13 organisations was studied and compared with simple benchmarks (e.g. the naïve forecast). Although contextual information discussed in meetings was relevant and valid, results showed that nine out of 13 companies' judgmental forecasts were worse than or no different from a forecast based on the last actual value (i.e., the naïve forecast) and most of them were worse than forecasts that were purely based on statistical models. They observed that forecasts were biased by organisational budgeting and incentive issues. Managers in the meetings could not rationally employ available contextual information because they had different objectives for the forecasts. As expected, when this organisational

bias was (statistically) removed, their forecasts tended to be significantly and substantially more accurate than the naïve benchmark.

Other evidence has shown that the motivational biases of experts interact with their forecast performance. With a substantial dataset that took into account the effects of special events and changes, Fildes et al., (2009) found that judgmental adjustments made to the model's forecasts did not all lead to similar improvements in the forecasts. There was a general bias towards optimism in demand forecasts; half of the final forecasts were made by adjusting outputs of statistical models in the wrong direction. Also, the effectiveness of domain knowledge has been criticised and been acknowledged as a "self-serving bias" (Lawrence et al., 2006). High accuracy of forecasts may arise because some forecasters (e.g. managers) exert control of the forecasted variables; for example, in sales (or earnings) settings, they may launch a promotion or some other activity to incentivise sales in order to meet their forecasts (i.e., targets). Therefore, Goodwin (2017) used the example of American weather forecasters to pinpoint one major criterion for being a reliable forecaster: no personal intention or motivation for making the forecasts.

#### *2.5.9. Simple Heuristics*

Many studies have shown that judgmental forecasts with contextual information perform better than the output of statistical models (Webby & O'Connor, 1996; Lim & O'Connor, 1996a). However, there are still some cases that do not fit this picture. In these cases, historical data were sufficient for making predictions but causal information could impair simple judgmental extrapolation because of imperfect human judgment (Webby et al., 2001). O'Connor and Lawrence (1998) summarized evidence from laboratory studies and practical evidence and pointed out that people have problems with information acquisition and utilization when they use single or multiple cues in addition to time-series data.

Human judgment tends to rely on some simple heuristics to reduce memory overload and task complexity (Tversky & Kahneman, 1974), especially when people do not fully understand causal relations between variables. For example, Andreassen (1991, cited in Lim & O'Connor, 1996a) compared judgmental forecast accuracy in two groups and found people in the condition of using time-series made more accurate forecasts than those in the condition using causal information. Harvey et al. (1994) reached a similar conclusion. They found future passenger predictions with additional causal information of the number of criminals were less accurate than predictions just based on the number of passengers' weekly data. This study cast light on the validity of the economic theory and challenged the rational expectations and adaptive models. Specifically, within-series forecasts were made by extrapolation based on the trend of last change in series. Cross-series forecasts appeared to be based on a rather simple but faulty conditional inference process: a coarse categorization. That is, "if criminal numbers were abnormally high or low, passenger numbers were thought to be high; otherwise, passenger numbers were assumed to be close to their average level (p. 219)". It

implied that the extrapolative heuristic works in a “procedurally rational” manner (Simon, 1976) to balance between the cognitive and ecological approaches in human reasoning. In a similar vein, the coarse categorization incorrectly simplified the cross-series rule in order to reduce cognitive burden.

Another type of bias has been attributed to the acquisition and utilization of cues since causal information cannot be properly weighted when there are multiple cues. When forecasting the sales of soft drink on a surf beach, Lim and O’Connor (1996a, b) showed that people placed excessive weight on initial forecasts or a statistical method and much less weight on causal information (a single piece), especially for high-reliability causal information. A similar pattern was revealed in another experiment (Handzic, 1997, cited in Armstrong, 2001), in which participants were asked to forecast the sales of ice cream at a famous surfing beach in Sydney. The ineffective use of information by people was identified by the extremely unequal weights put on three pieces of causal information (forecast temperature for the next day, the number of visitors likely for the next day at the beach and the ratio of sunshine hours to total daylight hours).

Poor ability to combine multiple pieces of information has also been found in experts, though they are good at selecting and coding relevant information. For example, Mintzberg (1973) found that they put too much weight on the one-off cue information. As a result, differences in performance outcomes between experts and novices are often small or non-existent (Kynn, 2008). Johnson and Sathi (1988, as cited in Johnson, 1988) found that experts had better ability to combine information than novices (new master students in finance) because news information significantly improved the forecast accuracy of experts more than novices. However, they also found that experienced company analysts over-weighted case-specific data (22 related variables included in the news) compared to base-rate data. Therefore, experts and novices were both inferior to a relatively simple linear regression. Johnson and Sathi (1988) recommended development of an aid to help decision-makers combine different pieces of information appropriately to produce an overall estimate of their effects.

#### *2.5.10. Presentation format*

Use of a suitable presentation format may reduce the cognitive burden of additional contextual information and improve forecast accuracy. Handzic et al. (2002) examined the effectiveness of graphical tools on forecasting tasks. Participants were given artificial time series data of sales of ice cream as well as contextual data comprising continuous local temperature and visitor’s numbers. Half participants received a scatterplot tool while the others received the bar chart tool. The scatterplot group exhibited better knowledge extraction (67% vs. 27%) and produced more accurate forecasts than the bar chart group. It seems that, with proper graphical tools, people can efficiently use two contextual cues. Thus, developments of such tools could facilitate learning and forecasting performance.

Presentation format may also provide an explanation for the inconsistent findings regarding the usefulness of causal information. The conclusions of Handzic et al. (2002) and Lim and O'Connor (1996a) were contrary to the findings of Harvey et al. (1994) and Andreassen (1991, cited in Lim and O'Connor, 1996a), who found that causal information was of little use. The former two studies provided causal information in graphs (scatterplot chart and bar chart). However, the latter two studies employed a tabular format to present casual information and were not described as a typical forecast task. Tabular presentation may serve to occlude the association between variables, thereby increasing cognitive processing load and producing poorer performance.

In summary, it is difficult to conclude that certain kinds of information are universally superior to others; differences in performance largely depend on which information is a good predictor of the future for the specific task. However, we can say that people have difficulties fully utilizing multiple pieces of casual information (O'Connor & Lawrence, 1998). Hence, approaches aimed at reducing cognitive load (presentation format, system aids, expertise, etc.) are essential for achieving accurate judgments and avoiding the use of inappropriate heuristics.

## **2.6. Forecasting under uncertainty**

Uncertainty can be divided into two classes with regard to its source: (1) environmental uncertainty, which originates from the probabilistic nature of the environment, and (2) internal uncertainty arises because of individuals' imperfect information, knowledge or skill (Gillies, 2000; Kahneman & Tversky, 1982; Peterson & Pitz, 1988; Kozyreva & Hertwig, 2021; Løhre & Teigen, 2016; Ülkümen et al., 2016). The perception and judgment of uncertainty of a task are highly subjective and contextual (Stephensen et al., 2021).

Accuracy and uncertainty are two essential components when talking about forecasting. The purpose of forecasting is to provide reliable information for decision makers and to improve the quality of decision-making when facing the pervasiveness of uncertainty in the world. Though point forecasting could be regarded as an average outcome when forecasters repeatedly make forecasts a number of times under the same conditions, when it comes to uncertainty, point forecasts provide little value.

Communication of uncertainty is critical (Armstrong, 2001; Önköl & Bolger, 2004; Papastamos et al., 2018). Some common practical ways of communicating uncertainty in forecasts have been investigated. Prediction intervals provide the upper and lower bounds within which outcomes are forecast to appear with a specified probability (Goodwin et al., 2010); probability forecasts indicate the probability that a known value or chosen event will occur (Whitcomb et al., 1995), probabilistic directional forecasts indicate the probability of change in a given direction (Bolger & Harvey, 1995; Budescu & Du, 2007; Önköl et al., 2003), and probability density forecasts provide a distributional measure of forecast uncertainty (Diebold et al., 1999). These methods convey uncertainty information and are important for providing

decision-makers with a complete picture of potential risks and for enabling them to plan for possible alternatives, especially in the domain of economics and finance (Taylor & Bunn, 1999; Clements & Taylor, 2003; Lahiri & Sheng, 2010a, b).

Although interval forecasting and probability density forecasting have increasingly been recognised as important, there are fewer studies addressing psychological issues concerning how they are judgmentally produced than there are for point forecasts (Lawrence et al., 2006). Existing research has mainly focused on three areas: preference for different forecast types from providers' and users' perspectives, factors impacting forecast accuracy, and forecasters' confidence judgments under different circumstances. I will discuss these topics in the context of interval forecasting and probability density forecasting separately. (Below, I have incorporated some research of probability forecasting under the section on interval forecasting because researchers made direct comparisons between them.)

## **2.7. Interval forecasting**

### *2.7.1. Effects of interval width*

Interval forecasting depicts uncertainty information about the future in a simple way for both providers and users of forecasts (Önkal-Atay et al., 2002). Its use is common in domains such as weather forecasting (Zarnani et al., 2019), economic forecasting (Moiseev et al., 2019), and financial forecasting (Filimonov, et al., 2017).

Some studies have shown a clear user preference for prediction intervals over point forecasts (Baginski et al., 1993; Pownall et al., 1993). Önkal and Bolger (2004) revealed that, in stock market settings, 95% prediction intervals were considered to be the most useful format, followed by directional predictions, 50% interval forecasts, and, finally, point forecasts, regardless of the forecasting role (either forecast providers or forecast users).

However, the preference of the coverage of interval forecast does not appear to be consistent across situations. Though a 95% forecast interval is most commonly used, Granger (1996) described 95% prediction intervals as often being 'embarrassingly wide' and argued that 50% intervals are more likely to be believable in the context of economic forecasts. Goodwin et al. (2010) found neither 50% prediction intervals nor 95% prediction intervals were beneficial for production planning but suggested use of 85% prediction intervals. Indeed, the accuracy-informativeness trade-off varies according to the forecast context (election candidates versus scientists) (Yaniv & Foster, 1995, 1997). People may sacrifice forecast accuracy for the sake of being informative. Thus, Goodwin et al. (2010) suggested that, rather than employing the standard intervals such as 50% or 95%, forecasters' choice of prediction intervals should take the nature of the task into consideration to meet the particular needs that are present.

Specified interval width (e.g., 95%, 50%) affects calibration because the degree of over-precision relates to it. People appear more overconfident with wider specified intervals

(Lichtenstein et al., 1982; Block & Harper, 1991). For instance, Budescu and Du (2007) found that, when individuals were asked to play the role of investors and were required to predict future stock prices which were randomly selected from a public database, 90% confidence intervals were over-precise and 50% intervals were under-precise whereas 70% intervals were well-calibrated. This pattern can be explained by a trade-off between the two competing objectives—accuracy and informativeness. That is, narrower intervals are less likely to contain the true value but are more useful for decision making.

Similarly, Önkal and Bolger (2004) found that, in the context of a stock price forecasting task, the overconfidence effect present with 95% prediction intervals disappeared when 50% prediction intervals were elicited from the same forecasters for the same stock price series. The difference between these two studies was that participants were required to take different roles: they were assigned to be forecast providers in Önkal and Bolger's (2004) task but to be investors in Budescu and Du's (2007) task.

### *2.7.2. Effects of providing interval information on accuracy*

A considerable amount of research has been devoted to obtaining reliable prediction intervals in order to improve the accuracy of judgments and decisions (De Gooijer & Hyndman, 2006; Diebold et al., 2017; Grounds et al., 2017; Mandel et al., 2020). However, evidence showing that provision of interval forecasts enhances judgment and decision making is inconsistent. Oliver (1972) and Keys (1978) found that bank loan officers made similar decisions regardless of what types of information they received in the financial statement (whether single figures or confidence intervals). Birnberg and Slevin (1976) also failed to reveal the significant difference between two types of information disclosure (five scores vs. five scores and 95% confidence intervals) on selecting a target performance level for MBA students.

However, Hawkins (1974) using a sample of financial analysts and Chen and Summers (1981) using MBA students, showed that investment decisions were facilitated by disclosure of additional probabilistic information (such as interval-scaled accounting figures with confidence levels). In a similar spirit, better decisions and more coherent answers were made when the flooding water level forecasters were provided with probabilistic information (probability of flooding) than when they were not (Ramos et al., 2013).

Johnson (1982) investigated the usefulness of interval forecasts. In a probability judgment task in which people were asked for their estimates of a stick's length, he manipulated environmental uncertainty as well as information uncertainties and used four report formats (point estimate, augmented point estimate (probability estimate), confidence interval (interval estimate) and conditional distribution (density estimate) in a pre-test-post-test design. Participants receiving point estimate reports about sample outcome estimates outperformed those given confidence interval information in both low and moderate uncertainty conditions. However, superior performance was achieved by the interval

information group when uncertainty was high. This suggests that interval forecasts are beneficial only under high uncertainty situations.

Can these conflicting results be reconciled? Financial statement analysis tasks may place such severe restrictions on responses that the information and environmental uncertainties provided by prediction intervals have little effect on performance. Bank loan officers given confidence interval information may have focused exclusively on the midpoint of interval (i.e. the best guess), thereby altering the task to make it identical to that of the control group in which participants were supplied conventional (point estimate) financial data (Birnberg & Slevin, 1976). On the other hand, when the decision environment is highly uncertain, people may consider other possibilities, such as the interval information conveyed in the current trial or even in proceeding trials. Thus, the benefit of providing interval information may depend on whether it induces additional and efficient information processing by the forecasters performing the tasks.

### 2.7.3. Factors influencing interval forecasts

One strand of the psychological research into interval forecasts has focused on how prediction intervals are affected by various factors. I considered characteristics of time series, question domains, the question format and expertise.

I will first consider *characteristics of time series*. Studies have shown that interval forecasting is influenced by trends, seasonality, randomness, forecast horizon, variability in data series and the presentation scale of a graphical plot used to present the time series (Lawrence & Makridakis, 1989; Lawrence & O'Connor, 1993; O'Connor & Lawrence, 1992). For example, prediction intervals become wider for trended time series (Eggleton, 1982), especially for downward sloping series (Lawrence & Makridakis, 1989). Directional probabilistic forecasts for exchange rates also exhibit a strong effect of trend (Thomson et al., 2003) and it was found that performance on ascending trends was superior to that on descending trends (Thomson et al., 2013). Prediction intervals also tend to be too narrow at a large presentation scale but too wide at a small presentation scale (Lawrence & O'Connor, 1993).

Regarding variability, more overconfidence is observed when data series contain high variability and this is reflected in narrower intervals. In contrast, intervals that are too wide, indicating under-confidence, are found for series with low variability (Lawrence & O'Connor, 1993; Lawrence & Makridakis, 1989). However, this effect of variability disappeared when horizontal graph lines and vertical value scales were removed (Lawrence & O'Connor, 1993). This fits with Beach and Solak's (1969) suggestion that individuals' prediction intervals are proportional to the correct answers (vertical scale values). Similarly, when compared to regression-based confidence intervals, forecast intervals are too narrow, signalling overconfidence, for series with high levels of randomness and too wide, indicating under-confidence, for series with low levels of randomness (Lawrence & Makridakis, 1989). However,

forecast interval width does not appear to be higher for series that are seen as more skewed (De Bondt, 1993).

How data points are presented in graphs influences interval production. For instance, continuous line graph presentation has been found to lead to narrower intervals than discrete point graph presentation (Theocharis et al. 2019). Discrete data presentation may reduce the degree to which people anchor on previous data points when producing their intervals.

It is generally accepted that accuracy level is higher for short forecast horizons than for longer ones. However, presence of this effect appears to depend on the type of forecast. Önkcal et al. (2003) found that one-day ahead *point* forecasts and *interval* forecasts were more accurate than one-week-ahead forecasts of those types. However, *directional* forecasts showed the opposite pattern. Thomson et al. (2004) also showed an interesting horizon pattern in terms of accuracy: one month < six months < three months. This was true for both experts and novices, and they suggested that this might be due to the high incidence of short-term noise (volatility) in the series so that the one-month horizon was difficult to forecast.

*Question domains* affect responses to interval forecasts and their tendency towards overconfidence (Soll & Kalyman, 2004). In particular, it is the familiarity of the domain that affects interval estimates. Block and Harper (1991, Experiment 1) required people to estimate 12 quantities. When people estimated unfamiliar quantities (i.e., those that participants had not perceptually experienced), they exhibited overconfidence with too narrow confidence intervals. However, confidence judgments were better with more familiar quantities (e.g. the height of a standard soft-drink can); they showing less overconfidence. These findings are consistent with those reported by Pitz's (1974): he found that participants were less overconfident when estimating familiar quantities than when estimating unfamiliar ones. It is possible that they perceived estimating unfamiliar quantities to be more difficult than estimating familiar ones. This is relevant because of the hard-easy effect.

The hard-easy effect refers to the fact that the degree of over-confidence depends on task difficulty: more over-confidence is found for more difficult tasks and under-confidence is found for very easy tasks. This phenomenon has been documented for general knowledge questions (Lichtenstein et al., 1982) as well as for forecasting questions (Bolger & Önkcal-Atay, 2004; Niu & Harvey, 2022, Chapter 5). People expect to be less accurate when they answer a difficult set of questions (Klayman et al., 1999; Suantak et al., 1996); they produce wider confidence intervals for unfamiliar quantities than for familiar ones (Block & Harper, 1991). Crucially, however, their lower expectations are not sufficient to compensate their inaccuracy.

In terms of *question format*, there are inconsistent findings regarding whether adding a point estimate to the task improves interval estimates and whether the effects of doing this are influenced by whether it is done before or after making the interval estimate.

Based on the assumption that people use the anchor and adjustment heuristic (Epley & Gilovich, 2006; Tversky & Kahneman, 1974), providing the best estimate first should result in worse calibration because the explicit anchor should lead to greater over-precision. However, results have not been conclusive. Russo and Schoemaker's (1992) showed that, when participants were asked to answer 20 trivia questions, overconfidence shown in their 90% prediction intervals was greater when participants had to report the best guess forecast prior to the prediction interval (61% misses) than when they reported interval forecast without first making a best guess (48% misses).

However, other studies have revealed the reverse effect (Block & Harper, 1991; Juslin et al., 1999). Using a series of experiments in which participants answered trivia questions, Block and Harper (1991) found that generating the best estimate and explicitly displaying it prior to eliciting a confidence interval (50% and 90%) improved the calibration of confidence by widening the intervals. Although supplying a person with another person's point forecast implicitly impacted on people's own forecast via an anchoring effect, this did not serve to decrease their overconfidence. The authors suggested that overconfidence may occur because people do not realistically assess their estimation ability. However, when generating and displaying the best estimate, people came to realize just how poor their estimation ability was. This can be seen as consistent with Epley and Gilovich's (2006) proposal that adjustment is sufficient when participants are motivated and able to engage in effortful thought.

Soll and Kalyman (2004) conducted a series of experiments and supported the benefit of adding a third estimate to an interval estimation task. In experiments 2 and 3, they required people to estimate the median value of a quantity as well as probability interval for that quantity. Their results showed that adding this median estimate improved the quality of the interval estimate. This can be interpreted as showing that requiring an estimate of median provoked people into searching their knowledge more deeply, improving the accuracy of their estimates and increasing the size of their confidence intervals.

In Soll and Klayman's (2004) studies, there was no clear evidence showing that the order in which best estimates and confidence intervals were elicited influenced accuracy and calibration in different question domains. Mandel et al. (2020) also found elicitation order had no significant effect on the accuracy of best estimates and the calibration of confidence intervals for general knowledge questions. Thus, it is still unclear exactly how point forecasts affect interval forecasts but it does appear that additional cognitive effort has a role to play in the process.

Regarding probability forecasting, the specific assessment method used affects different aspects of calibration performance. For example, in a cross-cultural study (Whitcomb et al., 1995), eliciting numerical probabilities resulted in higher Slope values than eliciting odds. (In other words, their discrimination was better.) However, participants who used the odds

method had lower Scatter scores than participants who used numerical probabilities or visual pie diagram methods. (In other words, they showed lower judgment noise.)

*Expertise* is another factor affecting interval forecasts. McKenzie et al. (2008) showed that, although novices and experts correctly reckoned that experts' interval estimates were narrower and their midpoints were closer to the truth than those of novices, they underestimated the overconfidence and overestimated the hit rate of experts. However, they made reasonable predictions for novices' performance. Menkhoff et al. (2010) even found that not all professionals were more sophisticated than laypeople regarding their private investment decisions: institutional investors were more sophisticated but investment advisors were not.

The accuracy level of experts' predictions depends on the forecast method used. In terms of accuracy, Önkal et al. (2003) directly compared experts' and sophisticated amateurs' ability to predict foreign exchange rates one day and one week ahead. Each participant was provided with the daily rate for the previous four months. Results showed that experts tended to outperform the novices in terms of point predictions and predicting directions of change (Whitcotton, 1996; Önkal & Muradoglu, 1996), but there were no differences in terms of the number of outcomes within their 90% confidence intervals.

Whether the forecast horizon is long or short also influences experts' performance. A good interval forecast represents a high ability to assess uncertainty. Therefore, one important factor for making accurate interval forecasts is an ability to identify factors that will determine the future of the series (e.g., trends) over a period long enough to reach the forecast horizon. Hamill and Wilks (1995) found that experienced forecasters had difficulty with short-term day-to-day uncertainty forecasts; there was no positive correlation between their interval forecast widths and their absolute errors. This contrasts with Winkler and Murphy's (1979) findings in which a much longer nine-month forecast period was studied. In this case, experienced weather forecasters had positive correlations between their 50% interval widths and their forecast errors, indicating a clear ability to assess uncertainty.

Concerning confidence judgment, it has been always found that people generate too narrow ranges: hit rates inside 90% confidence intervals are typically below 50%, implying that the confidence people have in their beliefs is inappropriate and always exceeds their actual accuracy level (Soll & Klayman, 2004; Moore et al., 2015). This conclusion holds across diverse levels of expertise (McKenzie et al. 2008; Atir et al., 2015; Deaves et al., 2010; Juslin et al., 2007; Glaser et al., 2013). Experts sometimes even demonstrate more overconfidence than novices (Glaser et al., 2007; Glaser et al., 2013), named "inverted expertise effects".

McKenzie et al. (2008) asked for 90% prediction intervals for questions about IT and about participants' own universities. They found that professionals and students exhibited equal overconfidence. This result is in agreement with Önkal et al.'s (2003) finding that experts and

novices produced confidence intervals for their foreign exchange rate predictions that showed similar levels of overconfidence. McKenzie et al. (2008) summarised findings as showing that experts and novices are about equally overconfident and have similar hit rates overall. Experts reported intervals that had midpoints closer to the true value thereby increasing their hit rate. However, their intervals were narrower which, in turn, decreased their hit rate. Thus the net effect was that there was no significant difference between novices and experts in their hit rate or their overconfidence. Lambert et al. (2012) also found that the degree of overconfidence was similar for experts (bankers) and novices when using interval evaluation questions to forecast 20 stocks listed on international markets.

One might presume that experts could use more contextual information to increase their forecast quality. Results from Whitecotton's (1996) study, showed that experience had a positive impact on accuracy. All participants were presented with the financial performance of companies and historical data (without labelling the firm's name or the time frame) and were asked to make probability forecasts (0 to 1) for the question that "the next earnings would exceed the indicated 4-year trend". Results revealed that experienced financial analysts and MBA students exhibited a higher accuracy level than undergraduates. Although financial analysts exhibited more biases than the other two groups, they still yielded the highest overall accuracy level. Whitecotton (1996) pointed out that these results conflict with those from a study by Yates et al. (1991) that revealed an 'inverted experience-accuracy effect'. In the Whitecotton (1996) study, forecasters were constrained in the information that they could use: they were limited to using a "prepacked" information set. In contrast, forecasters in the Yates et al. (1991) study were unconstrained in the information that they could use. Experts had more information but much of it was not directly relevant to their forecasting task. However, they did not realise this. Hence, allowing them to select and make use of their additional information damaged their performance because so much of it was irrelevant. As a result, using this information offset the advantage that they gained from their experience.

The overconfidence of experts may also be explained using Random Support Theory (Brenner, 1995, 2003). According to this approach, probability judgments reflect the balance of evidence captured by underlying distributions of support for both correct and incorrect hypotheses. Experts have more experiences which are in accordance with their forecasts which may result in their expressing greater confidence. This is consistent with the observed pattern of mis-calibration in expert judgment reported by Koehler et al. (2002).

#### *2.7.4. Interval construction*

Inconsistent findings about the sensitivity of the interval estimates to the confidence probability level have been found. For instance, using between-participant designs, Jørgensen et al. (2004) and Teigen and Jørgensen (2005) found that different groups of participants provided almost identical confidence intervals for 50%, 75%, 90%, and 99% levels. However,

some studies revealed contrasting findings. Budescu and Du (2007) showed that decision makers' judgments of quantiles were sensitive to the degree of confidence because they widened their range estimates according to the confidence level. This is consistent with earlier findings (Alpert & Raiffa, 1982; Juslin et al., 1999).

Nevertheless, people are able to learn to adjust the intervals to overcome overconfidence. In the first session of a four-session experiment (Bolger & Önkal-Atay, 2004), participants were firstly asked to choose a percentage confidence for their intervals between 50% to 90% inclusive and then make one-step ahead probability interval forecasts for 32 stock price time series in graph. In session 2 and 3, after a three-day interval, participants (1) received feedback from the last session, (2) chose a percentage confidence and (3) made interval forecast for a new set of 32 series. In the last session, participants were given feedback for session 3 and completed a questionnaire. Results showed that, initially, interval forecasts were overconfident. With calibration feedback over three sessions, calibration improved since forecasters were able to learn to match confidence-interval width with confidence percentage over those three sessions: the percent selection of confidence percentage categories (50–59%, 60–69%, 70–79%, 80–89% and 90–99%) shifted to higher percentage category over sessions, and specified averaged interval widths increased accordingly.

It is reasonable to expect that prediction intervals will be symmetric around the point forecast as the statistical methods (Taylor & Bunn, 1999). As Soll and Kalyman (2004) mentioned, it is hard to know where a judge's anchor lies within a range, but it is likely to be near the median (i.e., where the answer is believed to be equally likely to be above or below). Indeed, some researchers use the measure of midpoints of ranges to compared with the true values and some refer to it in task instructions (Önkal et al., 2003; Lee & Siemsen, 2017). For example, in a study concerning ordering decisions by newsvendors, Lee and Siemsen (2017, p. 3231) adopted a 90% interval estimate for the uncertainty estimate, and participants were instructed to provide 90% confidence intervals such that "you expect a 90 percent likelihood that the next demand will be in this interval. The midpoint of the interval is the point forecast".

However, several studies have shown that some factors lead to asymmetric judgmental prediction intervals (De Bondt, 1993; Önkal et al., 2013; O'Connor et al., 2001). Contextual information is one such factor. For instance, Spence (1996) found that both experts and novices produced asymmetric confidence intervals when they were required to provide a fair market value for residential properties and a range about their estimates. He also found that more asymmetric intervals and more biased estimates were produced by experts than by novices, suggesting that contextual environment or non-time series information may affect degree of asymmetry. Önkal et al. (2013) explored the effect of scenarios on judgment adjustments of model-based forecasts. Participants received graphs of historical demand for different products, together with a one-period-ahead model-based point forecast for the next month that served as forecast advice. Participants who received only best-case scenarios gave

the most symmetric intervals compared to groups who were given no scenarios, worst-case scenarios, or both best and worst scenarios.

Characteristics of time series also contribute to interval asymmetry. O'Connor et al. (2001) examined the asymmetry of confidence intervals in a laboratory setting by examining the position of bounds of 95% confidence intervals relative to point forecasts. They found that trends affected the asymmetry of the confidence interval, and the last actual value of a time series was a major determinant of the direction and size of the asymmetry. Therefore, it appeared that people adopted a hedging strategy with their point forecast being biased in one direction but with the confidence interval around it being biased in the opposite direction.

Pfajfar and Žakelj (2016) replicated these findings. When participants were asked to forecast inflation and to provide 95% confidence intervals around their point forecasts, they often chose an asymmetric confidence interval when given the option of doing so. Furthermore, participants given this option produced narrower intervals and were, therefore, less accurate than those who had to produce symmetric intervals. In general, participants who provide asymmetric confidence intervals typically place narrower confidence intervals on one side of the distribution and wider ones on the other side to hedge the risk. In particular, the lower limit of the interval was very inertial, while the upper interval actively reacted to economic conditions.

In these studies, the point forecast was elicited before the interval bounds. O'Connor et al. (2001) suggest that results might be different if the sequence of task elicitation were changed. Indeed, Budescu and Du (2007) showed that, when the point forecast was made after the confidence interval, the mean of median estimates “closely matches the midranges of the confidence intervals.

Beliefs about the average forecast of others could also lead to asymmetry in interval forecasts. Zhu et al. (2022) examined a hedging hypothesis for prediction interval formation in stock price forecasting. They were able to show that the asymmetry of interval forecasts arose because forecasters hedged their forecasts by adjusting the bounds of their prediction intervals in a way that reflected their forecasts of the average forecasts of other people.

All asymmetry research has been conducted under conditions in which production of interval forecasts was not the sole requirement. Additional contextual information has been provided or other types of forecasts have been required at the same time. As a consequence, asymmetries could have arisen because forecasters adopted hedging strategies. For future research, it would be interesting to investigate the asymmetry of interval production under different conditions separately using between-participant designs. It would also be useful to investigate how other characteristics of time series influence interval forecasts; this is because some of them are known to be major determinants of individuals' forecasts. These

include the slope of the last segment (O'Connor et al., 1993), the long-term mean, and the last data point of the series (Bolger & Harvey, 1993).

People adjust the two bounds of their interval forecasts as they receive more information (Önkal et al., 2013) but the way in which they do so reveals certain biases (Pfajfar & Žakelj, 2016). For example, Önkal et al. (2013) found that lower interval bounds are adjusted most and in an upwards direction (i.e., towards the centre-point of the interval provided by a model's predictions) when optimistic scenarios were received and that upper bounds are adjusted most and in a downwards direction (towards the centre-point of the interval) in the presence of pessimistic scenarios. Soll and Klayman (2004) found that when best guess questions were answered in conjunction with interval forecasts, intervals were widened and overconfidence was reduced; this supports the view that searching for additional information and realising the limitations to one's own forecasting ability significantly change the way that people form their interval forecasts.

Other research on construction of interval forecasts has focused on the diversity of elicitation formats: a single limit (an upper bound or a lower bound), a two-point range (two separate questions for upper and lower bounds) and an interval (a range).

Construction of the two bounds to an interval may rely on different processes when questions eliciting those bounds are framed in different ways. Soll and Klayman (2004) compared the procedures of generating a two-point range and a percentage interval range that was not explicitly characterized as a two-point range. These two formats may seem similar. For example, a 90% confidence interval can be modelled as two binary decisions: 95% certainty about the lower limit and 95% certainty about the upper one elicited separately should lead to the construction of a 90% confidence interval in which the true values will fall. However, the two-point range format, by asking for a sampling proportion, should not be biased towards overconfidence in the way that interval forecasts are. Experimental results supported this hypothesis; separate elicitation of two points produced wider intervals, higher hit rates, and better calibration than unitary elicitation of the interval.

In a similar vein, Teigen and Jørgensen (2005) compared the interval method and the single-limit method in a task in which participants were asked to estimate the populations of European capital cities. It turned out that participants in the single-limit conditions (termed as "more than" or "less than") were under-confident and generated much wider intervals than participants in the interval method. In another similar study of directional probability forecasts, Bolger and Harvey (1995) asked participants to estimate the probability of the next data point being above or below a specified reference value for trended and un-trended series. Results showed that participants appeared to be under-confident in their predictions.

Yaniv and Schul (1997) compared the effects of inclusion and exclusion elicitation strategies using general knowledge questions with 20 possible responses for each question.

Participating students were instructed to produce responses by either highlighting all the likely answers (inclusion) or unlikely answers (exclusion). The inclusion instruction elicited much smaller sets of responses (18–21% of the full set) than the exclusion instruction (43–50%). The exclusion elicitation strategy had the effect of reducing overconfidence and improving calibration.

Teigen et al.'s (2007) experiment used seven groups to investigate effects of question format (single, two-point, range) and term type (exclusive vs inclusive) on confidence. They replicated the findings of Teigen and Jørgensen (2005) and Soll and Klayman (2004) in a price estimation task. In later experiments, they found a moderating effect of term type on confidence level. Whether the single limit format yielded less over confidence than the two-bound format depended on whether the bounds were based on exclusive or inclusive terms. The degree of external variability of the outcome variable is also important. For example, the prices for flight tickets are not fixed but depend on the quality, date and market factors. When distributional questions had high variability, inclusive terms resulted in very extreme values; as a result, they were much further apart than when exclusive terms were used. Thus, the production of two boundaries of intervals is influenced by the question type and the variability of the outcome variable.

#### *2.7.5. Overconfidence in interval forecasts*

Overconfidence is a widespread phenomenon in interval forecasting and it is found consistently across domains (Glaser et al., 2007, 2013; Teigen et al., 2007; Parker & Fischhoff, 2005), expertise levels (Önkal et al., 2003; Jørgensen et al., 2004; Meikle et al., 2016), gender (Soll & Klayman, 2004), age (Hansson, Rönnlund, et al., 2008) and culture (Russo & Schoemaker, 1992). It is also found when groups estimate interval forecasts (Ang & O'Connor, 1991).

Some researchers are cautious about accepting excessively narrow intervals as evidence of a general tendency towards overconfidence. They aim to explain this bias in other ways. For example, Gigerenzer et al. (1991) and Juslin (1994) argued that overconfidence is produced by including too many misleading questions in the corpus, thereby reducing its ecological validity. This issue can be addressed by using a well-designed forecasting task not subject to that problem. However, overconfidence is still detected. For example, Glaser et al. (2013) asked three types of questions: (1) a stock price prediction task in an artificial environment to exclude the effect of information or knowledge, (2) general knowledge questions that were representative of the environment (Juslin, 1994), and (3) real-world stock market forecasts over horizons of two weeks. All three tasks required forecast intervals in a two-bound format. Overconfidence was confirmed over domains and over time. Furthermore, correlations of the overconfidence scores based on confidence interval questions were highly positive, indicating the stable individual differences in miscalibration. These findings were consistent with those of Klayman et al. (1999); overconfidence was found across several sets of tasks. However,

there was format dependence such that the levels of overconfidence varied from task to task, just as Soll and Klayman (2004) have documented.

One type of format dependence appears to have practical implications. Requiring people to assign a probability to a given interval (probability evaluation) systematically produces lower overconfidence than requiring them to assign an interval to a given probability (interval production) (Bolger & Harvey, 1995; Harvey, 1988; Hilton et al., 2011; Klayman et al., 1999; Teigen & Jørgensen, 2005). For example, Teigen and Jørgensen (2005) showed that individuals were less over-precise when they assigned a confidence level to an interval than when they produced an interval for a specified level of confidence (e.g., 90%). Clearly, performance of these two tasks relies on different cognitive processes (Hansson, Juslin, et al., 2008; Winman et al., 2004).

People are naïve with respect to biased estimators (Fiedler, 2000). Probability assessment (half-range and full-range formats) requires estimation of a proportion which is an unbiased estimator of the population (sample proportion). In contrast, interval production requires estimation of a dispersion of plausible values which is a biased estimator of the population equivalent (sample dispersion). Thus this type of format dependence provides support for the naïve sampling model (Winman et al., 2004): in the interval production task, people act as naïve statisticians and generate a “naïve sample” of relevant exemplars in response to the probe question (Juslin et al., 2007).

Budescu and Du (2007) reported no difference in the miscalibration between the two formats. However, this inconsistent finding could be explained in terms of specific features of their experiment design. For example, they used a within-participants design and participants were asked to judge multiple probabilities (the probabilities that the stocks will reach a certain threshold, provided intervals of 50%, 70%, and 90% confidence intervals of their future prices, and an estimated median). In addition, they made different types of comparison between performance in the two formats.

Consistent with the idea of removing dependency, thinking about each bound differently leads to more separated boundaries than setting two bounds simultaneously. This was confirmed by Teigen and Jørgensen (2005). They asked one group of participants to provide a typical range estimate and the other two groups to provide either a lower or upper bound. The two bounds of intervals generated by different persons were wider apart than the two bounds of intervals which were constructed by the same individual, thus suggesting that thinking about each bound independently is likely to lead to more disparate evidence being retrieved than when thinking of them simultaneously. However, this reduction in over-precision was itself subject to a format effect rather than changing their underlying thinking process (Winman et al, 2004): a later study by the authors used a different type of response (confidence interval production) and the above pattern of results was not replicated. Thus it

appears that the original pattern arose from how confidence was expressed via the responses rather than from how it was cognitively assessed.

Confidence assessments can be divided into local and global confidence assessments (Lieberman, 2004). The former refers to the confidence levels for individual estimates and the latter refers to the expected number of correct answers for all estimates that have been made in the task. Global confidence assessments show much less overconfidence than local confidence estimates (Gigerenzer et al., 1991; Griffin & Tversky, 1992; Liberman, 2004; Sieck & Arkes, 2005; Snizek & Buckley, 1991; Teigen & Jørgensen, 2005; Teigen et al., 2007). Indeed, global assessments can show under-confidence. Cesarini et al. (2006) required undergraduate students in Economics and Business to answer 10 numerical questions specific to the domain of economics. In the interval estimate stage, participants were required to give upper and lower limits for each question with 90% subjective confidence that would contain true value. In the later frequency estimate stage, they were asked to answer how many of their intervals contained the true value. Results revealed that using the frequency estimate reduced overconfidence by about 60 percent relative to using interval estimates. Monetary incentives failed to decrease this effect. Gigerenzer et al. (1991) explain the effect of assessment type by arguing that a different reference class is used when making these two types of judgment. For global judgments, the reference class is 'tests of this type' whereas each individual question in local assessments has its own reference class.

These results are also consistent with the finding that people are better at addressing frequency problems than at dealing with probabilistic questions. Many researchers (e.g., Hogarth & Soyer, 2015) have argued that humans have evolved to deal with frequencies whereas probabilities are a cultural construct of relatively recent vintage. Various phenomena support this notion. For example, the violation of the conjunction rule turns out to be much reduced when participants are provided with information in a frequency format (Fiedler, 1988; Hertwig & Gigerenzer, 1999; Tversky & Kahneman, 1983). Furthermore, people are better able to reason in accordance with Bayes' rule when all information is given as frequencies rather than as probabilities (Gigerenzer & Hoffrage, 1995; Sedlmeier, 1997).

However, this evolutionary argument has been challenged by dual-process theorists (Evans & Stanovich, 2013; Sloman, 1996). They have argued that use of frequencies helps people to clarify the underlying nested-set relations between the components of the probabilistic problems. As a result of this, cognitive processing shifts from use of a primitive associative judgment system to use of a slower but more effective deliberative rule-based system. This implies that the frequency/probability difference can be eliminated by making the underlying nested-set relationship evident and transparent in some other way in order to induce rule-based processing (Sloman et al., 2003).

### 2.7.6. Heuristics and biases

Moore and Schatz (2017) distinguished three facets of overconfidence: overestimation (compared with self), over-placement (compared with others) and over-precision (being too sure about knowing the truth). The former two are more driven by motivational bias and the last one is mainly caused by cognitive errors. Both overestimation and over-placement are affected by the task difficulty, which is called, the hard-easy effect. However, they occur in different ways: People tend to overestimate their performance on hard tasks and underestimate it on easy tasks (Lichtenstein & Fischhoff, 1977) whereas people believe they are worse than others on difficult tasks but better than others on easy tasks (Moore & Healy, 2008).

A common finding is that produced intervals are too narrow, so that the proportion of intervals that includes the correct value is way below the pre-stated confidence level, reflecting overconfidence. This effect is an example of the type of overconfidence termed over-precision according by Moore and Schatz (2017).

Various explanations of this overconfidence have been developed and can be divided into two categories: cognitive biases and motivational/affectional biases. I will discuss these in turn.

First, I will consider *confirmation bias*, a phenomenon that may lead people to comprehend new evidence in the direction of favouring their hypotheses (Tsai et al. 2008). This may occur because of the nature of human associative memory. People are most likely to retrieve information that has semantic connections with already-retrieved information rather than a random sample of evidence. As a result, the sample of information that people use to judge uncertainty is a small, biased subsample of the available data (Tong & Feiler, 2017). The confirmatory bias model of Koriat et al. (1980) is more specific. According to this approach, a general tendency towards overconfidence arises from people's inclination to seek confirmation while discounting disconfirming evidence (Eubanks et al., 2015; Klayman, 1995; Legoux et al., 2014). The stronger and the more affirmative the reasons that are recruited from memory to support their hypothesis, the greater the confidence expressed. This tendency may particularly affect assessments of highly self-relevant outcomes (Hoch, 1985; Koehler et al., 2002). However, confirmation bias might be intensified or be mitigated among experts; which of these happens depends on whether their extensive experience is biased or unbiased (Koehler et al., 2002).

Use of *the anchoring-and-adjustment heuristic* (Tversky & Kahneman, 1974) also provides a possible explanation for overconfidence. Tversky and Kahneman (1982) suggested that the reason why judgmental confidence intervals are too narrow is because of the effect of anchoring: point predictions serve as anchors and judges do not adjust away from them sufficiently to produce an interval that is appropriately wide (Jørgensen & Sjøberg, 2001). However, inconsistent findings about the effect of asking a point estimate before a confidence

interval implies that anchoring may not be the only determinant of overconfidence (Goodwin et al., 2010; Block & Harper, 1991; Mandel et al., 2020).

The *availability heuristic* is based on selective memory retrieval (Soll & Klayman, 2004; Tversky & Kahneman, 1974; Russo & Schoemaker, 1992). The heuristic may normally be adaptive. However, encountering information with a higher frequency than is representative of the natural environment (e.g., via media exposure to unusual events) may result in biased judgments of frequencies or probabilities. Furthermore, because sampling is naïve, confidence in those judgments cannot take account of the factors that resulted in them being biased. People will therefore be overconfident in their biased frequency judgments (e.g., Hoch, 1985; Koriat et al., 1980). Koriat et al. (1980) found that overconfidence was substantially reduced when people had to generate reasons why they might be wrong. Thus when people are instructed to engage in counterfactual reasoning, the effect of their automatically generated reasons that support their responses arising from the availability heuristic was counterbalanced. A related phenomenon is the explanation effect: it refers to the finding that requiring people to provide explanations for their predictions significantly increases their confidence of them (Koehler, 1991). This is because thinking about explanations increases the availability of the occurrence of the event that is already in mind and, hence, inflates judgments of its frequency. As Arkes (2001, p. 499) pointed out: “In the context of forecasting, it seems that just considering the likelihood of a future event makes the event seem more probable and thus increases the forecaster’s confidence in the prediction.”

*Hindsight bias* refers to people’s overestimation of their forecasting ability. They incorrectly believe that their past forecasts were more accurate than they, in fact, were: they suffer from a false belief that their forecasting was better than it was: this is often termed the “I-knew-it-all-along” effect (Fischhoff, 1975; Roese & Vohs, 2012; Russo & Schoemaker, 1992). As a result, forecasters’ confidence judgments about their past forecast accuracy are inflated (Roese & Vohs, 2012). This is likely to mean that they will over-estimate their ability to make future forecasts too. Pohl and Erdfelder (2016) pointed out that hindsight bias occurs because participants are unable to access their prior judgments (Erdfelder et al., 2007; Pohl & Hell, 1996); it is this that allows them to reconstruct their earlier estimates in a biased manner.

The increasing *the amount of information* available to forecasters may lead them to believe that this information will improve their ability to make forecasts and, hence, increase their confidence levels. Tsai et al. (2008) explored the effects of the amount of information on judgment accuracy and confidence level in three studies. When individuals received more information, their confidence increased faster than their accuracy: the accuracy of college football game predictions was significantly better than chance when provided with the first six cues but showed little improvement thereafter. Nevertheless, the confidence level continued to rise with the incremental addition of 24 more cues. This continued increment of

confidence without a corresponding change in accuracy led to confidence–accuracy discrepancies. A similar effect was demonstrated by Lusk and Hammond (1991). Meteorologists were assigned the task of forecasting microbursts and radar data was updated every 2.5 minutes. Though confidence rose with accumulation of information, no improvement was found in the forecast accuracy over time. Other studies have shown similar results (Oskamp, 1965; Peterson & Pitz, 1986). Similar confidence-accuracy discrepancies have been found in people who have more experience: those with more knowledge or more practical experience are typically more overconfident because their confidence has increased more than their accuracy (Arkes et al., 1986; Paese & Snizek, 1991; Dawson et al., 1993).

Other factors associated with *the provision of information* also affect confidence or probability judgment, including the order of receiving information, the length of a series of evidence items, response mode and the variability within a set of information (Hastie & Park, 1986; Hogarth & Einhorn, 1992; Kóbor et al., 2021; Voorhees et al., 2021).

Other factors that have been characterised as likely causes of overconfidence are better regarded as motivational or affective in nature.

The *informativeness-accuracy trade-off* (Yaniv & Foster, 1997) is one of these. Although wider intervals are more likely to contain the true values, people prefer narrow prediction intervals. Yaniv & Foster (1995) found that, even though participants were told that the true value of the number of countries of the United Nations in 1987 is 159, people preferred a narrow but wrong interval (140-150) over the wide but correct interval (50-300). In a real price stock forecasting task, Foong et al. (2003) found that forecasters used information provided by 50% intervals only when they were given performance-based incentives to do so. The authors postulated that the confidence interval information increased the impression that the task was algorithmic in nature and so made it appear more challenging. More broadly, wider intervals have been found to decrease forecasters' credibility (Du et al., 2011), reduce the degree to which they were seen as viewing "clarity of communication" as important (Önkál & Bolger, 2004, p. 37), give the impression of a lack of definitiveness and relevance (Yaniv & Foster, 1995, 1997), and result in forecasters being viewed as incompetent or lazy (Yates et al., 1996). Conversely, the use of high confidence percentages in judgmental probability forecasts is regarded as an indicator of expertise (Bolger & Önkál-Atay, 2004; Price & Stone, 2004). Keren (1997) showed that after knowing the outcome that it had rained on three out of four days, two-thirds participants preferred forecasters who gave four probabilities of 90 percent of the rain than the other forecaster who gave four probabilities of 75 percent. Although the latter forecaster demonstrated better calibration, people preferred the forecast with less uncertain probabilities irrespective of their correctness. In addition, Yates et al. (1996) reported that consumers tend to focus particularly on forecasts using extreme probabilities (close to 0% and/ or 100%). Such preferences for higher percentages by both the providers and the users reinforce the notion that confidence percentages are treated by users as a pointer to the providers' expertise.

The “*better-than-average*” effect refers to the finding that judgments about the self are typically better than the judgments about average peers (Beer & Hughes, 2010; Alicke & Govorlin, 2013; Taylor, 1989). Alicke and Govorlin (2013) and Moore and Schatz (2017) concluded that the effect is motivational in origin and demonstrated a reversed hard-easy effect: the increasing difficulty of a task leads to people to judge themselves to have done worse than others. This is in agreement with other findings (Lichtenstein et al., 1982; Larrick et al., 2007; Moore & Healy, 2008). Cesarini et al. (2006) found that it could be reduced by introducing monetary incentives or other techniques to motivate people to overcome it (e.g., feedback) (Harvey, Koehler, et al., 1997).

*Optimism bias* refers to overestimation of the probability of desirable events and underestimation of the probability of undesirable ones (Kahneman & Riepe, 1998; Weinstein, 1980; Sharot, 2011). It can lead to over-forecasting and overconfidence. This optimism may result from the selective focus on positive information over negative information (Krizan & Windschitl, 2007). It may be realised as wishful thinking, the prediction that desirable events are more likely to happen than undesirable ones (Babad, 1987; Kunda, 1990; Mayraz, 2011). In particular, desirable events will be judged as being more likely to happen to the forecasters themselves whereas undesirable events will be judged as being more likely to happen to others (Zakay, 1983). For example, Babad (1987) asked 1000 male participants to forecast game outcomes before and during a football game. Fans showed wishful thinking in their pre-game predictions and the level of wishful thinking was positively correlated with self-defined levels of ‘fanhood’ and team preference. Sharot et al. (2011) gave people desirable information about the future (such as telling that their likelihood of suffering from Alzheimer’s disease was lower than their own estimation of it). She found that people updated their beliefs to a greater degree than when they had been given undesirable information (such as learning that the likelihood of their house being burgled was higher than their estimation of it).

Individuals tend to give higher probabilities (more certainty) to personal event forecasts than to impersonal events forecasts (Wright & Ayton, 1989), suggesting that desirability and perceived controllability have a role in producing estimates (Zakay, 1983; Weinstein, 1980). With non-personal events (world-event questions), predictions for desirable events are less overconfident and better calibrated for the immediate time period than for later time periods (Wright & Ayton, 1992).

The *base-rate fallacy* (Meehl & Rosen, 1955; Tversky & Kahneman, 1982) is defined as inappropriately ignoring historical information concerning population likelihoods and giving undue weight to particular events. It has been observed in professionals’ forecasts as well as novices’ forecasts (Denrell & Fang, 2010; Ayton, 1992). The tendency to focus on particular information causes biased perception and beliefs and results in people overlooking uncertainty and, consequently, being overconfident in forecasts. As a result, people typically

deviate from Bayes' theorem when combining base rates with new information (Kahneman, 2011).

Using computational modelling at the level of individual participants, Stengård et al. (2022) systematically compared alternative accounts of base-rate neglect, including accounts of relying on a toolbox of heuristics (Gigerenzer & Hoffrage, 1995), using strategies for multiple cue tasks rather than probability theory (Juslin et al., 2009), and allowing for the influence of random noise (Costello & Watts, 2014). They found evidence supporting base-rate neglect but also obtained individual differences indicating that some participants were able to take account of base-rates.

The involvement of motivational biases in confidence judgments may depend on the *elicitation methods and measurements of confidence*. Hilton et al. (2011) argued that overconfidence of interval production tasks is caused by purely cognitive factors and does not involve any motivational influences. They examined a series of positive illusions regarding overconfidence, including unrealistic optimism, a general tendency to consider oneself "better-than-average", and two indexes of dispositional perception of control. They showed that average overconfidence was much higher in the interval production tasks (setting an interval appropriate for a given probability) than in probability evaluation tasks (providing a probability appropriate to a given interval), thereby replicating earlier findings (Klayman et al., 1999). However, in addition, they found that (a) measures of positive illusions are substantially correlated amongst themselves; (b) although positive illusions were not correlated to interval production, they could predict probability evaluation. This suggests that motivational biases have a role in producing miscalibration in probability evaluation tasks but not in interval production tasks. This is consistent with earlier findings reported by Oberlechner and Osler (2012), who found that miscalibration in professional traders' currency exchange rate predictions (measured using an interval production task) was not correlated with the tendency to consider oneself better than others at trading. Thus it appears that interval production measures judgmental overconfidence without influence from positive illusions. However, this conclusion may be too simple. Kruglanski et al. (2020) argue that motivational and inferential processes are always interrelated and always interact with one another. It may be that motivational biases other than the ones studied by Hilton et al. (2011) affect interval production.

Using *groups* of people to make forecasts should increase the chances of identifying a wider set of alternative outcomes and so may decrease overconfidence. However, research has revealed the opposite. Heath and Gonzalez (1995) demonstrated that group interaction prompts people to explain and defend their own beliefs; it serves only to increase their confidence in their forecasts. Sniezek and Henry (1989) also found that 98% of participants thought that their group's accuracy level was above the median. Group members believed that their discussion would lead to more accurate forecasts than those made by individuals but, actually, the accuracy of group discussion was not necessarily improved. As a result,

group discussion enhanced confidence more than accuracy. In line with this, Boje and Murnighan (1982) found that members of groups did increase their confidence over successive trials, but their accuracy diminished. However, much may depend in the nature of the group interaction. Schwenk and Cosier (1980) tested four methods of improving prediction performance. They found that a devil's advocate method (a person criticized in an objective, non-emotional way) could efficiently decrease overconfidence.

In summary, group discussion influences confidence and accuracy, but the interaction is complex and depends on the characteristics of the group and discussion procedure within the group. Certain types of group interaction do facilitate forecasting. For example, the Delphi method, which benefits from anonymity, improves judgment over that obtained with face-to-face groups (Rowe & Wright, 1999).

## **2.8. Probability density forecasts**

Density forecasting enables forecasters to provide more information and express their confidence levels for a number of values around the point forecast. Forecast users often appreciate receiving this additional information and expect to find it helpful for decision making. However, density forecasting still has some drawbacks: from the view of forecast providers, density forecasting is quite complex, difficult and time-consuming; from users' perspectives, it is more difficult to comprehend and it is not always clear how to use it to improve decision-making.

Probability density forecasting requires allocating probabilities into intervals or bins. In practice, it has been widely used in economics, finance and meteorology. For example, it is included in various surveys for measuring key macroeconomic variables among professional forecasters. These include the Survey of Professional Forecasters (SPF) in the US (Clements, 2014), the Survey of External Forecasters in the UK (Boero et al., 2008), and European Central Bank's Survey of Professional Forecasters (ECB SPF) (Coroneo et al., 2019).

Existing literature has mainly focused on the quality of calibration of professional forecasters. Relatively few researchers have made direct comparisons between the forecast accuracy between density forecasting and other formats of forecasting, except some studies that address (in)consistencies between point and density forecasts (Engelberg et al., 2009; Clements, 2010; Zarnowitz & Lambros, 1987).

Probability density forecasts obtained from professionals outperform simple benchmark models. For example, Coroneo et al. (2019) examined the real-time performance of the European Central Bank's Survey of Professional Forecasters and showed that density forecasts were better than simple benchmarks (uniform, Gaussian random walk and naïve forecast) for unemployment and real GDP growth at one-year horizon, but not for inflation, which was well-anchored on inflation targets (Henckel et al., 2019).

Probability density forecasts have been found to lead to better forecast performance compared with other forecast task formats. Önkal and Muradoglu (1996) found that task format led to different probability forecasts and that the effect of format was different for experts, semi-experts and novices. To forecast weekly price changes for 32 stocks, participants were asked to make multiple interval forecasts (similar to probability density forecasts) as well as dichotomous forecasts (a directional forecast and a probability forecast). Multiple interval forecasts were transformed into sign-aggregated multiple interval forecasts so that they were comparable to the dichotomous forecasts. Results showed that experts' forecasts derived from multiple-interval forecasts showed a better calibration and a better overall accuracy level than dichotomous forecasts. This may have been because the multiple interval format allowed experts to employ their richer cognitive representations of the task (Murphy & Wright, 1984). In contrast, the dichotomous scale produced better performance in people who had limited knowledge (e.g., semi-experts and novices). It is possible, however, that the relatively poor forecast performance of semi-experts and novices was due to their unfamiliarity with or difficulty in using the multiple interval format.

There is also some evidence that density forecasts outperform point forecasts. In a professional forecast survey (the Deloitte CFO Survey Switzerland), Phillot and Rosenblatt-Wisch (2018) asked for both point and density questions without forcing the order of these responses. Density forecasts yielded more accurate forecasts than point forecasts and were less affected by the order in which the different forecasts were made. Providing the density forecast before the point forecast resulted in an approximately 5% increase in inconsistencies (the discrepancies between point and density forecasts), mostly by affecting the point forecasts. Hence, the authors suggest that point forecast should be provided first when both types of forecasts are elicited but that survey designers should opt for the density forecasts when only one type of forecast is elicited.

Haran and Moore (2014) reviewed the shortcomings of point forecasting and interval forecasting. They describe an approach to eliciting forecasts which offers better accuracy levels and greater flexibility for decision-makers; they term their new approach SPIES (Subjective Probability Interval ESTimates) (Haran et al., 2010). It provides a way of producing predictions that is identical to the procedure used in probability density forecasting. In one study, researchers asked participants to estimate the highest daytime temperature in Washington D.C. in one month's time. Those who were asked for 90% SPIES achieved a significantly higher hit rate (74%) than those who reported forecasts with 90% confidence intervals (29%) (Haran et al., 2010). Moreover, a carryover effect was found: use of the SPIES method changed participants' subsequent responses in different formats, implying that the method produced a learning effect.

Over-precision in inventory settings causes biased newsvendor orders but it has been found that the bias can be reduced or even eliminated by implementing the SPIES approach (Ren & Croson, 2013). In a simulated supply chain and ordering experiment, the control group

provided two forecasts at the beginning of every five rounds: (1) the best guess of demand, (2) their beliefs about the 5th percentile and the 95th percentile of the demand distribution (90% confidence intervals). In the treatment group, they were asked to respond to the SPIES questions before answering those best guess and interval questions. Results revealed that participants who used the SPIES approach to forecast future product demand achieved 3.1% higher profits than those who made their forecasts using 90% confidence intervals.

People's probability density forecasts are less biased than their interval forecasts. In Diebold et al.'s (1999) study, the mean probability forecasts (for each bin) of annual inflation were obtained from the Survey of Professional Forecasters (SPF). The distributions overestimated the uncertainty of inflation since they were too dispersed and most of the realizations fell squarely in the middle of the forecast density distribution. These findings were the opposite of those obtained with interval forecasts, where people expressed overconfidence by providing ranges that were too narrow. This suggests that probability density forecasts provide better calibrated measures of confidence and decrease overconfidence relative to that seen with interval forecasts. This is consistent with Lee and Siemsen's (2017) finding that use of the SPIES approach reduces overconfidence. I provide further support for the advantages of density forecasts (Niu & Harvey, 2022c, Chapter 6).

Using data from SPF, Engelberg et al. (2009) showed that point predictions were quite close to the central tendencies of forecasters' subjective distributions. But for those inconsistent forecasts, professional forecasters showed the tendency to report point predictions that give a more favourable view of the economy than do their subjective means/medians/modes derived from distributions. Similarly, García and Manzanares (2007) reported that forecasters tended to be biased towards favourable outcomes in that they forecast GDP growth estimates that are too high and inflation rate estimates that are too low. Given the fact that optimistic bias is most systematically observed with point forecasts, they suggest that density forecasts should be elicited and central tendency and uncertainty measures derived from the resulting distributions.

More generally, benefits of density forecasting are in line with Goodwin et al.'s (2018) suggestion that effective demand forecasting should involve gathering all relevant information and creating a probability distribution of future demand outcomes. Without that, vague uncertainty information may allow planners to make biased demand judgments and decisions.

Some drawbacks of probability density forecasting have been identified. In an inflation study, Diebold et al. (1999) examined several features of the density forecasts obtained in the Survey of Professional Forecasters in relation to realized inflation. The density forecasts of inflation were found to be non-uniform and autocorrelated. The authors argued that forecasters were expecting that the inflation series to be highly persistent and not to change rapidly, or that they are systematically optimistic or pessimistic. This is congruent with the finding that people

assume there is a correlation between current inflation and the next year's inflation that arises because inflation rate increases over time (Niu & Harvey, 2022a, Chapter 4).

In terms of performance accuracy, Clements (2010) evaluated the consistency of these two types of forecasts made simultaneously in the US Survey of Professional Forecasters and found they were inconsistent sometimes: point forecasts were more accurate because of delayed updating of the histograms relative to the point forecasts. This is in line with the proposition of Mankiw and Reis (2002) and Carroll (2003) concerning bounded rationality: forecasters do not continually update their forecasts as the arrival of new information but use simple forecasting rules. However, in Clements (2010) study, those point forecasts and density forecasts were elicited together and so a hedging strategy could have been employed – I discussed this above for the case in which point forecasts and interval forecasts are made together. As a result, it is still unclear what can be concluded about their relative accuracy levels.

The number of bins assigned with probabilities serves as a proxy for measuring the variance of underlying probability density forecasts (Clements, 2010; Giordani & Söderlind, 2003). Although the mean dispersion of density forecasts has been found to be positively related to the standard deviations of point forecasts, the latter tend to understate uncertainty (Zarnowitz & Lambros, 1987). Thus, inconsistencies between different measures of uncertainty in survey respondents have been found. It has been argued that the published aggregate density forecasts are inconsistent with the measure of disagreement of point forecasts (or the measure of an aggregation of individual measures of dispersion) and so are inadequate indicators of uncertainty in both UK surveys and US surveys (Engelberg et al., 2009; Boero et al., 2008; Clements, 2010).

### *2.8.1. Factors affecting density forecasts*

Though, as we saw above, question order does not affect density forecasts, other factors do. I review them in this section.

*Experts* generally outperform novices in terms of probability density forecasts (e.g., Önköl & Muradoglu, 1996; Murphy & Winkler, 1977) because they are more knowledgeable and more accustomed to the complex question formats. Nevertheless, a few studies that recruited various levels of expertise have shown a reversed expertise effect (Yates et al., 1991; Staël von Holstein, 1972; Önköl & Muradoglu, 1994) in financial forecasting with a multiple-interval task format. In the study of Yates et al. (1991), participants were provided with background information of each forecast target and then they were asked to make probabilistic forecasts (using six intervals) of the per-share stock prices and earnings of 31 companies. Forecasts of undergraduate students contained less variability and were independent of the actual price and earnings activities, thus they were more accurate than those of graduate students. This was consistent with results of a study by Staël von Holstein (1972) in which bankers exhibited

worse performance than other four groups (stock market experts, statisticians, business teachers and students). Yates et al. (1991) suggested that the effect arose because more experienced forecasters misused the additional cues (e.g., weaker cues) and the additional cues made the task more difficult.

Across *forecast horizons*, professionals' probability forecasts showed an accuracy that was comparable or even worse than that of a uniform distribution (i.e., equal probabilities assigned to all possible outcomes). For example, Bartos (1969) showed that security analysts' probability distributions were consistently inferior to a uniform distribution over forecast horizons of one, three and six months although the mean errors of the one-month horizon were smaller than those for the three- and six-month horizons. Similarly, Staël von Holstein (1972) found only three of 72 participants' average scores performed better than the uniform distribution for the closing stock prices of 12 stocks over 2-week intervals. However, he suggested that the accuracy could be improved if the most suitable forecast horizon was used. In particular, forecasters might perform better with an ecologically valid time horizon such as a few days or 6 months rather than the one (14 days) that he used in his experiment. Some support for this notion was provided by Kabus (1976) who found that banking executives made accurate probability interest rate forecasts for a horizon of 3 months, the forecast horizon that they habitually used for their forecasts.

*Rounding* behaviour is often observed in survey responses. Researchers usually classify forecasters as rounders if their point forecast are multiples of an integer number. For example, Binder (2017) defined consumers as rounders when their point forecasts were a multiple of five. Similarly, Manski (2004) noted that probabilistic forecasts are frequently rounded to the nearest 5 percent on the 0-100 percent scale and Clements (2011) observed that the probabilities reported in the FED-SPF tend to be multiples of five or ten.

Based on density forecasts of the Survey of Professional Forecasters for the Euro area and the US, Glas and Hartmann (2022) found that non-rounders used more of the available bins and reported higher variances than the rounders for all forecast horizons, though the difference became less pronounced as forecast horizon shortened. This agrees with Levenko's (2020) finding that rounding in density forecasting results in lower variance by reducing the number of bins used. However, Glas and Hartmann (2022) found that forecast error levels were not significantly different between rounders and non-rounders. Both Engelberg et al. (2009) and Binder (2017) have suggested that rounding has little impact on the mean of forecasters' subjective distributions.

Some researchers question whether a tendency to use rounding can serve as a measure of uncertainty (e.g., Krifka, 2002). Binder (2017) argued that rounding can act as a proxy measure for consumers' inflation uncertainty because it implies less knowledge. However, Clements (2021) disagreed with respect to US professional forecasts; his view is that more rounding behaviour is correlated to worse accuracy but cannot be used to measure perceived

uncertainty. Also, Levenko (2020) was of the view that the decreased variance in the density forecasts produced by rounding says little about the uncertainty perceived by forecasters. Engelberg et al. (2009) provides evidence against hypothesis that inconsistencies between point and density forecasts are related to rounding.

*Motivational factors* (incentives) might influence the quality of density forecasts. For example, Ehrbeck and Waldmann (1996), Laster et al. (1999), and Ottaviani and Sørensen (2006) have suggested that forecasters may react strategically to balance forecast accuracy with other objectives, such as convincing the market that they are well-informed or attracting media attention.

Compared with research on point forecasts and interval forecasts, research about influential factors of probability density forecasting is scarce. The factors involved in interval forecasts could be considered for future research into density forecasts.

### *2.8.2. Density forecast construction*

Goldstein and Rothschild (2014) have shown that laypeople understand density distribution and have an accurate mental representation of it. However, their responses may be inaccurate because of the difficulty that they have in articulating their answers or because they suffer from cognitive biases, such as optimism (Engelberg et al., 2009), primacy and recency effects (Deese & Kaufman, 1957), and anchoring effects (Tversky & Kahneman, 1974).

Goldstein and Rothschild (2014) presented participants with 100 numbers from one of six distributions and then allocated them to one of five elicitation conditions: a graphical method (using a graphical user interface to elicit frequencies into corresponding values), a standard fractile method, a standard mean method, a standard average method and an average range method. Results showed that at both individual and aggregate levels, laypeople's probability distributions were significantly more accurate when they used the graphical approach than when they used any of the other standard methods. This implies that individuals' intuitions about probability distributions are highly accurate, that they can form accurate mental representations of probability distribution, and that they are able to express them via the graphical elicitation method. All the other elicitation methods were significantly affected by the presentation order of samples and this increased their errors relative to those observed with the graphical method. Use of graphically elicited probability density forecasts could contribute to higher accuracy levels and less biased confidence judgments than those observed with single interval forecasts.

Hasher and Zacks (1979, 1984) have shown that frequency encoding of events is relatively automatic and accurate. Kaufmann et al. (2013) applied this finding to probability density forecasting. They showed that people who experience frequencies over time in a graphical risk tool show a better understanding of the characteristics of a risky asset than those who

were simply provided with summary statistics or a static distribution. Hogarth and Soyer (2011) also found that people exhibit more accurate assessments of probabilities when those assessments are generated on the basis of simulated experiences rather than on the basis of forecasting models.

The construction of the probability density forecasts by both experts and novices could be developed further and efforts in this direction could serve to uncover why this method is associated with superior performance.

## **2.9. Improving judgment**

Forecasts are suboptimal because they are affected by two types of error: inconsistency and bias. Inconsistency is an unsystematic error that is also termed as a random error. Bias is a type of systematic error. Improving judgmental forecasts and judgments of confidence in them are important topics for research.

### *2.9.1. Learning and education*

Although forecasting is difficult, there are some domains, such as weather forecasting, where it is relatively accurate. This accuracy appears to arise because forecasters are provided with a “kind” environment with a large sample, available cues, past data to learn from, and provision of immediate and clear feedback (Zsombok & Klein, 2014). Without these task features, people show very little learning over time (Fildes & Stekler, 2002; Lim & O'Connor, 1996b). Through two longitudinal experiments, Morewedge and colleagues (2015) showed that playing an immersive video game with feedback or watching a training video reduced six cognitive biases immediately and persistently. Also, Chang et al. (2016) showed that a one-hour cognitive de-biasing module providing training for probabilistic reasoning training boosted the accuracy of probability forecasts in a four-year series of geopolitical forecasting tournaments. These two studies indicate that acquiring information about biases and having the opportunity to put that knowledge into practice serve to improve judgmental forecasting accuracy.

Dohmen et al. (2009) showed that education and schooling can directly affect people's capability for making probability judgments. They investigated the process of perceiving probabilistic information in a sample of the German population. A third of participants were unable to make correct probability judgments concerning the tossing a fair coin based on a sequence of realizations. They did not know that each toss of a coin does not depend on outcomes of previous tosses. Additionally, people suffered from two well-known cognitive biases: the gamblers' fallacy (i.e., the tendency to overestimate the occurrence of alternations in random sequences) and the hot hand fallacy (i.e., the tendency to overestimate the occurrence of streaks in random sequences). Ability to give correct responses was significantly related to education: “It turns out that the effect of schooling is large and significant: an additional year of schooling is related to an increase in the probability of giving the correct

answer of about 4.5 percentage points ( $p < 0.01$ ), controlling for cognitive ability” (pp.909). Moreover, the impact of schooling was relevant not only to the coin game but also to participants’ financial decision-making and job search decisions.

### *2.9.2. Judgmental adjustment*

Use of judgment to modify statistical forecasts can improve accuracy. Fildes and Hastings (1994) interviewed forecasters and found that 84% of them stressed the importance of judgmental adjustment. Indeed, making such revisions to the model forecasts does often improve accuracy (Clements, 1995; Turner, 1990; Wallis & Whitley, 1991) though sometimes make forecasts worse (Fildes et al., 2009). One reason for this advantage is that judgmental adjustment enables omitted information to be incorporated into the forecasts (Bunn & Salo, 1996). For example, Wolfe and Flores (1990) showed that judgmental adjustment of statistical forecasts of earnings led to higher accuracy. Of course, the converse is also true: providing statistical forecasts to judgmental forecasters also improves their performance; as a result, decision support is generally beneficial for forecasting performance (Kremer et al., 2016).

### *2.9.3. Combining forecasts*

Combining forecasts is common in practice and is based on the rationale of broadened information of domain knowledge (Granger, 1989; Lobo & Nair, 1990), removing systematic biases (MacDonald & Marsh, 1994), and reducing random error (Kahneman et al., 2021). Fan et al. (1996) combined forecasts to predict movements of the Hong Kong stock market and found that combinations were superior to individual forecasts. Combined forecasts have also been found to outperform other approaches in the M1 to M5 competitions (Makridakis et al., 2021).

Goodwin et al. (2018) emphasise that forecast combination is most effective when the constituent forecasts are negatively correlated, so that combination increases the variety of information. Consistent with this, Graham (1996) pointed out that when the forecasts of economists are correlated, the effective number of independent forecasts is reduced. In summary, the benefits of combining forecasts are greatest when those forecasts are generated by different methods or on the basis of different theories (Batchelor & Dua, 1995).

### *2.9.4. Group judgment*

Group judgment also enhances accuracy because it increases the diversity of available information (the wisdom of crowds). For example, Ang and O’Connor (1991) and Sniezek (1989, 1990) found that interactive groups produce more accurate forecasts than nominal groups (i.e., simply averaging the individual pre-group judgements). However, the degree of improvement relies heavily on the way the group is structured; aggregation of judgments across group members is less important when they all have access to the same information (the use of a group is less useful) (Sniezek, 1990). This is in line with the finding that combining

variables/information from different sources improves accuracy more when different methods have been used to produce the individual forecasts (Clement, 1999).

Group discussion enhances confidence levels more than accuracy levels, thus leading to an increase in overconfidence. Brenner et al. (2005) found that people are more optimistic about predictions generated through group discussion than about those generated individually. The increased optimism arose because group members paid more attention to factors that prompted success. Other work has shown that there is a tendency of group members to reinforce each other's opinions (instead of raising scepticism) and that this causes an escalation in the groups' confidence (Janis, 1972; Heath & Gonzalez, 1995). However, as the group's diversity increases, the average probability forecast becomes more spread out and more under-confident (Hora, 2004; Ranjan & Gneiting, 2010).

#### *2.9.5. Feedback and advice*

Different types of feedback have been studied in order to improve forecasting performance. Feedback provides forecasters with information about the outcome or their performance. They are then able to evaluate that information and make modifications for the next forecast. Simple outcome feedback consists of presenting the true outcome whereas performance feedback provides people with a measure of how well they have performed. Petropoulos et al., (2017) found that performance feedback about bias (signed percentage errors) displayed individually for each of the most recent forecasts improved forecasting performance significantly. Legerstee and Franses (2014) showed that the accuracy of demand forecasts in a pharmaceutical company improved after forecasters received formal training and were provided with combined feedback types. Without that information, hindsight bias and confirmation bias occurred which impaired forecasting and forecasters' judgements of their own forecasting ability.

There has been a debate about the effectiveness of outcome feedback and cognitive feedback. Simple outcome feedback appears to be more effective in situations where the cue-outcome relations used as a basis for forecasting are fairly simple and straightforward (Tape et al., 1992) than in situations where more than three predictor variables have to be taken into account (Schmitt et al., 1977; Hammond et al., 1973). However, Buchheit et al. (2012) have proposed that outcome feedback and financial incentives complement each other to improve performance in relatively complex iterative tasks. Cognitive feedback is more highly processed information that informs people of particular aspects of their performance (e.g., their average bias). It may be more useful than outcome feedback when forecasts are based on a larger number of cues or on non-linear relations in MCPL tasks (Doherty & Balzer, 1988; Önkal & Muradoğlu, 1995; Fischer & Harvey, 1999).

Effectiveness of feedback has been examined for probabilistic forecasting (Benson & Önkal, 1992; Murphy & Winkler, 1984), interval forecasting (Bolger & Önkal-Atay, 2004), and density

forecasting (Önkal & Muradoglu, 1995). For example, in a dynamic stock price forecasting experiment, participants gave probability forecasts for the closing stock prices of 34 companies listed in the Istanbul Stock Exchange (Önkal & Muradoglu, 1995). The questions were formatted as a multiple interval task so that people gave their probabilities for each of a number of designated categories. They were randomly assigned to three feedback groups: (1) simple outcome feedback, (2) task-formatted outcome feedback, and (3) performance feedback about accuracy and calibration. Performance feedback improved forecast accuracy and was superior to outcome feedback. In addition, all forms of feedback improved the forecasters' calibration. But it is not clear if performance still improved if no feedback had been given (a control condition).

Feedback information is dependent on forecasters' past performance; advice or guidance is provided before their performance and regardless of their own judgment performance. However, forecasters put more weight on their own beliefs than on the advice that they are given (Bonaccio & Dalal, 2006; Yaniv, 2004). Also, advice has a stronger effect on improving accuracy and calibration of confidence judgments when people do not make their own judgment, perhaps because they have a reluctance to change their minds (Yaniv & Choshen-Hillel, 2012).

#### 2.9.6. Decomposition

Decomposition is a task structuring technique that can have beneficial effects when tasks are complex. Its aim is to overcome the limitations of human cognitive processing capacity and thereby help to produce accurate judgment performance (Goodwin et al., 2018). MacGregor (2001) has suggested that when the number of cues cannot be reduced, a complex judgment task can be decomposed into several simpler tasks. For example, a time series can be graphically decomposing into trend, seasonal, and residual components, and people can be asked for separate judgmental forecasts of each one. Edmundson (1990) found that this method led to more accurate point forecasts once these separate forecasts were re-integrated. Similarly, Lee and Siemsen (2017) compared two groups of newsvendor order decisions in three laboratory experiments. In the Task Decomposition (TD) treatment, participants prepared point forecasts, uncertainty estimates, and in-stock probability decisions for every period; these components were then mechanically converted into an order quantity using an aggregation mechanism that was defined *a priori* by an equation. In the Direct Order (DO) treatment, subjects directly entered an order quantity without receiving any instructions on the components and aggregation mechanism. As expected, results showed that the decomposition techniques used in the TD group led to performance improvements compared with the group that reported order quantities directly.

## 2.10. Techniques used to improve realism of confidence judgments

To overcome overconfidence, Russo and Schoemaker (1992) proposed five techniques: accelerated feedback; asking for counter-argumentation; paths to trouble (reasons); paths to future (conjunctions); awareness alone. Here I focus on the two most practical and commonly used techniques: feedback and elicitation questions.

### 2.10.1. Feedback

The usefulness of different types of feedback is has been explored in various tasks. A large number of studies (Stone & Opel, 2000; Lichtenstein & Fischhoff, 1980; Bornstein & Zickafoose, 1999) have reported that overconfidence in probability forecasting tasks is reduced by providing performance feedback and information about how forecasts relate to the cues on which they are based (i.e., functional validity information).

Outcome feedback has also been shown to improve the calibration of judgmental interval forecasts (Goodwin et al., 2011; O'Connor & Lawrence, 1989). Usually, judgmental prediction intervals are too narrow for the stated level of confidence, but outcome feedback is effective in widening them. Bolger and Onkal-Atay (2004) conducted a four-session experiment with three-day intervals between each session. Participants were asked to make one-step-ahead probabilistic interval forecasts for each of 32 graphically presented time series. Three days later, they received feedback about their performance and were required to make forecasts for new series. Results showed calibration improved over sessions due to the better ability to allow for the variability in time series. Participants achieved good calibration simply by increasing the width of the confidence intervals they gave, thereby improving their hit rates. The difference between the outcome and the original forecast draws attention to the degree of the uncertainty that is associated with the forecasting task (Goodwin et al., 2011), thereby producing performance improvements. However, Teigen and Jørgensen (2005, Experiment 5) showed that interval predictions improved with feedback rather slowly. Furthermore, they found an asymmetric effect of feedback: it rendered overconfident participants slightly less certain but under-confident and well-calibrated participants much more confident.

Apart from task type, the effectiveness of feedback also depends on aspects of the question set. It is not effective when items in the question set are not related (e.g., general knowledge questions) (Keren, 1991) but it is effective when they are related. As an example of the latter case, forecasters learn to be well-calibrated, such as bridge players (Keren, 1987), weather forecasters (Murphy & Winkler, 1984) and horse race bettors (Johnson & Bruce, 2001). The availability of consistent, timely, and applicable outcome feedback provides people with an excellent opportunity for learning to be better calibrated (Winman & Juslin, 1993).

Feedback cannot reduce overconfidence when people never experience 'errors'. Dawson et al. (1993) required physicians to forecast three measures of heart functioning immediately

after they had inserted a catheter into their patients and to assign confidence levels to their estimates. Feedback was inefficient in the situation where their prior predictions were not clearly remembered; instead, feedback acted to heighten confidence. Specifically, participants were not able to benefit from their forecast errors (do not remember/know the distance between their original predictions and true outcomes). Therefore, increased confidence was due to the manifestation of the hindsight bias instead of learning from feedback. Nevertheless, once the learning effect happens, Hoch and Loewenstein (1989) showed that feedback can be very helpful even to those forecasters who distort the feedback because of the hindsight bias: learning from feedback can still pierce the bias to inform the forecaster of the relative difficulty of various tasks.

In a word, in order to achieve a better calibration in judgmental forecasting, feedback information should encourage and facilitate learning underlying the whole forecast process.

### *2.10.2. Elicitation question*

Quality of calibration varies with different types of question format (Juslin et al., 1999; Seaver et al., 1978; Soll & Klayman, 2004; Teigen & Jørgensen, 2005). Using single limit estimates produces much wider intervals than the traditional range method (Teigen & Jørgensen, 2005). Employing the SPIES method by asking forecasters to consider a complete picture of possibilities (probability distribution) increases calibration (Haran et al., 2010). Interval evaluation, based on probability judgments of fixed intervals, produces less overconfidence than interval production (Winman et al., 2004).

These findings can be taken as support for knowledge sampling theory (Klayman et al., 2006). According to this approach, the more times we sample our knowledge to create an estimate, the less our overconfidence in that estimate. Speirs-Bridge et al. (2010) combined the results of Soll and Klayman (2004) and Teigen and Jørgensen (2005) to examine reductions in overconfidence related to question format. In three experiments (in the domains of infectious diseases and marine ecology, they asked experts to write down their estimates using either a three-point elicitation procedure (upper limit, lower limit and best guess) (Soll & Klayman, 2004) or a four-step procedure (by adding an additional step that was to rate their anticipated confidence in the interval produced - interval evaluation) (Teigen & Jørgensen, 2005). Meta-analysis of overall results showed that 80% intervals derived from the 4-step elicitation procedure showed significantly less overconfidence (11.9%) than 3-point intervals. This indicates that the more steps of there are in the elicitation procedure, the more accurate people's estimates of their uncertainty in their forecasts. This finding may form the basis for developing techniques that allow people to express their uncertainty more accurately.

In summary, different factors affecting people's judgmental forecasting accuracy and confidence judgment have been identified. However, there are few studies investigating how these factors influence inflation expectations produced by lay people and how lay inflation

expectations can be improved. In the last two experimental chapters, I examined (1) the role of outcome feedback in improving the performance of inflation judgment accuracy and the confidence judgments in them, and (2) the differences in forecast performance between different types of elicitation methods (point forecast, interval forecast and probability density forecast).

## Part 2 Experimental chapters

### Introduction

Inflation is regarded as a self-reinforcing process at an aggregate level. Theoretical frameworks and economic models have been developed to understand the way lay people seek relevant information and form their inflation judgments. Systematic patterns of over-estimation and high heterogeneity of their responses have been identified. Some researchers have attributed these phenomena to factors such as the over-weighting of personally experienced price changes, media amplification and demographical and socio-economic characteristics. However, their explanations for them have not been consistent and so the challenges of accounting for them still exist. Though the utilization of heuristics in human judgment and decision-making and the occasional biases that result have been acknowledged, there appear to be no studies that have considered them as the underlying psychological mechanisms responsible for producing laypeople's inflation expectations. They could, however, produce inaccurate survey responses to questions about those expectations.

After examining the current literature in the fields of inflation expectation and judgmental forecasting, my review revealed that no study has been conducted that a) investigates how lay people form expectations about inflation; no cognitive model has been developed to explain how people process price change information stored in their memories to produce inflation expectations, b) examines how well lay expectations are measured in surveys and reveals reasons for the differences that have been observed in the responses obtained from lay and professional surveys, c) explores ways of improving measurement of inflation expectations, specifically, by use of training (giving outcome feedback) as a means of reducing judgment errors and improving confidence judgment components, and d) identifies reasons for differences in expectations produced by the three commonly used response formats.

### Overview of experiments

The overall objective of the thesis is to contribute to the literature in economic psychology on factors that affect laypeople's inflation expectations by addressing the four issues listed in the previous section. This will involve discussion of a number of previously identified cognitive processes, including use of judgment heuristics, memory priming, 'wisdom of crowds' effects, context effects, overconfidence in judgments and the use of outcome feedback.

## Chapter 3 How lay people form expectations about inflation

### 3.1. Introduction

Lay expectations of inflation are recorded in surveys such as the Michigan Survey of Consumers (MSC), the Federal Reserve Bank of New York's Survey of Consumer Expectations (SCE), and the Bank of England's Inflation Attitudes Survey (IAS). Central banks need to know about these expectations because they are assumed to determine future levels of inflation<sup>4</sup>. In other words, they feed into the inflation forecasts made by the banks and therefore influence policy. There are various ways in which lay expectations of inflation could influence actual levels of inflation. Two examples will suffice. First, the more that consumers expect inflation to increase, the more likely it is that they will decide to bring forward their purchases of durable goods; the resulting increase in demand will push up prices for those goods. Second, the more that wage earners expect their cost of living to increase, the more effort they will put into securing increases in their wages; in some cases, these efforts will be successful and result in higher costs for producers, who consequently then need to increase the price of their products.

According to rational expectations theory (Muth, 1961), rational economic agents form their expectations in line with what macroeconomic theorists specify as rational. Hence the inflation expectations of experts and lay people should be the same. However, they are not. Sometimes median inflation expectations obtained from lay surveys have turned out to be more accurate than those obtained from professional surveys (Ang et al., 2007; Hafer & Hein, 1985). More often, lay expectations are less accurate and much more heterogeneous than those of experts (Mankiw et al., 2003; Palardy & Ovaska, 2015). Research into possible causes of the differences in accuracy has been reviewed by Niu and Harvey (2022a).

Unlike experts, lay people may draw on their own personal experience of price changes when forming their expectations for future levels of inflation. This means that differences in personal experience are likely to contribute to the high degree of heterogeneity observed in lay expectations of inflation (e.g., Jungermann et al., 2007; Malmendier & Nagel, 2016). People who judge price rises for specific products to be high would expect inflation to be high, thus leading to a correlation between judgments of specific price rises and judgments of inflation.

I report three experiments designed to test these predictions. In each one, participants completed two judgment tasks. In the direct inflation estimation task, they estimated the average price change for the upcoming year (expected inflation)<sup>5</sup>. In the specific price

---

<sup>4</sup> King (2022, p3), a former Governor of the Bank of England, has criticised central banks' over-reliance on "a family of theoretical models which rely on the assumption that expectations drive inflation, and central banks drive expectations" (Woodford, 2003), so that "inflation in the long run is determined by the official inflation target".

<sup>5</sup> In some experiments, they also estimated current inflation.

assessment task, they answered a separate set of questions for each of the 12 consumer product divisions on which the Consumer Price Index (CPI) is based. In each set, they were asked to assess the average cost of a single payment in the category in the current year, in the previous year, and in the following year. Answers to these questions were used to calculate the average expected inflation rate across all 12 categories. This indirect derivation of expected inflation rate was then compared with the rate obtained from the direct estimation task.

The first experiment, carried out in the summer of 2019 produced results that were not compatible with those previously reported by Bruine de Bruin, Van der Klaauw, et al. (2011) and which I discuss in detail in the section below. The most obvious difference between my experiment and theirs was that mine was carried out when inflation across all 12 product categories contributing to the CPI was uniformly low whereas theirs was carried out when inflation in some of those categories was very high. An opportunity to determine how important this difference in the inflation environment was in producing the difference in results arose in the summer of 2022 by which time inflation had risen markedly. I repeated my experiment twice and found that, on both occasions, results were indeed different from those that I obtained in 2019. Together, findings from my experiments and those of Bruine de Bruin, Van der Klaauw, et al. (2011) indicate that the way that lay people form their estimates of inflation depends on the prevailing inflation environment. In the next section, I discuss three ways in which these estimates can be made.

### *3.1.1. Three models of the development of expectations of inflation*

Do people use their memory for information about past and current prices of different categories of product to estimate future inflation? Broadly speaking, there are three possibilities. First, they do not. Instead they use information that they have gained from the media and other secondary sources. This is a price-free model of the basis of expectations about inflation. Second, they do use information drawn from memory about past and current prices to estimate price rises across all different categories of product or across a representative sample of those categories. This is a price-recall model of the formation of expectations about inflation. Third, they identify highly salient (typically the largest) price rises that they have recently noticed and use this information to develop their expectations of future inflation. This is a price-salience model of inflation expectations: specific price information is used but it is restricted to selected non-representative categories of consumer product.

### *3.1.2. Review of previous work*

One study reported by Bruine de Bruin, Van der Klaauw, et al., (2011) contained three conditions and involved a priming manipulation. In the first condition, the initial priming task required participants to think of *any* price change that they had noticed over the past 12 months, to specify the service or good that they had thought of, and then to estimate the

percentage change in the price of that service or good. In the second condition, participants were asked to think of the *largest* price change that they had noticed over the past 12 months, to specify the service or good that they had thought of, and then to estimate the percentage change in the price of that service or good. In the third condition, participants were asked to think of the *average* change in prices that they had noticed over the last 12 months and to estimate the percentage change in that average price (without specifying what goods and services they had included in the average).

After the priming task, all participants expressed their inflation expectations for the next 12 months using the “prices in general” question employed in the Michigan Survey. Specifically, they were first asked “during the next 12 months, do you think that prices in general will go up, or go down, or stay where they are now”. Then, if they had said they expected prices would go up or go down, they were further asked “by what percent do you expect prices to go [up/down] on the average, during the next 12 months”. Bruine de Bruin, Van der Klaauw, et al. (2011) measured the median absolute deviation of inflation expectations from 2%, the Federal Reserve’s implicit inflation target and the European Central Bank’s explicit targeted inflation ceiling. This was 3.50% in the first condition (primed with thinking about any price rise), 5.50% in the second condition (primed with thinking about the largest price rise), and 2.00% in the third condition (primed with thinking about the average price rise): estimates in the first two conditions did not differ significantly from one another but they were both significantly higher than estimates in the third condition. Thus, recall of a *single* price rise (i.e., any one price rise or the largest price rise) produced a priming effect. Also, those who recalled larger price changes in the priming task tended to produce inflation expectations that were further from 2%. This effect explained differences in inflation expectations between the first two conditions and the third one.

Bruine de Bruin, Van der Klaauw, et al. (2011) acknowledge that their research leaves some questions unanswered<sup>6</sup>. It is notable that, in a second study in which no priming task was performed, they found that the extremeness of inflation expectations averaged across all 92 participants (i.e., 2.00%) was identical to the extremeness of inflation expectations in the first study when the priming task was to think about think about the average price rise (i.e., 2.00%). One way of interpreting this is to conclude that being required to think about the average price rise brought to mind no specific price rises but was based on media reports (the price-free model). Another possibility is that it led people to sample price rises in a representative set of consumer product categories (the price-recall model). In either case, I would expect no priming effect after judging the average price rise.

Bruine de Bruin, Van der Klaauw, et al. (2011, p. 839) pointed out that their results suggest that “memories for the past year’s changes in prices are biased towards those goods and

---

<sup>6</sup> For example, they point out that: “Another limitation of Study 1 is that it provides no insights into whether or not participants who were instructed to recall the average change in prices tried to remember specific changes in prices” (Bruine de Bruin, Van der Klaauw, et al., 2011, p 840).

services that have shown the largest price changes, affecting the extremeness and dispersion of subsequently reported inflation expectations". In other words, they were consistent with the price-salience model.

### **3.2. Experiment 1**

This experiment was designed primarily to test hypotheses arising from the price-free and price-recall models but certain patterns of results could have implications for the price-salience model. Participants in one condition produced inflation expectations with no prior priming task. Those in a second condition produced those expectations only after completing a priming task in which they answered questions about the previous, current and future cost of items in all 12 consumer product divisions on which the CPI is based.

Neither the price-free model nor the price-recall model predicts a priming effect. Inflation expectations should be similar in the two conditions. According to the price-free model, I would not expect a difference between them because estimating inflation is not based on recall of specific prices and so not influenced by them. According to the price-recall model, I would not expect a difference between the two conditions because estimates of overall inflation depend on recall of all specific price rises anyway; priming people by asking them to recall those price rises would merely be asking them to carry out the processing that they would carry out anyway to estimate inflation.

If a priming effect is obtained such that overall inflation is expected to be higher after answering questions about prices in all 12 consumer product categories, it could be explained by the price-salience model. I would need to assume that going through all 12 categories highlights the one with the largest price change as particularly salient and that price changes in this single category are used to estimate overall inflation.

According to the price-recall model, people normally use an average of all price rises (or a representative sample of them) to generate their inflation expectations. This suggests that I should expect there to be no difference between people's inflation expectations and the average of all price rises that they had produced in the priming task. There should also be a correlation between these two measures: people whose price rise judgments are, on average, high should also have high expectations of inflation. This is because they use the former to produce the latter.

In contrast, if the price-free model holds and inflation expectations are not normally based on recalled price rises for specific categories of product, there is no reason to expect that overall inflation expectations will be the same as the judged price rises for the following 12 months averaged across all 12 consumer product divisions. Any difference between these two measures would be consistent with the price-free model but not be expected with the price-recall model. Furthermore, finding no correlation between these two measures would also

be consistent with the price-free model (though this model could also accommodate such a correlation).

In summary, consistent with the price-free model but not with the price-recall model, I hypothesise that there will be a significant difference between the mean of inflation expectations and mean of price rise judgments across all 12 consumer product divisions on which the CPI is based ( $H_1$ ). Consistent with the price-recall model, I hypothesise that there will be a significant correlation across individual participants between the mean of inflation expectations and mean of price rise judgments across all 12 consumer product divisions on which the CPI is based ( $H_2$ ). Finally, consistent with the price-salience model but not with either the price-free model or the price-recall model, I hypothesise that judgments in the inflation estimation task will be affected when people first perform the specific price assessment task ( $H_3$ ).

### 3.2.1 Method

The experiment used a mixed design with a within-participant and a between-participant variable. Task type was varied within-participants: the tasks were a) estimating overall inflation for the current and following year, and b) making five judgments, including assessments of price changes, about each of the 12 consumer product divisions on which the CPI is based. I term these tasks inflation estimation and specific price assessment. The between-participant variable was task order: half the participants performed the inflation estimation task followed by the specific price assessment task and the rest performed them in the opposite order.

*Participants* One hundred and twenty-three participants (76 female, 47 male) were recruited for the study. Their mean age was 25 years ( $SD = 8$  years). Twenty-one were recruited from the participant pool at University College London and given 0.25 credits for taking part; 50 were recruited from China via Qualtrics.com and paid 3 RMB; 52 were recruited via Prolific.com and paid £1.25. Thirty-four participants came from countries other than the UK and China: Australia (1), Austria (1), Bulgaria (1), Canada (4), Germany (1), Greece (1), Italy (3), Mexico (7), Portugal (8), USA (7)<sup>7</sup>. Participants were randomly assigned to the two conditions. This resulted in 62 of them being in the condition in which specific price assessment followed inflation estimation and 61 of them being in the condition in which inflation estimation followed specific price assessment. Data were collected between 1st June 2019 and 31 August 2019.

*Stimulus materials* In the inflation estimation task, participants were initially provided with a simple definition of inflation and price change. They then completed two inflation estimation tasks. One asked them to provide their inflation expectations: “What is your prediction of the average price change (as a %) from the year June 2018-June 2019 to the year June 2019-June

---

<sup>7</sup> Inflation estimates in Western and Eastern (Chinese) subsamples were not significantly different. See Appendix 2.

2020. (Note: You should not add a percentage sign after your estimate but you can use a decimal point.)” The other asked them to provide their estimate of current inflation: “What is your estimate of the average price change (as a %) from the year June 2017-June 2018 to the year June 2018-June 2019.”

In the specific price assessment task, participants answered 60 questions. These comprised five questions in each of the 12 consumer product divisions on which the CPI is based. These divisions were extracted from the Consumer Price Indices Technical Manual published by the UK Office for National Statistics. Each screen in this task was headed by a label corresponding to one of the CPI consumer product divisions. Underneath each heading, there was a specification of the items that the category included. For example, underneath the heading “Item: Clothing and footwear”, participants were told “The category includes Garments (e.g., Coat, Trousers – formal, casual, Casual jacket, Jeans, Various shirts, T shirts, Swimwear, Exercise leggings, Underwear, Tights, Nightwear), Man’s tie, Knitting wool, Lady’s scarf, Cycle helmet, Cleaning, repair, and hire of clothing, Shoes, Boots, Training shoes, Sandals”. After this, participants answered the following five questions for each product division: 1) How many times do you buy this item per month, 2) What is the average cost of a single payment this year (June 2018-June 2019)? 3) What was the average cost of a single payment last year (June 2017-June 2018)? 4) What will be the average cost of a single payment next year (June 2019-June 2020)? 5) What percentage of your knowledge of prices in this category comes from your personal experience of them rather than hearing about the prices from other people or the media? Respond using the sliding scale below where 0% is all knowledge is from other people or the media and 100% is all knowledge is from my personal experience.

The experiment was programmed in two versions: English and Chinese. To ensure that these were comparable, an initial translation was followed by an independent back-translation which was then matched to the original version. This process was iterated until a good match was obtained.

*Design* Participants were randomly assigned to one of the two conditions with the constraint that there were an equal number of people in each one. The order of the two questions in the inflation estimation task was randomised separately for each participant. The order of the 12 consumer product divisions in the specific price assessment task was also randomised separately for each participant.

*Procedure* Participants accessed a link to the experiment via the Qualtrics website. Before the start of the study, each one was asked whether they had lived in the same country for the last two years (i.e., from June 2017 to the date of the participant’s response). Those who had not done so were excluded from the study. The participants who remained were then shown an information screen that described the experiment in broad terms and a consent screen that gave them details of the ethical permission for the study and elicited their consent for taking part in it. They completed the two tasks, filled in demographic details, and were debriefed.

*Data processing* I entered people's judgments in the inflation estimation task directly into my analyses. To obtain inflation estimates from price judgments in the specific price assessment task, two steps were required. To produce an estimate of expected inflation for each of the 12 consumer product divisions, the first step was to subtract the assessed price for the current year in a particular category from the assessed price for the next year in that category and then divide this difference by the assessed price for the current year in that category. To obtain a percentage estimate for expected inflation in that category, the result of this calculation was multiplied by 100<sup>8</sup>.

A second step was needed to produce an overall estimate of inflation from the individual estimates of inflation for each category. To do this, I calculated the average of the separate estimates from the 12 categories, each weighted by the individuals' estimates of their frequency of purchasing items in that category. The frequency-weighted average is a more appropriate measure than other alternatives such as using equal weights or weights based on expenditure share. This is because it is derived from people's experience of recent price rises contingent on their actual purchasing behaviour and because D'Acunto, Malmendier, Ospina, et al. (2021, p. 1615) have shown that the "weights consumers assign to price changes depend on the frequency of purchase, rather than expenditure share"<sup>9</sup>.

Estimates of inflation for some participants were extremely high. For example, one of them made price assessments that resulted in estimates of current and future values of inflation of 1022% and 1586%, respectively. Another made direct estimates of current and future inflation of 105% and 100%, respectively. Bruine de Bruin, Van der Klaauw, et al. (2011) also report some unrealistically high estimates of inflation. I excluded participants who produced inflation estimates and purchasing frequencies beyond 3.0 standard deviations from the mean value. After exclusion according to this criterion, my sample comprised 92 participants (53 female, 39 male) with a mean age of 26 years (SD = 9 years). Forty-four were in the condition in which people performed assessed specific price assessment task before the inflation estimation task and 48 were in the condition in which tasks were performed in the opposite order. It is, however, worth emphasising that the effects that were significant when outliers were excluded were identical to those that were significant when they were not excluded.

To examine how their experimental manipulations affected the quality of participants' current inflation estimates, Bruine de Bruin, Van der Klaauw, et al. (2011) extracted the absolute deviation of those estimates from 0%. This measure showed a high correlation with the absolute deviation of the estimates from the actual CPI (1.1%) in the 12 months preceding their survey. To obtain corresponding measures of the quality of participants' inflation

---

<sup>8</sup> An analogous procedure was used to extract an estimate of current inflation for each of the 12 consumer product divisions.

<sup>9</sup> I also analysed the simple averages of the separate estimates from the 12 categories. Results of those analyses were not substantially different from those that I report here.

expectations, they calculated the absolute deviation of the participant's estimates from 2%. (This was both the Federal Reserve's implicit inflation target and the European Central Bank's explicit targeted inflation ceiling.) I was in a different position.

When I performed data analyses, official inflation rates for all countries from which my participants were drawn had been published and were obtained from the website [inflation.eu](http://inflation.eu), the Office for National Statistics (UK), and the Australian Bureau of Statistics (Australia). I averaged the 12-month inflation rates published each month over the 13-month period (e.g., June 2018 - June 2019) for which participants were asked to estimate inflation to obtain criteria against which participants' judgments were assessed. (I averaged over 13-month rather than a 12-month period because my instructions did not specify a date within the month of June. In practice, there is minimal difference between criteria based on 12-month and 13-month periods.)<sup>10</sup>

I extracted two error measures for each participant. The constant error or bias was the arithmetic difference obtained by subtracting the official inflation rate (CPI) from the participant's estimate. The absolute error was calculated as the absolute deviation of the participant's estimate from the CPI.

### 3.2.2. Results

Here I report analyses needed to test my hypotheses. For additional analyses of this and later experiments, please see Appendix 2.

I compared direct inflation estimates with those obtained by taking the frequency-weighted average of inflation estimates derived from specific price assessments in the 12 product categories. I carried out three three-way mixed model analyses of variance (ANOVAs) on judgments, constant errors and absolute errors, each of which used Task Order (inflation estimation first versus specific price assessment first) as a between-participants factor and Task Type (inflation estimation versus specific price assessment) and Judgment Type (current inflation estimate versus expected inflation) as within-participant factors. There was only a main effect of Task Type in the analyses of Judgment ( $F(1, 90) = 20.77$ ;  $p < 0.001$ ,  $ges = 0.0595$ ) and Constant Error ( $F(1, 90) = 20.77$ ;  $p < 0.001$ ,  $ges = 0.0594$ )<sup>11</sup>. There was a main effect of Task Type on Absolute Error ( $F(1, 90) = 41.47$ ;  $p < 0.001$ ,  $ges = 0.1108$ ) and an interaction effect of Task Type and Task Order ( $F(1, 90) = 3.96$ ;  $p = 0.05$ ,  $ges = 0.0118$ ). However, simple effect analyses showed an effect of Task Type both when inflation estimation occurred first ( $F(1, 47) = 31.41$ ,  $p < 0.001$ ,  $ges = 0.2325$ ) and when specific price assessment occurred first ( $F(1, 43) = 11.86$ ,  $p = 0.001$ ,  $ges = 0.0992$ ). The simple effect of Task Order was not found for

---

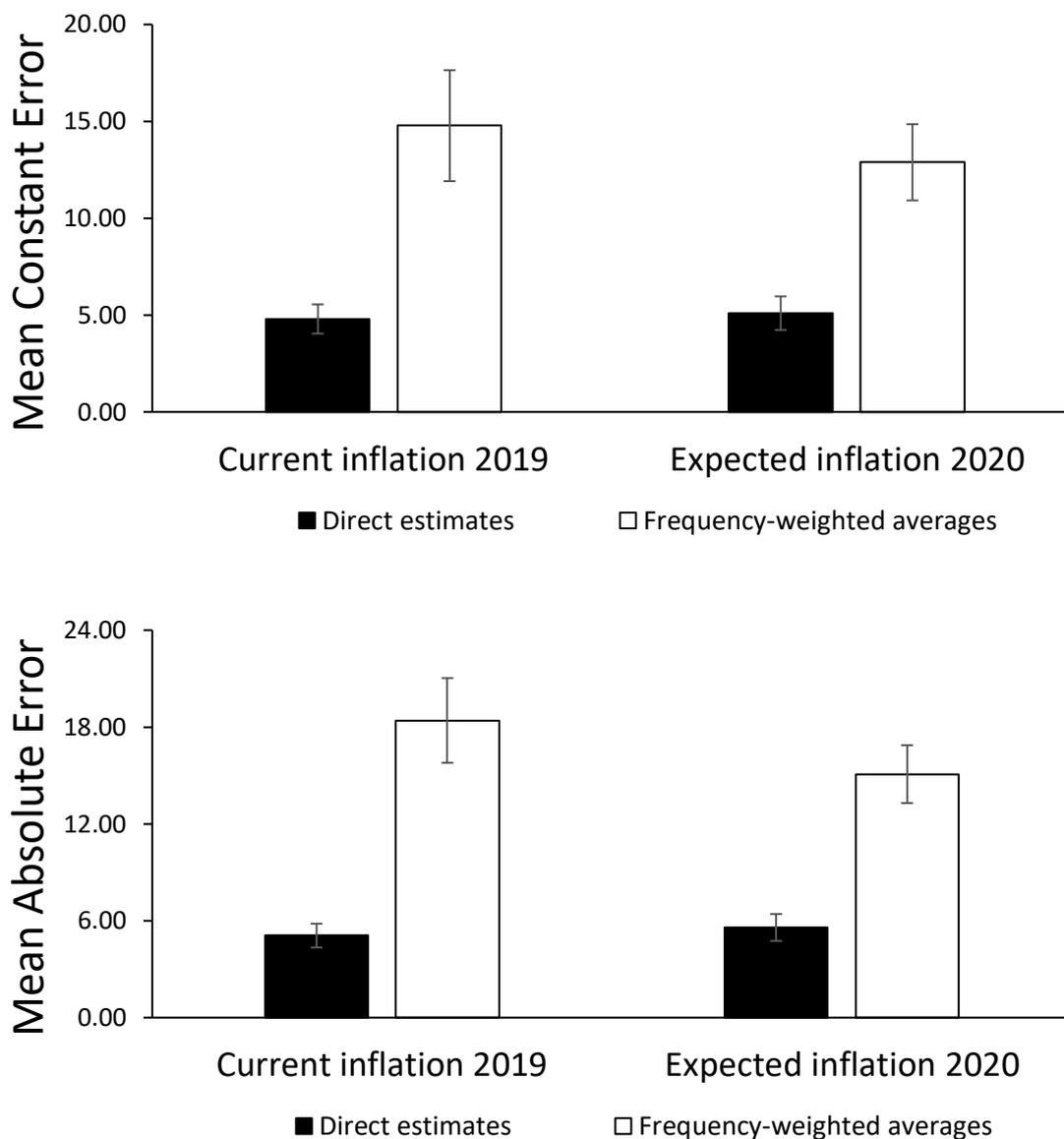
<sup>10</sup> Australia, where one of my participants lived, publishes annual inflation rates quarterly but not monthly. Criteria for this participant were obtained by averaging over four quarters, July – September, October – December, January – March, and April – June.

<sup>11</sup> I used generalised eta squared ( $ges$ ) to measure effect size (Olejnik & Algina, 2003).

either Task Type. Effects of Task Type on Mean Constant Error and Mean Absolute error are shown in Figure 3.1.

**Figure 3.1**

*Experiment 1: Effects of Task Type on constant errors and absolute errors in judgments of current and expected inflation.*



*Note.* The figure shows the errors in frequency-weighted averages of inflation estimates derived from assessments of specific prices and in direct estimates of inflation.

Separately for judgments of current and future inflation, I then calculated correlations across all 92 participants (i.e., ignoring task order) between absolute error of direct inflation estimates and absolute error of inflation estimates derived indirectly from specific price

assessments. Neither of these correlations was significant. I repeated this procedure but using measures of constant error. Again, neither correlation was significant.

### 3.2.3. Discussion

Figure 3.1 shows a clear difference in the inflation rate that people directly estimated and the one that was derived from their estimates of prices in all product categories ( $H_1$ ). Also, errors were much greater in the second case. There were no significant correlations between inflation rates or errors in those rates across the two tasks for judgments of current inflation and future inflation ( $H_2$ ). Also, there was no evidence that first assessing the levels of prices in different years for each category of goods or services that contribute to the CPI influenced later direct estimates of inflation ( $H_3$ ). This combination of results is more consistent with the price-free model of inflation estimation than with the price-recall or price-salience models.

Direct estimates of overall inflation were somewhat too high. Why was this? Even if media reports and memory for prices are reasonably accurate and people have some ability to extrapolate from the past into the future, the price-free and price-recall approaches can still produce expectations of inflation that are too high. This is because people expect inflation rate (as distinct from prices) to go up even when it does not do so. Niu and Harvey (2022a) found that when separate groups of people estimated current and future inflation rates, the rates they gave were not significantly different. However, when the same group of people estimated both current and future inflation rate, their mean estimate for the future inflation rate was significantly higher than that for the current rate. Niu and Harvey (2022b) suggest that people's use of the availability heuristic leads them to expect inflation to rise: media coverage of potential and actual rises in inflation is much more extensive than corresponding coverage of falls in inflation. This is an example of media-based risk-amplification.

Dietrich, Knotek, et al. (2022) reported results of a survey carried out between July 2020 and September 2021. They found that direct estimates of expected inflation were higher than indirect estimates obtained by aggregating expectations across product categories. They argue that this difference shows that aggregate (i.e., direct) estimates of inflation are not derived from a weighted average of beliefs about inflation in each category. Specifically, "the heuristics involved in expressing aggregate expectations may differ from the processes underlying *aggregated* inflation expectations" (p. 25). Thus, Dietrich, Knotek, et al. (2022) argue that both direct and indirect (aggregated) inflation expectations are derived from respondents' personal experience with prices but via different cognitive processes. But, in their study, direct estimates of inflation were higher than inflation expectations in any one product category. What sort of heuristic based on experienced price information could have produced such a result? Dietrich, Knotek, et al (2022, p. 25) simply say that it must be nonlinear. I would argue that, like us, they have produced evidence consistent with the price-free model. Direct estimates of inflation in their study were based on recall of media reports and other secondary sources. During the period of their study, the experts whose views

formed the basis of such reports considered that inflation would rise faster than the price experience of lay people indicated.

### **3.3. Experiment 2a and 2b**

Bruine de Bruin, Van der Klaauw, et al. (2011) found that first making a judgment about a single specific price change over the previous 12 months ('any price change' or 'the largest price change') resulted in higher and significantly less accurate inflation expectations for the next 12 months. However, first asking people to provide the average change in prices over the previous 12 months did not have that effect. I have shown that asking people for many specific price changes also does not have that effect. This implies that the original finding was not due to the price specificity of the priming task but to the fact that just a single price was required. This is consistent with Bruine de Bruin, Van der Klaauw, et al.'s (2011, p 835) view that their original finding was produced by an anchoring effect and is what is to be expected according to the price-salience model.

Findings from Experiment 1 are consistent with the price-free model but, as I have seen, Bruine de Bruin, Van der Klaauw, et al.'s (2011) results were consistent with the price-salience model. What could have caused this difference? There are two possibilities. First my experiment was carried out when inflation rates were low across all product categories: for the UK and China, overall inflation rates were 2.19% and 2.20% respectively. The inflation environment was very different when Bruine de Bruin, Van der Klaauw, et al. (2011) carried out their work. In the 12 months before they released their surveys in August 2010, overall inflation was low at 1.1% but, for certain commodities, it was very high. For fuel oil, it was 10.6% and for used cars and trucks, it was 15.5% (Bureau of Labor Statistics, 2010a). Twelve-monthly Inflation rates for these and related commodities had been high and changing for some time; for example, corresponding figures in March 2010 for fuel oil and for used cars and trucks were 27.2% and 16.3%, respectively even though overall inflation rate was just 2.3% (Bureau of Statistics, 2010b). In situations such as this, it is likely that consumers would closely monitor prices only in those categories where they showed saliently high increases. Information obtained from this limited price monitoring may have then primed (biased) their overall estimates of inflation via use of an anchor-and adjust heuristic just as Bruine de Bruin, Van der Klaauw, et al. (2011) suggest.

There is a second possibility. In my experiment, respondents were given a definition of inflation in terms of price change and then asked for their expectation of the average price change (as a %) for the next year. In contrast, Bruine de Bruin, Van der Klaauw, et al.'s (2011) participants reported their year-ahead inflation expectations using the Michigan Survey's "prices in general" question. Bruine de Bruin et al. (2012) found that inflation expectations were lower when respondents were asked to estimate inflation than when they were asked to estimate "prices in general" or "prices you pay". In addition, participants reported thinking more about "prices of things you usually spend your money on" and less about "the US

inflation rate” when answering questions about “prices in general” or “prices you pay” than when answering questions that specifically included the term “inflation”.

The questions in Experiment 1 did not include the term “inflation”; they asked respondents to estimate “the average price change”. Of the three question formats that Bruin de Bruine et al. (2012) compared, this appears most similar to the one that asked people to estimate changes in “prices in general”. Any difference between asking for the average price change across all products (Experiment 1) and the average change in prices in general (Bruine de Bruin et al., 2011) is subtle indeed. Hence I would not expect that the difference between my results in Experiment 1 and those of Bruine de Bruin, Van der Klaauw, et al. (2011) arose from this minor difference in question format. However, given Bruine de Bruin et al.’s (2012) results, it remains possible that use of a question format that specifically includes the term “inflation” would lower direct estimates of inflation, reduce recall of specific prices, and thereby increase evidence in favour of the price-free model.

An opportunity to test the hypothesis that the prevailing inflation environment affects how people estimate inflation arose in summer 2022. At that time, CPI had risen sharply first to 9.1% in May 2022 and then to 10.1% in September 2022. It is reasonable to expect that when inflation rate is so high, changing rapidly and affecting many product categories, people monitor price changes in either all those categories or in those categories showing particularly large increases. In other words, results should be consistent with the price-recall or the price-salience model.

To examine whether this would be so, I carried out two versions of an experiment and tested the same hypotheses as before. In Experiment 2a, I employed the second question format used by Bruine de Bruin et al. (2012) in which questions specifically mentioned the term “inflation”. In Experiment 2b, I employed the same question format that I used in Experiment 1. Thus differences between Experiment 1 and Experiment 2b can be attributed to the difference in inflation environment and differences between Experiments 2a and 2b are likely to arise from differences in question wording. In both experiments, I test the same hypotheses examined in Experiment 1.

In addition, I tested two further hypotheses. To test the hypothesis (H<sub>4</sub>) that initially assessing prices for all 12 categories of consumer product does not influence the later estimation of inflation whereas assessing the price of a single product category does do so, I varied the priming task in different experimental conditions. I compared the effects of priming with the task of estimating prices in all 12 product categories (as in Experiment 1) with a no-priming control condition and with the effects of priming with a task of estimating prices in a single product category (as in Bruine de Bruin, Van der Klaauw, et al., 2011).

To examine the role of memory in identifying the product category with the largest price change, I used two versions of the ‘single product category’ priming task. In the first of these,

the 12 product categories were displayed and participants chose (i.e., recognised) the one with the largest price changes before estimating prices in that category. In the second, the categories were not displayed but participants simply entered (i.e., recalled) the name of the product category with the largest price changes before estimating prices in that category. This second 'recall' version corresponds closely to the 'largest price change' condition examined by Bruine de Bruin, Van der Klaauw, et al. (2011). This allowed me to examine whether explicitly providing categories in the 'recognition' version of the task facilitates their identification of a category with a higher price change. If it does, this would result in higher inflation estimates in the 'recognition' version of the task than in the 'recall' version (H<sub>5</sub>).

### 3.3.1. Method

In both Experiment 2a and 2b, there were four conditions. In the no-priming control condition (Condition 1), people simply carried out the main task of estimating overall inflation rates. In the other three conditions, they carried out a priming task before this main task: a) in the 12-product category priming task (Condition 2), they estimated current and future inflation in each of those categories; b) in the 'recognition' single product category priming task (Condition 3), they were provided with the list of the 12 product categories, chose the one that had shown the largest price rise over the previous 12 months, and then estimated current and future inflation in just that category; c) in the 'recall' single product category priming task (Condition 4), they were asked to identify the product category that had shown the largest price rise over the previous 12 months, and then estimated current and future inflation in just that category.

*Participants* In Experiment 2a, 225 participants were recruited via Prolific.com. Four of them were excluded because they provided incorrect answers to three simple attention screening questions (e.g., what colour is the sky with options of orange, blue, green and red). The remaining 221 participants (151 females, 69 males, 1 prefer not to say) were paid £1 for their participation. All of them were from UK. Their mean age was 41 years (SD = 15 years). Data were collected between 9 May 2022 and 13 May 2022.

In Experiment 2b, 260 participants were recruited via Prolific.com. Three of them were excluded because they provided incorrect answers to two simple attention screening questions. The remaining 257 participants (153 females, 100 males, 4 prefer not to say) were paid £1 for their participation. All of them were from UK. Their mean age was 29 years (SD = 9 years). Data were collected between 23 September 2022 and 29 September 2022.

*Design* In both experiments, participants were randomly allocated to one of four conditions. In Experiment 2a, 56 participants allocated to Condition 1, 55 to Condition 2, 56 to Condition 3, and 54 to Condition 4. In Experiment 2b, 58 participants allocated to Condition 1, 82 to Condition 2, 61 to Condition 3, and 56 to Condition 4. In Condition 2 of both experiments, the order of the 12 categories was randomized separately for each participant.

*Stimulus materials* In both experiments, participants carried out the main inflation estimation task that required them to forecast overall inflation. In Experiment 2a, they were asked: “During the next 12 months, do you think that inflation will go up, go down, or stay where it is now?” and, then, those who answered that it would go up or would go down were further asked to state the percentage by which it would change. In Experiment 2b, they were asked: “What is your prediction of the average price change (as a %) from the year September 2021-September 2022 to the year September 2022- September 2023?”

In the specific price assessment task in Condition 2 of Experiment 2a, participants were presented with the 12 product categories one at a time (in a different random order for each participant) and, for each category, they were asked 1) whether prices they paid over the past 12 months had gone up, gone down or stayed the same and, if they had changed, by what percentage they had changed and 2) whether prices they thought they would pay over the next 12 months would go up, go down or stay the same and, if they thought they would change, by what percentage they expected them to change. After they had answered these questions for each of the 12 categories, they saw a screen on which those categories were listed. They filled in cells adjacent to each category label to report their monthly purchasing frequency for each category in turn. In Condition 2 of Experiment 2b, the wording of the questions was the same as in Experiment 1 but with the dates changed to reflect the fact that the experiment took place in September 2022 rather than June 2019. For example, to elicit assessments of current prices, participants were asked: “What is the average cost of a single payment this year (September 2021-September 2022)?”

In Condition 3, the 12 consumer product categories were displayed to participants and they were asked to click on the category that had the largest price change over the past 12 months. After that, they responded to the same questions as were put to the participants in the 12 consumer product priming task but they did so only for the single category that they had specified as showing the largest price change.

In Condition 4, participants were first asked to “think of the largest price change you have noticed over the past 12 months” and to enter in a blank cell on the screen the specific good or services category that produced that change (c.f., Bruine de Bruin, Van der Klaauw, et al., 2011). Then, as in Condition 3, they responded to the same questions as were put to the participants in the 12 consumer product priming task but they did so only for the single category that they had specified as showing the largest price change.

*Procedure* Participants first responded to demographical questions about their age, gender, nationality, country of residence, occupation, education level, income level, ethnicity and marital status. They were then provided with a simple definition of inflation before completing their experimental tasks.

### 3.3.2. Results of Experiment 2a

Sixteen participants were excluded because their direct inflation forecasts or their estimates of price changes were more than 3.0 standard deviations from the mean value. As a result, the analyses were carried out on 205 participants (139 females, 65 males, 1 prefer not to say) whose mean age was 41 years ( $SD = 15$  years). Of these, 56 were in Condition 1, 44 were in Condition 2, 53 were in Condition 3, and 52 were in Condition 4.

Below, I report analyses of the level of inflation judgments and of both CE and AE levels in those judgments. Inflation outcomes for the 12 months following the date of the experiment were not available when these analyses were performed. Hence, to calculate errors in inflation expectations, I used the Monetary Policy Committee's annual inflation rate forecast (6.65%) for the second quarter of 2023 (Monetary Policy Committee, 2022a). Data are shown in Figure 3.2.

To compare the effect of different types of priming on the main task of producing direct estimates of future inflation (forecast, CE, AE) in the four groups, I carried out one-way ANOVAs using Task Type (no prime and the three priming conditions) as a between-participants factor. The analyses of forecasts and CEs<sup>12</sup> (upper panel of Figure 3.2) and AEs (lower panel of Figure 3.2) showed no significant effect of this factor. In summary, these analyses fail to show any effect of priming on the main task.

To obtain the indirect estimate of expected inflation for the 12 months from May 2022 in Condition 2, I proceeded in the same way as I did in Experiment 1: I took the frequency-weighted average of the changes in the estimated prices in each of the 12 categories of consumer product. I then used two-way mixed ANOVAs with Task Type (i.e., the three different types of priming task) as a between-participant factor and Estimate Type (Direct versus Indirect) as a within-participant factor to examine the relation between the two types of estimate in the three priming conditions. To obtain the indirect estimates of inflation in Conditions 3 and 4, I used the estimate of the price change that participants had specified for the single product category that they had selected as the one showing the largest price change.

The analyses of forecasts and CEs (upper panel of Figure 3.2) showed main effects of Task Type ( $F(2, 146) = 13.88, p < 0.001, ges = 0.0950$ ) and Estimate Type ( $F(1, 146) = 60.56, p < 0.001, ges = 0.1568$ ), together with an interaction between these variables ( $F(2, 146) = 14.06, p < 0.001, ges = 0.0795$ ). The simple effect of Estimate Type was significant in Condition 3 ( $F(1, 52) = 46.80, p < 0.001, ges = 0.2907$ ), and Condition 4 ( $F(1, 51) = 22.43, p < 0.001, ges = 0.1670$ ). The simple effect of Task Type was significant only for indirect estimates of inflation derived from differences in assessments of specific prices ( $F(2, 146) = 15.45, p < 0.001, ges$

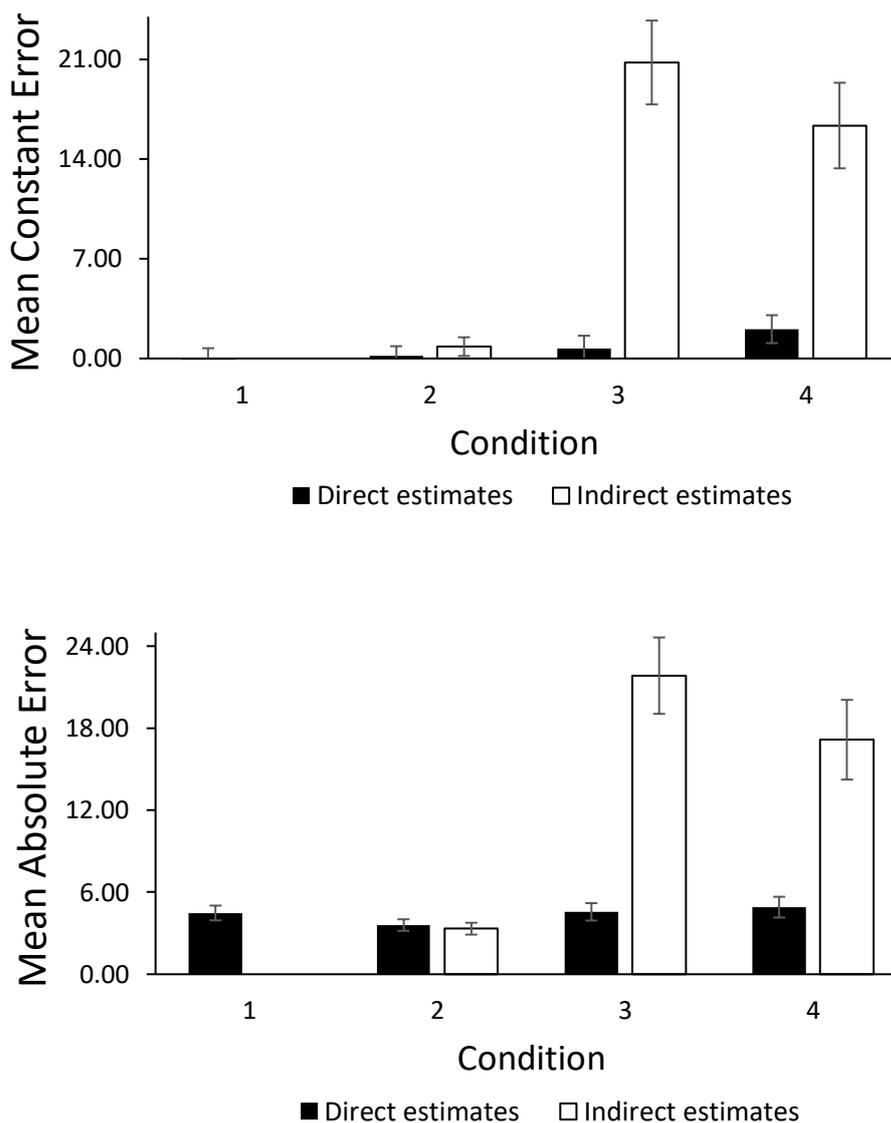
---

<sup>12</sup> These analyses produce the same results because CE equals the forecast minus a constant value, the true inflation rate.

= 0.1747). Pairwise comparisons using the BH adjustment method (Benjamini & Hochberg, 1995) showed significant differences between Condition 2 and the other two priming tasks: Condition 3 ( $p < 0.001$ ), and Condition 4 ( $p < 0.001$ ).

**Figure 3.2**

*Experiment 2a: Effects of Task Type on the constant errors and absolute errors of the direct and indirect inflation forecasts in each of the four conditions*



*Note.* Constant errors in Conditions 1 and 2 were very close zero and so only the top half of their error bars are visible in the figure.

The analysis of AEs (lower panel of Figure 3.2) showed main effects of Task Type ( $F(2, 146) = 14.57, p < 0.001, ges = 0.0937$ ) and Estimate Type ( $F(1, 146) = 44.95, p < 0.001, ges = 0.1293$ ), together with an interaction between these variables ( $F(2, 146) = 12.28, p < 0.001, ges =$

0.0750). The simple effect of Estimate Type was significant in Condition 3 ( $F(1, 52) = 36.97, p < 0.001, ges = 0.2591$ ), and Condition 4 ( $F(1, 51) = 17.34, p < 0.001, ges = 0.1394$ ). The simple effect of Task Type was significant only for indirect estimates of inflation derived from differences in assessments of specific prices ( $F(2, 146) = 14.30, p < 0.001, ges = 0.1638$ ). Pairwise comparisons using the BH adjustment method showed significant differences between Condition 2 and the other two priming conditions: Condition 3 ( $p < 0.001$ ), and Condition 4 ( $p < 0.001$ ).

For Condition 2, there were correlations between direct and indirect estimates of future inflation ( $r = 0.56, t(42) = 4.37, p < 0.001$ ) and between the CE values in those estimates ( $r = 0.56, t(42) = 4.37, p < 0.001$ ), but not between the AE values in them. For Conditions 3 and 4, there were no significant correlations between direct and indirect estimates, between CE values in those estimates, or between AE values in them.

### 3.3.3. Results of Experiment 2b

Thirty-two participants were excluded because their direct inflation forecasts or their estimates of price changes or purchasing frequency were more than 3.0 standard deviations from the mean value. As a result, the analyses were carried out on 225 participants (134 females, 87 males, 4 prefer not to say) whose mean age was 29 years ( $SD = 9$  years). Of these, 58 were in Condition 1, 59 were in Condition 2, 57 were in Condition 3, and 51 were in Condition 4.

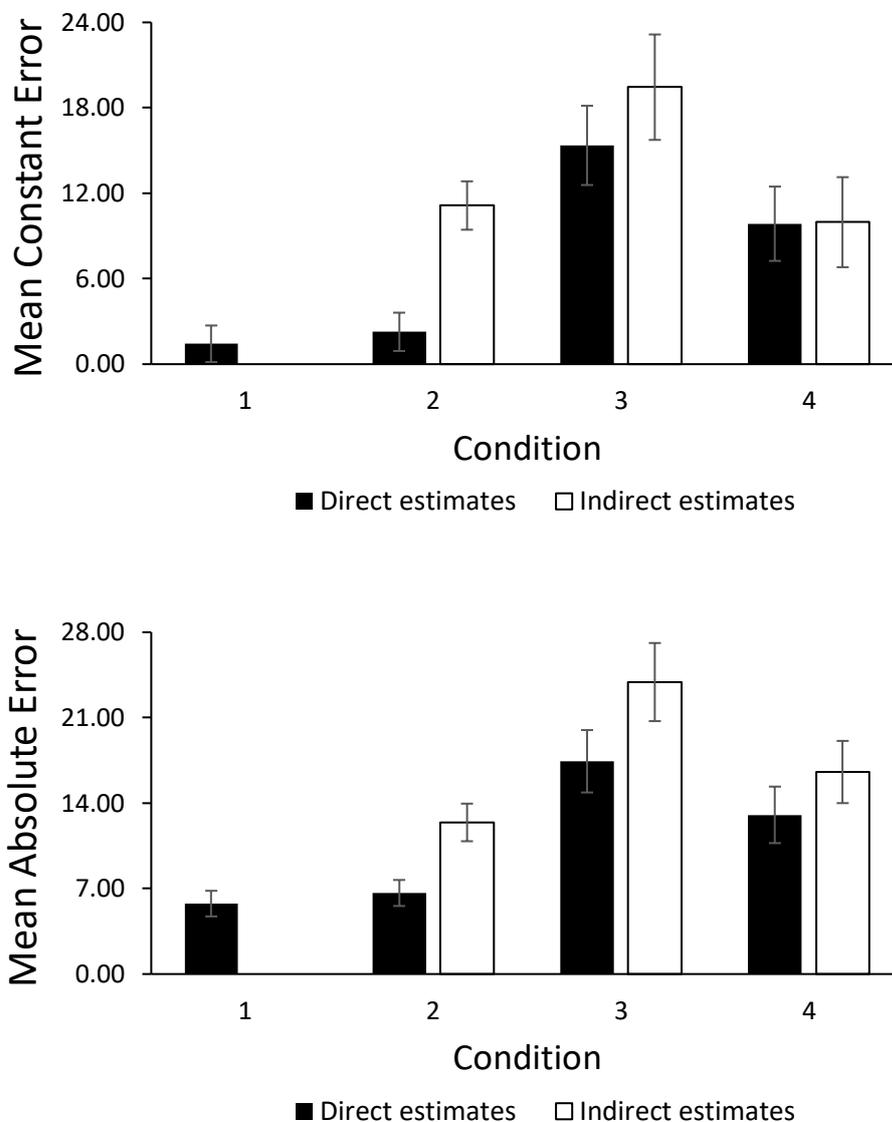
Below, I report analyses of the level of inflation judgments and of both CE and AE levels in those judgments. The actual inflation in September 2022 was 10.1% (Office for National Statistics, 2022b). Inflation outcomes for the 12 months following the date of the experiment were not available when these analyses were performed. Hence, to calculate errors in inflation expectations, I used the Monetary Policy Committee's annual inflation rate forecast (9.53%) for the third quarter of 2023 (Monetary Policy Committee, 2022b).

Levene's test showed that the ANOVA assumption of homogeneity of variance was violated. When sample sizes are similar and variances are proportional to means (as here), ANOVA seems fairly robust to violation of this assumption (Field et al., 2012). Below I report results of ANOVAs but, in Appendix 2, I also report results from robust ANOVAs (Wilcox, 2017). Conclusions do not differ for the analyses that I report below.

To compare direct estimates of inflation in the four conditions, I used one-way ANOVAs with Task Type (no prime and the three priming conditions) as a between-participants factor. Results for CE (and forecasts) revealed a significant effect of Task Type:  $F(3, 221) = 10.32, p < 0.001, ges = 0.1229$ . Pairwise comparisons using BH adjusted method showed significant difference between Condition 1 and condition 3 ( $p < 0.001$ ), between condition 1 and condition 4 ( $p = 0.01$ ), between condition 2 and condition 3 ( $p < 0.001$ ) and between condition 2 and condition 4 ( $p = 0.02$ ).

**Figure 3.3**

*Experiment 2b: Effects of Task Type on the constant errors and absolute errors of the direct and indirect inflation forecasts in each of the four conditions*



Results for AE showed a similar pattern. There was a significant effect of Task Type:  $F(3, 221) = 9.19$ ,  $p < 0.001$ ,  $ges = 0.1109$ ). Pairwise comparisons using BH adjusted method showed significant difference between Condition 1 and condition 3 ( $p < 0.001$ ), between Condition 1 and condition 4 ( $p = 0.01$ ), between condition 2 and condition 3 ( $p < 0.001$ ) and between condition 2 and condition 4 ( $p = 0.02$ ).

To compare direct and indirect estimates of future inflation, I used two-way mixed ANOVAs. The analysis of CE (and forecasts) revealed a significant effect of Task Type:  $F(2, 164) = 6.63$ ,  $p = 0.002$ ,  $ges = 0.0517$ ). Pairwise comparisons using BH adjusted method showed significant difference between condition 2 and condition 3 ( $p < 0.001$ ), and between condition 3 and

condition 4 ( $p = 0.009$ ). There was also a significant main effect of Estimation Type:  $F(1, 164) = 6.22$ ,  $p = 0.01$ ,  $ges = 0.0122$ ).

Analysis of AE showed a significant main effect of Task Type:  $F(2, 164) = 9.52$ ,  $p < 0.001$ ,  $ges = 0.0700$ ). Pairwise comparisons using BH adjusted method showed significant difference between condition 2 and condition 3 ( $p < 0.001$ ), between condition 2 and condition 4 ( $p = 0.02$ ), and between condition 3 and condition 4 ( $p = 0.02$ ). There was also a significant main effect of Estimation Type:  $F(1, 164) = 11.19$ ,  $p = 0.001$ ,  $ges = 0.0234$ ).

For participants in all three priming conditions, there were correlations between direct and indirect estimates of future inflation ( $r = 0.38$ ,  $t(165) = 5.25$ ,  $p < 0.001$ ), between CE values in those estimates ( $r = 0.38$ ,  $t(165) = 5.25$ ,  $p < 0.001$ ), and between the MAE values in those estimates ( $r = 0.35$ ,  $t(165) = 4.80$ ,  $p < 0.001$ ).

### 3.3.4. Discussion

In Experiment 2a, there was no evidence of any priming effects. With respect to Condition 2, this replicates the finding in Experiment 1 ( $H_3$ ). With respect to Conditions 3 and 4, it represents a failure to replicate the findings reported by Bruine de Bruin, Van der Klaauw, et al. (2011): overall estimates of inflation in the main task were not elevated when people were first asked to identify the product category with the largest price change in the previous year and to estimate the size of that change ( $H_4$ ). Furthermore, the lack of a priming effect in either of these conditions meant that providing people with a list of candidate product categories from which to select the one with the largest price rise did not lead to a greater priming effect than requiring them to search their memory for the category with the largest price rise ( $H_5$ ).

There were two ways in which the results of Experiment 2a differed from those of Experiment 1. First, whereas indirect estimates of inflation based on the frequency-weighted average of annual changes in judged prices of each of the 12 product categories were much higher (and, hence, less accurate) than direct estimates in Experiment 1 (Figure 3.1), there was no significant difference between them in this experiment (Figure 3.2, Condition 2). Thus, whereas the previous experiment produced evidence inconsistent with the view that assessments of specific prices (across all product categories) are used to form people's overall estimates of inflation ( $H_1$ ), no such evidence was obtained in the present experiment. In fact, the close similarity of people's direct estimates of overall inflation with the indirect estimates based on their earlier assessments of price changes in all 12 categories implies that the former was based on the latter.

Second, whereas in the previous experiment, there was no correlation across participants between indirect and direct estimates of inflation ( $H_2$ ), the present experiment did reveal significant correlations between these variables in Condition 2. If the price-recall model is correct, I would expect the correlation to be significant in Condition 2. I would not expect it to be significant in Conditions 3 and 4 if people recognise that the prices that they have

recalled for only the single product category with the largest price change should not be used as a guide for their estimate of overall inflation. In summary, whereas the results of Experiment 1 were best explained by the price-free model, results from Experiment 2a are most consistent with the price-recall model.

Findings in Experiment 2b are somewhat different. Priming effects occurred when direct estimates of inflation were primed by recall of price changes in a single product category (Conditions 3 and 4) but not when they were primed by recall of price changes in all product categories (Condition 2). This is consistent with results reported by Bruine de Bruin, Van der Klaauw, et al. (2011). Furthermore, despite the lack of a priming effect in Condition 2, indirect estimates of inflation based on a frequency-weighted average of all product categories were significantly higher and less accurate than direct estimates of inflation (Figure 3.3). There are two ways of explaining this. First, direct estimates of inflation were not based on recall of price changes in individual product categories but were based on media reports and other secondary sources; in other words, as in Experiment 1, the price-free model holds. Alternatively, when making direct estimates of inflation, people may have recalled prices that they had paid for various types of product but their recall of price changes in all types of product enabled them to allow for the effects of their own relatively high purchase of products that showed particularly high price increases (food, fuel).

Can I select between these two accounts? In contrast to Experiment 1, there were quite strong correlations between direct and indirect estimates of inflation. These would certainly be expected if people based their direct estimates on recall of prices. However, they could also be explained by a price-free model if people who attend more to expert estimates of inflation transmitted via the media are also better at estimating price changes in individual product categories. I cannot exclude this possibility. But only a price-salience model explains the priming effects in Conditions 3 and 4. It also most easily accounts for correlations between direct and indirect estimate of inflation. However, it must be elaborated somewhat to explain the difference between direct and indirect estimates of inflation in Condition 2.

Why were results from Experiment 2b somewhat different from those of Experiment 2a? The 'inflation' wording used in Experiment 2a can explain why estimates were generally lower (and, hence, more accurate) than those in Experiment 2b, where the 'price change' wording was used. Bruine de Bruin et al. (2012) showed that inflation wording produces lower estimates. However, wording effects do not easily account for other differences, such as the appearance of priming and the difference between direct and indirect estimates in Experiment 2b but not in Experiment 2a. These are better interpreted as effects of a switch from a price-recall model (Experiment 2a) to a price-salience model (Experiment 2b). Why would such a switch have occurred between May and September 2022? In that period, inflation in just one or two crucial product categories that are purchased by everyone had increased very rapidly; in particular, inflation for food and non-alcoholic beverages has increased from 8.6% to 14.5%. This is the type of phenomenon that was present when Bruine

de Bruin, Van der Klaauw, et al. (2011) carried out their work that showed priming after recall of price changes for single products.

### 3.4. General Discussion

In Experiment 1, direct estimates of overall inflation for the following year were different from indirect estimates obtained by averaging the differences in people's estimates of the costs of each of the 12 categories of consumer product for the next year and their estimates of the costs of the same product for current year (Figure 3.1). Furthermore, there was no correlation between these different types of estimate. These results fit well with the price-free model. In Experiment 2a, results were different: direct estimates of inflation were very similar to indirect estimates derived from people's estimates of the costs of all categories of item, the two types of estimate were quite highly correlated ( $r = 0.56$ ), and there was no evidence of priming effects. These results are what I would expect from the price-recall model. In Experiment 2b, results were different again. As in Bruine de Bruin, Van der Klaauw, et al. (2011), there was strong evidence of a priming effects when people recalled the single product category with the highest price rise over the previous year. Direct and indirect estimates of inflation were again quite highly correlated but the latter were somewhat higher and less accurate than the former. These results are most easily accommodated within the price-salience model.

These differences in the cognitive processing underlying the formation of people's expectations for future inflation appear to be related to the inflation environment prevailing during the period when the judgments were made. Experiment 1 was carried out in the summer of 2019 when levels of inflation were historically very low, when they had not changed substantially for many years, and when they were not expected to change much in the future. Media reports ensured that people had this information available to them. As a result, they had no need to monitor prices of different categories of consumer product. Consequently, their estimates of overall inflation were not based on knowledge of how the costs of different types of product were changing: they used a price-free approach to making those estimates.

Experiment 2a was carried out in May 2022. At the time, inflation had risen very rapidly to 9.13% (Office for National Statistics, 2022a), a 40-year high, and was forecast to rise to 10.19% in the fourth quarter of 2022, before dropping back to 3.56% by the end of 2023 (Monetary Policy Committee, 2022a). However, MPC inflation forecasts have been changing rapidly: in May 2021, the forecast for the fourth quarter of 2022 was just 2.02%; now it is in double figures. At the time of writing (November 2022), there is great uncertainty about how long high inflation will last and the notion that it will be as transient as the May 2022 forecasts suggest is increasingly regarded as optimistic as concerns about an inflationary wage-price spiral increase. Also, while inflation for certain product categories (fuel, food) was higher than that for others in May 2022, inflation was increasing rapidly in all categories. In such circumstances, people need to carefully monitor price changes themselves to produce

sensible decisions about their consumer behaviour. Given this, it is not surprising that the price-recall model provided the best account of the data from Experiment 2a.

Experiment 2b was carried out in September 2022. Results replicated the priming effects that Bruine de Bruin, Van der Klaauw, et al. (2011) obtained in support of a price-salience model. What aspects of the inflation environment were common to the period when they carried out their studies and when I carried out Experiment 2b? In the 12 months before they released their surveys in August 2010, inflation was low at 1.1%. For most products, it had changed little in the recent past. However, for certain commodities, it was very high and changing rapidly. For example, monthly inflation rate for fuel oil, the product category with the highest level of inflation, differed by 16.6% between March and August 2010 (Bureau of Labor Statistics, 2010a, b). When I carried out Experiment 2b in September 2022, inflation rates for different product categories varied widely but, generally, they were fairly stable. However, monthly inflation rate for food and non-alcoholic beverages was changing rapidly. In August 2010 and September 2022, there was little need to monitor all product categories as most of them were fairly stable. However, in both cases, inflation rate for an important product category was not only high but also changing very rapidly. Consumers needed to monitor price changes in these salient categories. Following Bruine de Bruin, Van der Klaauw, et al. (2011), I conclude that recall of inflation levels in these categories produces judgment anchors and that, as a result of under-adjustment (Tversky & Kahneman, 1974), direct estimates of overall inflation are biased upwards.

In summary, the model that best explains consumers' use of information about specific prices when estimating overall inflation rates depends on the inflation environment at the time. Rather than thinking of the price-free, price-recall, and price-salience approaches as competing models each of which is aimed at explaining all estimates of inflation, we should think of them as different cognitive strategies, each of which is adapted to a different inflation environment.

This approach assumes that people obtain some broad-brush information about how much prices are changing in different product categories. They may obtain this from the media or from very occasional sampling of prices in different categories. They then use it to decide whether to monitor prices in any or all of the product categories more thoroughly. Comprehensive price monitoring may be effective for estimating overall inflation but it is likely to be resource-demanding. Consumers are unlikely to carry it out unless it is worth their while. More intensive processing of price information will improve their economic decision making but the advantages gained may outweigh the effort that it requires only when inflation rates are high, unstable, difficult to predict, and variable across product categories.

Do I have evidence of an effect of the wording used in the estimation tasks? Bruine de Bruin et al. (2012) found that asking people to estimate inflation produce lower values than asking them to estimate changes in "prices in general" or "prices you pay". I found a similar effect:

in Experiment 2a, direct estimates of inflation were unbiased when people were asked to estimate “inflation” as a percentage (Figure 3.2) whereas, in Experiments 1 and 2b, they were biased upward when people were asked to estimate the average price change as a percentage (Figures 3.1 and 3.3). However, wording does not seem to influence other aspects of the data. For example, clear correlations between direct and indirect estimates of inflation appeared both when the “inflation” wording was used (Experiment 2a) and when the “price change” wording was used (Experiment 2b). Features of the data other than the absolute level of estimates are more simply explained in terms of the prevailing inflation environment.

#### *3.4.1. Limitations*

Bruine de Bruin, Van der Klaauw, et al. (2011) recognized potential problems with the priming methodology. Asking someone to complete a priming task (e.g., estimating the price rise in the product category with the largest price rise) may influence their later performance on the main task (i.e., estimating overall inflation rate). But that does not mean that when carrying out the main task without the priming task, they would prime themselves by implicitly thinking of the largest price rise. This needs to be borne in mind when interpreting evidence from Experiment 2b that is consistent with the price-salience model.

An alternative approach adopted by Bruine de Bruin, Van der Klaauw, et al. (2011) in their second study was to ask people to estimate overall inflation rate without any priming task but afterwards to ask them whether they had thought of specific price rises when making their estimates. When I used this approach in my second experiment, I obtained data consistent with the rest of my results (Figure A3.5 in Appendix). However, these data were not crucial for testing my hypotheses and, as I mentioned above, people’s introspections about their thought processes may represent conscious (System 2) post-hoc rationales for the way they used their intuition (System 1) to estimate inflation rate (Evans, 2008).

#### *3.4.2. Implications*

People made overall estimates of inflation that were 3 - 7% too high when they used a price-free strategy based on media reports and other secondary sources (Experiment 1; Figure 3.1). However, when they used a price-recall or price-salience strategy based on their own purchasing experience, this bias was almost absent (Figure 3.2) or much smaller (Figure 3.3). This supports my view that the bias in the former case largely arises from media-based risk amplification: media reports to which people attend more frequently deal with rises than falls in inflation.

When people use a price-salience strategy, biases in estimates of overall inflation will be much larger. Without priming, Bruine de Bruin, Van der Klaauw, et al. (2011, Study 2) found mean overall inflation rate was expected to be 3.66%. However, when respondents first estimated the largest price change that they had noticed over the previous 12 months, they expected mean overall inflation rate to be 7.11% (Study 1). Similarly, in Experiment 2b, mean overall

inflation was expected to be 10.96% without priming but 24.89% and 19.39% when people first estimated price rises in the category that they had recognised or recalled to be the one with the highest price rise (Figure 3.3). This suggests that consumer decisions will be particularly impaired during periods in which people use a price-salience strategy. For example, compared with those using a price-recall strategy, the wage rises that they seek will be higher and seen as less reasonable by employers.

This suggests that we need to consider the factors that lead consumers to switch from one strategy to another. The first is the inflation environment. As inflation increases from a previously stable and low level, it is likely to affect one or two product categories (e.g., fuel, food) first: this is likely to trigger a switch from a price-free strategy to a price-salience one. If inflation continues to increase, more product categories will show high price rises and those price rises are likely to be less predictable: this is likely to trigger a switch from a price-salience strategy to a price-recall strategy. The second factor is that monitoring prices places a demand on cognitive resources. If the gain that monitoring produces no longer exceeds the cost that it incurs, people are likely to switch from a price-recall to a price-salience strategy or from a price-salience strategy to a price-free strategy.

### *3.4.3. Conclusion*

Data reported here, together with earlier findings reported by others, are consistent with the view that the degree to which people recall past and present prices of specific products when forming their expectations for future levels of inflation depends on the inflation environment at the time. If inflation has been low and its level is not expected to change, people do not use their personal experience of specific prices; instead their expectations are based on media reports and other secondary sources. Due to media-based risk amplification, expectations may then be somewhat too high. If inflation is generally low and stable but rises to high levels for some product categories, they monitor the prices in those categories. The results of this monitoring are used to produce their estimates of the overall level of future inflation; consequently, these estimates may be much too high. If inflation is generally high, unpredictable, and possibly variable across product categories, people monitor prices in all product categories. Because media reports are not used with this strategy, media-based risk amplification does not occur. As a result, inflation estimates are unbiased (Figure 3.2, upper panel).

## Chapter 4. How lay expectations are measured in surveys

### 4.1. Introduction

In chapter 3, I have shown that the formation of lay inflation expectations could base on different types of information in terms of the inflation environments, which resulting different accuracy levels of inflation expectations. It is reasonable to assume that this disagreement between lay and expert forecasters may also arise, in part, because they base their expectations on different types of information.

First, lay people are not exposed to or do not attend to information of the quality absorbed by experts (Binder & Rodrigue, 2018; Cavallo et al., 2017). News media comprise their main source of economic information and heterogeneity of their inflation expectations can be partly attributed to exposure to different reports (Lamla & Maag, 2012). Also, news media are likely to treat larger price rises for some items as more newsworthy than smaller rises for the majority of items: lay judgments of inflation are likely to be biased in an upward direction by this 'social amplification' process (Soroka, 2006). In contrast, experts are relatively well-informed and use similar datasets to update their beliefs (Coibion, Gorodnichenko, Kumar, et al., 2020; Gábrriel et al., 2014).

A second difference is that only lay people draw on their own personal experience of price changes when forecasting inflation. As a result, differences in personal experience contribute to the greater heterogeneity observed in their inflation expectations (Bates & Gabor, 1986; Brachinger, 2008; Jungermann et al., 2007; Lein & Maag, 2011; Madeira & Zafar, 2015; Malmendier & Nagel, 2016; Ranyard et al., 2017).

A third difference concerns the way in which information about certain other economic variables (e.g., inflation rates, unemployment rates) can be used to forecast inflation. Experts can use their macroeconomic models for this purpose. Lay people, without access to these models, may exploit their own naïve theories of how the economy works or use simple heuristics, such as the good-begets-good heuristic (Leiser & Aroch, 2009; Leiser & Krill, 2018). These lay approaches are likely to be less effective at forecasting inflation than the models used by experts.

These three factors can explain why inflation judgments by experts responding to surveys directed at them are superior to and more homogeneous than inflation judgments by lay people responding to surveys targeting them. Crucially, however, experts and lay people have been required to respond to *different* surveys. The notion that there is a difference between lay and expert judgments that is in need of explanation is predicated on the assumption that these different surveys are equally good at eliciting judgments of inflation. It is possible that this assumption is not valid. For example, if I asked experts to answer the consumer surveys normally given to lay people and lay people to respond to the surveys designed for professional respondents, I might find that the latter group are now more accurate and less

homogeneous than the former one. While this outcome may not seem likely, the possibility that it could occur emphasises the importance of investigating the effects of survey format on the accuracy and homogeneity of inflation judgments. There have already been a number of studies of this issue.

#### *4.1.1. Effects of survey format*

Various surveys have been developed to elicit inflation expectations from lay respondents and experts. Format varies across consumer surveys in a number of ways. In some cases, questions prompt point forecasts but, in other cases, they elicit probability density functions (Armantier et al., 2013; Bruine de Bruin, Manski, et al., 2011). Sometimes people are asked to estimate 'inflation' whereas, on other occasions, they are required to estimate 'general price change' (Armantier et al., 2017; Bruine de Bruin, Potter, et al., 2010; Bruine de Bruin et al., 2012; Bruine de Bruin et al., 2017). In some surveys but not others, respondents are given the opportunity to revise their answers (Bruine de Bruin et al., 2017). Main and interactive effects of these factors influence the inflation forecasts that people provide (Bruin de Bruin et al., 2017).

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

#### *4.1.2. Information context: Differences between surveys of lay and expert forecasters*

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

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

on treasury bonds), and various other indicators. Again, surveys to which lay people respond do not provide this contextual information.

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

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

#### *4.1.3. Judgment heuristics in forecasting: The nature of the information available*

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

When no external information is provided, judgmental forecasters must rely on relevant information held in memory. The availability heuristic is appropriate to such circumstances (Kahneman & Tversky, 1973). For consumers, extreme price changes are more salient and

available to memory. Hence they have an inordinate influence on judgments of inflation or 'general price change' (Bruine de Bruine, Van der Klaauw, et al., 2011).

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

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

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

Would I expect forecasts to improve if I gave people contextual information about other economic variables from the period prior to the one being forecast? While it is not unreasonable to expect that additional information will improve performance, it is possible that the two heuristics used for forecasting in this situation lead people further astray. Although the good-begets-good heuristic can be regarded as a lay version of the professional view that economies can be classified on a continuum from good to bad using a measure such as the 'misery index' (Barro, 1999), it is also possible to see how use of this heuristic could be misleading. For example, Phillips (1958) found an inverse relationship (the Phillips curve) between inflation rate and unemployment rate; in other words, low inflation ('good') begets high unemployment ('bad'). However, since the 1970s, the relation described by the Phillips curve has become less clear, arguably because inflation expectations have had more of a role

in determining inflation (Phelps, 1969). Hence, use of the good-begets-good heuristic may not lead people astray as much as it would have done in earlier times. However, use of the representativeness heuristic in the manner outlined above may also introduce error into inflation forecasts. Relations between inflation rate and other variables are subject to uncertainty and so I should expect some regression to the mean when using the latter to forecast the former. However, forecasts based on representativeness do not allow for this effect.

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

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

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

#### **4.2. Experiment 3**

Lay people made a series of four inflation judgments either for the current year (inflation perception) or for the upcoming year (inflation expectation). Their first judgment was made without any additional information. They made their second forecast with provision of

information about either the interest rate or the unemployment rate (randomly chosen) on the period prior to the one for which the inflation forecast was required. They made their third forecast with provision of information about the variable (either interest rate or unemployment rate) that had not been provided for the second forecast; again, this information pertained to the period immediately prior to the one for which the inflation forecast was required. They made their fourth forecast after additional information was provided about the level of inflation on the period immediately prior to the one for which the forecast was required.

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

*H<sub>1</sub>: Mean value of inflation forecasts will be too high.*

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

*H<sub>2</sub>: Second, third and fourth inflation judgments will be more accurate than the first ones.*

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

*H<sub>3</sub>: The fourth forecast will be more accurate and less variable than any of the earlier forecasts.*

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

*H<sub>4</sub>: The fourth forecast will be more accurate when people are given data about the previous five periods than when they are given data about just the immediately preceding period.*

In Ranyard et al.'s (2017) model, experienced price changes, media reports and official statistics produce inflation perceptions via a nowcasting process. These inflation perceptions, together with expert forecasts and inferences produced by naïve models of the economy, then produce inflation expectations via a forecasting process. This implies that inputs to inflation perceptions (e.g., experienced price changes) then go on to influence inflation

expectations. In line with this, Dräger (2015) found strong effects of structural shocks to inflation perceptions on inflation expectations.

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

*H<sub>5</sub>: Effects of contextual information on inflation perceptions will be similar to its effects on inflation expectations but perceptions will be more accurate.*

#### 4.2.1 Method

*Participants* One hundred and forty-eight people (40 men, 108 women), all of whom had been living in the United Kingdom for at least two years, were recruited via the participant recruitment platform, Prolific.com. Table A4.1 in Appendix 1 shows their demographic characteristics. Each participant was paid £0.60 to complete the study. Data were collected between 7 March and 4 April 2020.

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

*Stimulus materials* Participants made estimates of the UK inflation rate for 2019 or predictions of the inflation rate for 2020 by entering their judgments into empty cells of tables presented to them (Figure 4.1). Contextual information was supplied by entering values into appropriate cells in the tables for the last three inflation judgments and comprised UK historical data for base interest rates, unemployment rates, and CPI inflation rates for the years 2014 to 2019.

All data used in the experiment were obtained from UK official reports published by the Office for National Statistics and the Bank of England.

#### Figure 4.1

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

#### Task instructions

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

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

#### Economic indicators: Annual data (%)

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

#### Economic indicators: Annual data (%)

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

*Procedure* After people had been informed about the nature of the study, been given details of the ethical permission that it had received, and been told that they could withdraw from it at any time, they gave their consent to participate. They were then supplied with simple definitions and examples of the three economic indicators involved in the study (base interest rate, unemployment rate, CPI inflation rate). They were randomly allocated to one of the four experimental groups. For each of the four judgments that they made, they were instructed to provide the inflation judgment appropriate to their group (Figure 4.1)<sup>13</sup>. After all judgments

<sup>13</sup> In this and later experiments, after participants had entered each of their inflation judgments, they gave an estimate of the likelihood that it would be within 10% of the true value. These estimates showed that people

had been completed, basic demographic details were collected (gender, age, highest level of education qualification obtained, primary academic discipline, working experience related to economics, and primary country of residence over the previous two years).

#### 4.2.2. Results

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

The upper panel of Table 4.1 shows means and standard deviations of levels of people's raw inflation judgments in the four experimental groups. To measure errors in 2019 inflation judgments, I used the 1.8% value for the year 2019 reported by the Office for National Statistics as the correct one. To measure errors in 2020 inflation judgments, I used the forecast of 1.5% for the year 2020 that was issued by HM Treasury and based on forecasts they received from many different institutions between 1<sup>st</sup> March and 17<sup>th</sup> March 2020.

Consistent with  $H_1$ , judged inflation rates were too high (Table 4.1, Middle panel). Directional errors were significantly above zero on the first judgment ( $t(134) = 4.72; p < 0.001$ ), the second judgment ( $t(134) = 3.86; p < 0.001$ ), the third judgment ( $t(134) = 4.88; p < 0.001$ ) and the fourth judgment ( $t(134) = 9.85; p < 0.001$ ).

A three-way mixed analysis of variance (ANOVA) on the directional errors with Task (inflation perception, inflation expectation) and Contextual Information (one year, five years) as between-participant variables and Judgment Number (first, second, third, fourth) as a within-participant variable showed only an effect of Judgment Number ( $F(2.32, 303.79) = 3.01; p = 0.043; ges = 0.009$ )<sup>14</sup>.

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

Absolute error scores are shown in the lower panel of Table 4.1. A three-way mixed ANOVA using the same factors as before showed a main effect of Judgment Number ( $F(2.10, 275.10)$

---

were overconfident in their inflation judgments. As this phenomenon was not my present concern, I do not report data demonstrating it here. I discuss overconfidence in inflation judgments in Chapter 5.

<sup>14</sup> When Mauchy's test showed a deviation from sphericity, Greenhouse-Geisser corrections were used to adjust degrees of freedom. Generalised eta squared (*ges*) measured effect size (Olejnik & Algina, 2003).

= 13.16;  $p < 0.001$ ;  $ges = 0.0434$ ) and an interaction between Judgment Number and Contextual Information ( $F(2.10, 275.10) = 3.36$ ;  $p = 0.034$ ;  $ges = 0.011$ ). The simple effect of Judgment Number was significant both for when there was one year of contextual information ( $F(2.49, 174.30) = 5.00$ ;  $p = 0.004$ ) and when there were five years of contextual information ( $F(1.79, 113.02) = 8.87$ ;  $p < 0.001$ ). These effects are consistent with  $H_2$  and are shown in Figure 4.2.

**Table 4.1**

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

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

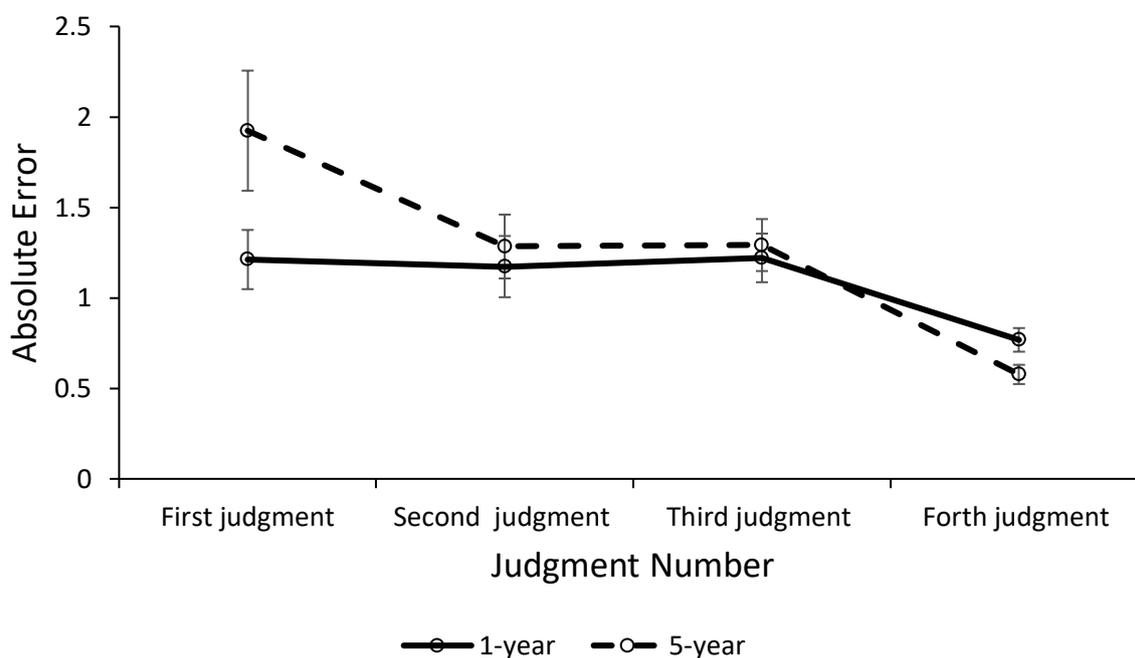
Multiple Bonferroni pairwise comparisons showed absolute error was lower for the fourth forecast than for the first forecast (one-year information:  $p = 0.02$ ; five-years information:  $p$

< 0.002), the second forecast (one-year information:  $p = 0.08$ ; five-years information:  $p = 0.006$ ) and the third forecast (one-year information:  $p = 0.002$ ; five-years information:  $p < 0.001$ ). These results provide evidence consistent with  $H_3$ : the fourth forecast was more accurate than the preceding ones.

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

**Figure 4.2**

*Experiment 3: Interaction between Contextual Information and Judgment Number in the analysis of absolute error (together with standard error bars)*



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

more accurate when they are given data about the previous five periods than when they are given data about just the immediately preceding period.

#### *4.2.3. Discussion*

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

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

### **4.3. Experiment 4**

In Experiment 3, different groups of people judged current inflation for 2019 and expected inflation for 2020. The distinction between the perception and expectation tasks was not made salient to either group. People performing these different tasks may have used very similar procedures to estimate the required inflation rate but, not being aware of the other task, may have failed to make allowances for the quality of and the uncertainty in the data on which they were basing their estimates. If I make people aware of the difference between the two tasks, they may respond differently to them. This reasoning provided the rationale for Experiment 4.

Different surveys ask people to estimate inflation for different combinations of years, I term it here as Task context. The MSC asks people to estimate the percent increase in prices over the next 12 months and to estimate the average percent increase over the next five to 10 years. The SCE asks for percentage estimates of inflation over the period between the present and a date 12 months later and over the period between a date 24 months from the present and a date 36 months from the present. The IAS asks people to estimate change in prices over the last 12 months, over the next 12 months, over the 12 months after that, and over the longer term (five years). The US-SPF asks experts for their estimates of inflation rate for the current year and the two following years. The EU-SPF solicits experts' views on inflation rate for the current year and the two following years. All these surveys obtain inflation estimates

for different years from the *same* respondents. This may be the reason those surveys produce different estimates from different years. Experiment 3 suggests that, had they used different respondents to obtain inflation estimates for different years, the differences between those estimates would have been much reduced.

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

In Experiment 3, people evaluated current and future inflation rates separately. Important differences between inflation perception and inflation expectation were not made salient. In Experiment 4, participants evaluated current and future inflation rates together by providing their estimates of inflation for 2019 and 2020 on the same screen. I anticipated that this would make the differences between the two tasks more salient and that people would better understand the different factors influencing each one. As a result, they should weight factors more heavily in perception than in expectation judgments when those factors are better predictors of current than future inflation (e.g., recent price rises). Hence,

*H<sub>6</sub>: Judgments of current inflation will be more accurate than those of future inflation.*

#### 4.3.1 Method

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

*Participants* Eighty-seven people (24 men, 63 women), all of whom had been living in the United Kingdom for at least two years, were recruited via the participant recruitment platform, Prolific.com. Table A4.1 in Appendix 1 shows their demographic characteristics. Each participant was paid £0.60 to complete the study. Data were collected between 19 August and 20 August 2020.

*Design* The design was the same as that used for Experiment 3 except that Task (inflation perception versus inflation expectation) was a within-participant variable instead of a

between-participant variable. Thus, Task and Judgment Number were within-participant variables and Contextual Information was a between-participant variable. Participants were randomly allocated to Group 1/2 (one year of contextual information) or Group 3/4 (five years of contextual information).

### Figure 4.3

Experiment 4: Response tables ready for a) entry of the first inflation judgment in group 1 (upper panel) and b) entry of the fourth inflation judgment in group 3 (lower panel)

#### Task instructions

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

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

Economic indicators: Annual data (%)

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

Economic indicators: Annual data (%)

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

*Materials* The screen into which participants entered their responses was similar to the one used for Experiment 3 except that they filled in two empty cells, one for 2019 and one for 2020. There was no constraint on the order of responding. Examples of the response screen are shown in Figure 4.3.

*Procedure* The procedure was identical to that used in Experiment 3.

#### 4.3.2 Results

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

The upper panel of Table 4.2 shows means and standard deviations of levels of people's raw inflation judgments in the four experimental groups. To measure errors in 2019 inflation judgments, I used the same criteria for correctness as before.

As in Experiment 3, judged inflation rates were too high: directional errors were significantly above zero on the first judgment ( $t(75) = 3.60$ ;  $p = 0.001$ ), the second judgment ( $t(75) = 3.19$ ;  $p = 0.002$ ), the third judgment ( $t(75) = 3.53$ ;  $p = 0.001$ ) and the fourth judgment ( $t(75) = 8.76$ ;  $p < 0.001$ ).

A three-way mixed ANOVA on the directional errors with Contextual Information (one year, five years) as a between-participant variable and Judgment Number (first, second, third, fourth) and Task (inflation perception, inflation expectation) as within-participant variables revealed a main effect of Task ( $F(1, 74) = 14.04$ ;  $p < 0.001$ ;  $ges = 0.0170$ ). In contrast to Experiment 3, overestimation was significantly greater for expected inflation in 2020 than for perceived inflation in 2019. There was also a main effect of Judgment Number ( $F(1.85, 136.75) = 5.31$ ;  $p = 0.007$ ;  $ges = 0.0226$ ). As the middle panel of Table 4.2 shows, directional error decreased over the four judgments.

Absolute error scores are shown in the lower panel of Table 4.3. A three-way mixed ANOVA using the same factors as before showed a main effect of Task ( $F(1, 74) = 19.29$ ;  $p < 0.001$ ;  $ges = 0.0235$ ). Thus, consistent with  $H_7$ , inflation perception was more accurate than inflation expectation in this experiment. There was also a main effect of Judgment Number ( $F(1.80, 133.20) = 7.46$ ;  $p = 0.001$ ;  $ges = 0.0337$ ). Post-hoc comparisons showed significant differences between the first judgment and the second judgment ( $p = 0.03$ ), the third judgment ( $p < 0.003$ ), and the fourth judgment ( $p < 0.001$ ), between the second judgment and the fourth judgment ( $p = 0.002$ ), and between the third judgment and the fourth judgment ( $p = 0.001$ ). Thus, provision of contextual information again improved judgment but, in contrast to Experiment 3, this effect was shown not only by the last judgment being better than the three earlier ones but also by the first judgment being worse than the three later ones. In other words, inflation judgments were helped by providing people with past information about macroeconomic variables other than inflation but were helped even more by giving them information about previous values of inflation (Figure 4.4).

The upper panel of Table 4.2 indicates that, as in Experiment 3, provision of contextual information resulted in judgments of inflation rate becoming more homogeneous. Mean

variance of the fourth judgment was lower than that of the third judgment ( $F(75, 75) = 4.00$ ;  $p < 0.01$ ), the second judgment ( $F(75, 75) = 6.88$ ;  $p < 0.01$ ) and the first judgment ( $F(75, 75) = 13.61$ ;  $p < 0.01$ ), mean variance of the third judgment was lower than that of the second judgment ( $F(75, 75) = 1.72$ ;  $p < 0.025$ ) and the first judgment ( $F(75, 75) = 3.40$ ;  $p < 0.01$ ), and mean variance of the second judgment was lower than that of the first judgment ( $F(75, 75) = 1.98$ ;  $p < 0.05$ ).

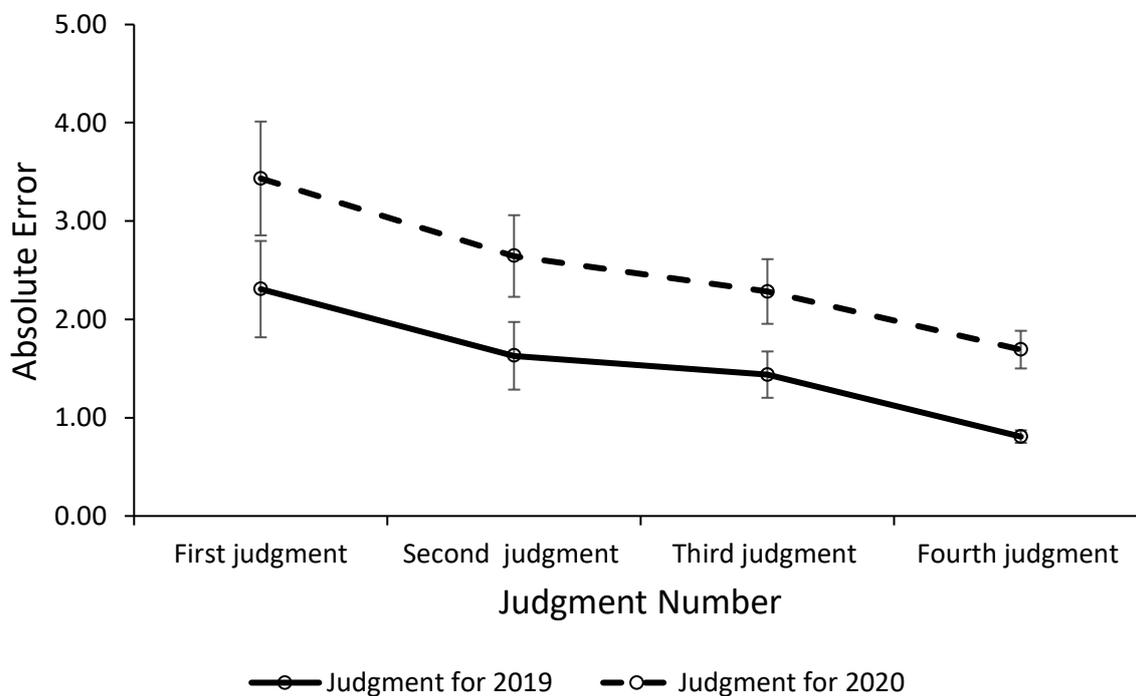
**Table 4.2**

*Experiment 4: Means and standard deviations (in parentheses) of inflation judgments, their directional errors, and their absolute errors*

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

**Figure 4.4**

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



#### 4.3.3 Discussion

Use of joint evaluation was effective in rendering the difference between the inflation perception and inflation expectation tasks salient. As expected, the former was now performed more accurately than the latter. Also, as in Experiment 3, contextual information reduced absolute error in judgments. As Figure 4.4 shows, this was evidenced by lower judgment error when information about past values of macroeconomic variables other than inflation were provided relative to when no information was provided and lower judgment error when information about past values of inflation were provided relative to error when information about past values of macroeconomic variables other than inflation were provided.

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

In all major surveys, people make joint rather than separate evaluations of inflation rates in different years: estimates of inflation for one or more later years are made in the context of already having made an estimate of inflation for at least one earlier year. As a result, people's expectations about how inflation changes from one year to the next influence their judgments of inflation for later years. My results imply that people expect inflation to increase over time,

even when it does not do so. (Compare the bottom rows of the upper panels of Tables 4.1 and 4.2.)

#### 4.4 Experiment 5

There is one final issue that needs to be resolved. The experiments have shown that, relative to when no contextual information is provided, judgment error was lower when people are given information about past values of macroeconomic variables other than inflation (Figure 4.4). Furthermore, relative to when information about past values of macroeconomic variables other than inflation is provided, judgment error was lower when people are given information about past values of inflation (Figures 4.2 & 4.4). The issue is whether these improvements occurred a) because people had received *more* information when making later judgments than when making earlier ones, or b) because they had received *more useful* information when making later judgments than when making earlier ones. My data already support the latter proposition. In neither experiment was judgment accuracy higher on the third judgment than on the second one. In other words, providing more information about the past values of an additional macroeconomic variable had no effect. It was only when more useful information in the form of past values of inflation was provided on the fourth judgment that an additional improvement in accuracy was observed in both experiments.

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

##### 4.4.1. Method

The experiment was similar to the Experiment 4 except that contextual information was varied between participants. There were four groups of participants, each of which made judgments for both 2019 and 2020. Within each set, each group was given just one of four different types of contextual information: no contextual information; base interest rate information for the preceding five years; unemployment rate information for the preceding five years; inflation rate (CPI) information for the preceding five years. Thus, if accuracy is found to be higher in the fourth group than in the second and third group, it cannot be because participants in that group had more information than those in the second and third groups. It would have to be because participants in that group had more useful information than those in other groups.

*Participants* Three hundred and fifty-two people (108 men, 244 women), all of whom had been living in the United Kingdom for at least two years, were recruited via the participant recruitment platform, Prolific.com. Table A4.1 in Appendix 1 shows their demographic characteristics. Each participant was paid £0.22 to complete the study. Data were collected between 4 September and 14 November 2020.

**Figure 4.5**

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

### Task instructions

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

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

---

Economic indicators: Annual data (%)

	2019	2020
CPI Inflation rate (%)	<input type="text"/>	<input type="text"/>

---

Economic indicators: Annual data (%)

	2014	2015	2016	2017	2018	2019	2020
Unemployment rate (%)	<input type="text" value="5.50"/>	<input type="text" value="4.40"/>	<input type="text" value="3.80"/>	<input type="text" value="4.20"/>	<input type="text" value="3.20"/>	<input type="text" value="No data"/>	<input type="text" value="No data"/>
CPI Inflation Rate (%)	<input type="text" value="No data"/>	<input type="text"/>	<input type="text"/>				

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

*Stimulus materials* The screen into which participants entered their responses was similar to the one used for Experiment 3 except that each of them responded to just one table by entering their judgments for 2019 and 2020. There was no constraint on the order of responding. Examples of the response screen are shown in Figure 4.5.

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

*Procedure* The procedure was identical to that used in Experiment 3.

#### 4.4.2. Results

In this between-participants experiment, Levene's test showed that the ANOVA assumption of homogeneity of variances was violated ( $p < 0.05$ ) at each level of Year for all three dependent variables (judgment score, directional error score, absolute error score). (This was true even after outliers more than three standard deviations from the mean had been excluded.) Hence, I carried out a robust two-way mixed ANOVA (Wilcox, 2017) on each dependent variable using Information Type as a between-participants factor and Year as a

**Table 4.3**

*Experiment 5: Means and standard deviations (in parentheses) of inflation judgments, their directional errors, and their absolute errors*

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

within-participants factor. Data were analysed in R using robust tests on 20% trimmed means (to reduce skew) and a bootstrap procedure ( $nboot = 2000$ )<sup>15</sup> to obtain empirically-derived critical values ( $p < 0.05$ ) against which test statistics were compared.

As the robust analyses trim means, these ANOVAs were performed on the complete data set ( $n = 352$ ) with no outlier exclusion: 89 people (32 men, 57 women) with a mean age of 32

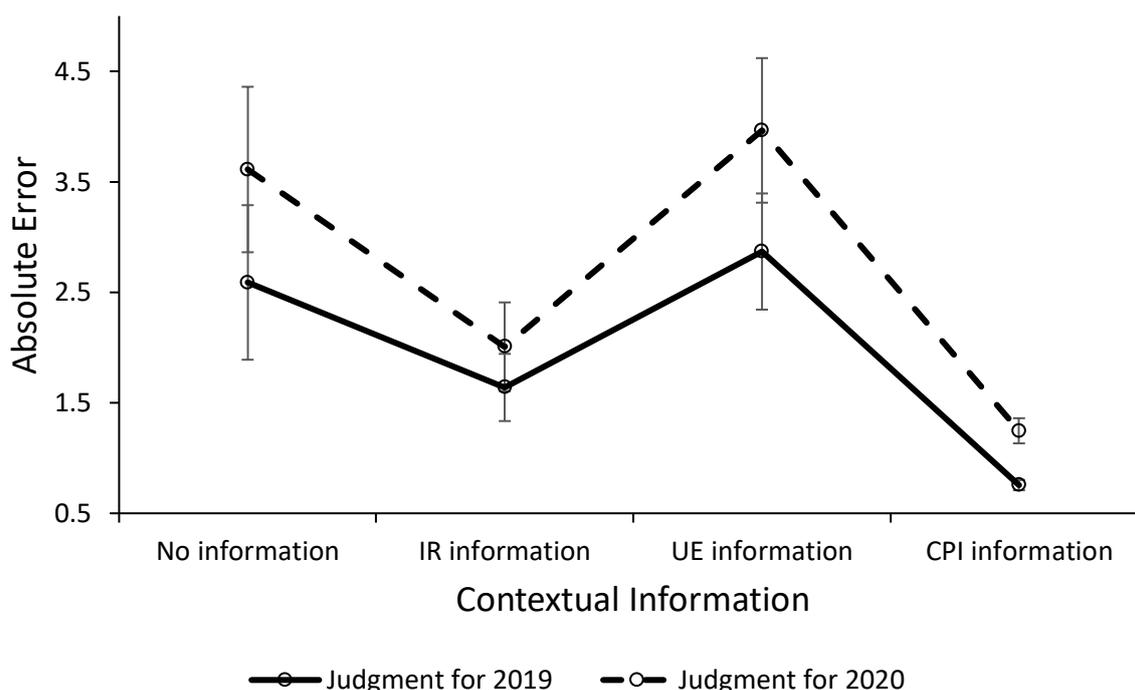
<sup>15</sup> In this section, terms in italics refer to R functions in Wilcox (2017).

years ( $SD = 11.23$  years) were in the group without additional information, 86 people (27 men, 59 women) with a mean age of 31 years ( $SD = 10$  years) were in the group with interest rate information, 86 people (23 men, 63 women) with a mean age of 33 years ( $SD = 10$  years) were in the group with unemployment rate information, and 91 people (26 men, 65 women) with a mean age of 33 years ( $SD = 11$  years) were in the group with inflation rate (CPI) information. Means and standard deviations of the three dependent variables in each of the four conditions are shown in Table 4.3.

Analysis of directional error scores using the same factors as before revealed main effects of Information Type ( $Q = 23.70$ ,  $p < 0.001$ ) and Year ( $Q = 27.53$ ,  $p < 0.001$ ) but no interaction between these variables. Post hoc analyses between each pair of Information Types revealed significant differences in every case ( $p < 0.05$ ).

**Figure 4.6**

*Experiment 5: Effects of Year and Information Type on absolute error (together with standard error bars)*



The same type of analysis performed on absolute error scores (Figure 4.6) revealed a main effect of Information Type ( $Q = 4.20$ ,  $p < 0.02$ ), a main effect of Year ( $Q = 30.36$ ,  $p < 0.001$ ), and an interaction between these variables ( $Q = 7.47$ ,  $p = 0.006$ ). The simple effect of Year at each level of Information Type was examined using Wilcox's (2017) *ydbt* function to extract bootstrap confidence intervals: absolute error was significantly greater for 2020 inflation judgments than for 2019 inflation judgments when no information was supplied ( $p = 0.038$ ), when past unemployment rates were provided ( $p = 0.003$ ), and when past CPI information

was given ( $p < 0.001$ ). Wilcox's (2017) *t1waybt* bootstrap function showed a simple effect of Information Type for both 2019 ( $Ft = 6.77, p < 0.001$ ) and 2020 ( $Ft = 6.01, p = 0.01$ ). Follow-up post hoc tests using Wilcox's (2017) *linconb* function revealed that, for 2019, all paired comparisons were significant ( $p < 0.05$ ) except that between no information and CPI information and, for 2020, they were all significant except for the comparisons between no information and CPI information and between no information and interest rate information.

#### 4.4.3. Discussion

In this between-participants experiment, absolute error was again higher for 2020 judgments than it was for 2019 judgments. This replicates the effect that I obtained in Experiment 4, where, as in the current experiment, the same participants made judgments for both 2019 and 2020. These effects of Year found in Experiments 2 and 3 contrast with the lack of such an effect in Experiment 3, where participants made judgments for just a single year – some made them for 2019 and others made them for 2020. Thus results here are consistent with a task context effect: when people make judgments for two years, their first judgment provides a context for, and thereby influences, their second one. For example, people viewing inflation as generally increasing over time will ensure that their judgment of its value for next year is higher than their judgment of its value in the current year (Tables 4.2 and 4.3). In contrast, when people make judgments for a single year, no task context effect can operate: as a consequence, inflation judgments made by people producing judgments for only next year are no different from those made by people producing judgments for just this year (Table 4.1).

There was again an effect of information context: the type of information given to forecasters influenced their inflation judgments. Specifically, when people were given information about past values of inflation, their estimates of the values of inflation later in the series were better than when they were given past values of other macroeconomic variables, such as interest rates or unemployment rates (Figure 4.6). This indicates that the effect of providing inflation rate information in the fourth judgments of Experiments 1 and 2 arose not (or not only) because people received *more* information when making later judgments in those experiments but because they receive *more useful* information when making later judgments in those experiments.

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

To use such information effectively, they would have had to make use of a mental model of the economy that was at least approximately correct. But, as Leiser and Krill (2018) have shown, they do not do this. One possibility is that they use a good-begets-good heuristic by assuming that when interest rates and unemployment rates are low, inflation is also low. This could explain why people judged inflation to be low (2.09 – 2.50 %) when they were told that interest rates were low (0.29 – 0.60%) but why the judged inflation rate to be moderate (4.28 – 5.05%) when they were told that unemployment rates were moderate (3.20 – 5.50%). This pattern of results is also consistent with Kahneman and Tversky's (1973) account of how the representativeness heuristic is used in cross-series forecasting. It may also be explained by an anchoring effect: higher judgment anchors (unemployment rates) produced higher judgments of inflation than low ones (interest rates).

#### 4.5. General discussion

Inflation judgments were systematically too high, a finding that replicates what has been found in previous studies using lay participants (Bruine de Bruin, van der Klaauw, et al., 2011; Bryan & Venkatu, 2001a, b; Georganas et al., 2014). When different people made inflation judgments for the current year or for the following year, mean values of these judgments did not differ (Experiment 3) but when the same people made judgments for both those two years, inflation judgments for 2020 were higher than those for 2019 (Experiment 4). This task context effect, triggered by joint evaluation, implies that people (wrongly) expected inflation rate to increase over time. As a result, inflation expectations for 2020 were worse than inflation perceptions for 2019 in Experiments 2 and 3.

Information context effects were found in both experiments though their nature differed somewhat. In Experiment 3, the fourth judgment, the only one that benefitted from provision of the inflation rate in the year immediately prior to the year for which inflation rate had to be estimated, was more accurate than the three earlier judgments. In Experiment 4, the fourth judgment was again superior to the previous three judgments but, in addition, the first judgment was less accurate than the three later judgments. It is likely that this difference is related to the fact that, for the fourth judgment in Experiment 4, information about the immediately preceding inflation rate could be provided only for inflation judgment for 2019; it could not be provided for the inflation judgment for 2020 because participants provided it themselves when estimating the inflation rate for 2019. In contrast, for the fourth judgment in Experiment 3, information about the immediately preceding inflation rate was explicitly provided for the inflation judgments of both 2019 and 2020. (Compare the lower panels of Figures 4.1 and 4.3.)

This suggests that Experiment 3 provides a purer comparison of the difficulties in using (and benefits arising from) the heuristics responsible for cross-series forecasting (second and third judgments) and within-series forecasting (fourth judgment). Cross-series forecasting, reliant on use of the representativeness (Harvey, 2007) and good-begets-good (Leiser & Krill, 2018) heuristics, is difficult and often ineffective (Harvey et al., 1994): comparison of the second

and third judgments with the first judgment shows that it produced little improvement over memory-based forecasting. In contrast, within-series forecasting, based on the anchor-and-adjust heuristic (Harvey, 2007) or on knowledge of temporal patterns in the ecology (Harvey & Reimers, 2013; Reimers & Harvey, 2011), is more effective: comparison of the fourth judgment with the first three judgments shows the advantages it has over memory-based and cross-series forecasting.

Information context also influenced degree of judgment homogeneity. Thus, in Experiment 3, variance of the fourth judgments was lower than that of each of the three earlier judgments and, when five years of contextual information was provided, variances of the second and third judgments were lower than the variance of the first judgment. In all conditions of Experiment 4, variance of the fourth judgment was significantly lower than that of the other three judgments and variances of the second and third judgments was lower than variance of the first judgment.

#### *4.5.1. Potential limitations*

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

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

Demographic factors, including gender, education, and financial literacy are known to influence inflation judgments (Bruine de Bruin, van der Klaauw, et al., 2010; Souleles, 2004). Differences in demographic characteristics could therefore potentially explain differences between results obtained in different experiments (including the task context effect revealed

by the difference between the first and second experiment). In fact, as Table A4.1 in Appendix 1 shows, the demographic characteristics of the samples in the three experiments were highly comparable.

#### *4.5.2. Implications*

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

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

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

Task context effects also have implications. When people judged inflation rates for two successive years (Tables 4.2 and 4.3), their estimate for the later year was higher and less accurate than it was when they made a single judgment for that later year (Table 1). In other

words, they did not make their judgment for the later year in the same way that they made it for the earlier year. Instead of making their judgment using only their memory and the contextual information they were given, they were also influenced by their expectation about how inflation would change from one year to the next. Expectations about how inflation is going to change over time adds another potential source of error to judgments of inflation. Currently, all major surveys require respondents to judge levels of inflation for a number of different years. Their responses, especially for later years, would be likely to be more accurate if they were asked for their estimate for a single year, with different respondents supplying estimates for different years.

#### *4.5.3. Conclusions*

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

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

## Chapter 5 How does feedback improve measurement of inflation expectations

### 5.1. Introduction

Human performance of high-level cognitive tasks such as the forecasting (e.g., Lawrence et al., 2006), monitoring (e.g., Brown & Steyvers, 2009; Fiedler et al., 2016), and control (e.g., Frensch & Funke, 1995; Kerstholt, 1996; Osman, 2010) of complex systems is typically sub-optimal. All these tasks involve judgment and many methods of enhancing its quality have been studied. Unaided judgmental forecasting is required in the surveys of inflation expectations that are carried out as part of the process used by central banks to forecast inflation and other important economic variables (e.g., unemployment levels, interest rates, etc.). It is reasonable to assume that estimates of inflation would be of more use to central banks if they were more accurate.

Chapter 4 investigated how consumers' inflation judgments are affected by the survey formats and showed that consumers' inflation judgments were still overestimated even additional information about inflation rate was provided when they were asked to make judgements. However, few studies investigate how the bias in inflation expectation can be reduced or eliminated.

Here I examine whether they can be improved by providing people with training using OFB. In other words, after each inflation forecast, people were told what the actual value of inflation turned out to be. I review previous work on the effects of OFB on two aspects of judgmental forecasting. First, feedback may affect the accuracy of forecasts. Second, it may affect the confidence that people have in their forecasts. Both of these are important for planning purposes.

#### *5.1.1. Effects of outcome feedback on forecast accuracy*

Judgmental forecasting relies on information held in memory, on information from a single point in time about values of a set of variables other than the one to be forecast, or on information about past values of the variable to be forecast. I term these memory-based forecasting, cue-based forecasting, and time series forecasting, respectively. Though particular forecasting tasks carried out by practitioners may draw on more than one of these types of information, they have been studied separately by psychologists. More specifically, different types of heuristic have been used to account for characteristics of performance in each case (Harvey, 2007).

Studies of memory-based forecasting have focussed on sports forecasting (e.g., Ayton et al., 2011; Pachur & Biele, 2007; Serwe & Frings, 2006), financial stock forecasting (e.g., Andersson & Rakow, 2007; Borges et al., 1999) and geopolitical forecasting (e.g., Chang et al., 2016;

Mellers et al., 2014). A major concern has been to identify the nature of the memory-based heuristic used to make forecasts. Examining effects of OFB has not been a major concern. However, Ayton et al.'s (2011) study is relevant to my work here. They asked 50 Turkish students to make forecasts for the full-time results of 32 English football matches. Once they had done this, they were given the half-time results for 19 of those matches and then they re-forecast the full-time results. The half-time results provided partial OFB. This information did influence the forecasts that people made but did not increase the accuracy of those forecasts: 62.5% were correct without feedback and 60% were correct with it.

Research on cue-based forecasting has used the multiple-cue probability learning (MCPL) paradigm (Cooksey, 1996). In MCPL tasks, people estimate the values of a criterion (e.g., future examination grades) for each of a number of instances (e.g., different students) after being given the values of a set of cues for each of those instances (e.g., past examination grades, past marks from continuous assessment, number of absences from teaching classes).

Many studies have shown that, when more than two or three cues are given or when the relation between cues and criterion is non-linear, OFB fails to produce learning (Deane et al., 1972; Goldberg, 1968; Hammond, 1971; Hammond & Boyle, 1971; Hammond & Summers, 1972). In more complex tasks, it can impair performance (Hammond et al., 1973; Holzworth & Doherty, 1976; Schmitt et al., 1976; Schmitt et al., 1977). Reviewers (e.g., Klayman, 1988; Karelaia & Hogarth, 2008) confirm this impression.

There has been some debate about the reasons for the ineffectiveness of OFB in MCPL tasks. In these tasks, a random error term is added to (or is assumed to be present in) the expression that generates the criterion from the cue values. Brehmer (1980, p233) claimed that "people simply do not have the cognitive schemata needed for efficient performance in probabilistic tasks". In contrast, Todd and Hammond (1965) suggested that OFB fails because it does not specify what people need to do in order to improve their performance: finding that a forecast for an examination was too high does not tell the forecaster how to change the weights on the different cues. However, in advice tasks where cues are suggestions for the values of the criterion made by different advisors, OFB does provide information about how to re-weight the cues. The finding that OFB does permit more rapid learning in such tasks than in MCPL tasks (Fischer & Harvey, 1999; Harries & Harvey, 2000) favours Todd and Hammond's (1965) approach over that of Brehmer (1980).

There is a problem with studying OFB in time series forecasting. It is that, for a given time series, recurrent forecasting of a particular series means that the actual outcome that has just been forecast must be revealed so that it can be used when making the next forecast. Thus, OFB is always present. As a result, OFB conditions have been used as a baseline control against which the effectiveness of more elaborate types of feedback, such as performance feedback<sup>16</sup>,

---

<sup>16</sup> For example, forecasters could be informed of their root mean squared prediction error.

have been assessed (e.g., Goodwin & Fildes, 1999; Remus et al., 1996). In considering the effectiveness of these different forms of feedback, Lawrence et al. (2006, pp. 507-8) conclude that “studies have tended to show that outcome feedback is the least effective form” and, following Klayman (1988), they suggest that this “is probably because the most recent outcome contains noise and hence it is difficult for the forecaster to distinguish the error arising from a systematic deficiency in their judgement from the error caused by random factors”. In other words, OFB is relatively ineffective because forecasters have difficulty filtering out the contribution that system noise makes to the values of the realized outcomes of the variable being forecast.

The question that remains to be answered is whether OFB is effective to any degree. Does it facilitate forecasting, impair forecasting, or have no effect whatsoever? As far as I can ascertain, no studies have addressed this issue. Yet it has clear relevance to forecasting practice. For example, after sales forecasts have been made and used for planning purposes, it is often the case that they are not retained. Is this a sensible strategy or would improvements in forecast accuracy produced by OFB be worth the additional costs of retaining forecasts and comparing them with outcomes? To determine whether OFB is effective, the standard paradigm of requiring recursive forecasts from the same data series is of no use for the reasons that I outlined above. Instead, a single forecast (or a set of simultaneous forecasts) must be made from each of a number of data series by participants who either receive OFB or do not do so. It is this approach that I take in the experiments reported below.

### *5.1.2. Effects of outcome feedback on confidence in forecasts*

Confidence in forecasts is usually assessed by asking people to estimate the probability that their forecast will be correct or that it will fall within certain pre-specified bounds. The probabilities they give can then be compared with objective frequencies obtained from a sample of their judgments. For each forecast,  $f$ , (e.g., there is a 60% chance of my sales forecast being within 20% of the actual sales), I set an outcome index,  $d$ , at 1.00 when the event occurs (the forecast is within 20% of the outcome) and at 0.00 when the event does not occur. The probability score,  $PS$ , is then calculated from the forecast (expressed as a probability rather than a percentage) as follows:  $PS = (f - d)^2$ . The probability score is also known as the Brier score (Brier, 1950) and gives a measure of the quality of these confidence judgments with lower mean probability scores indicating better judgments. Someone who judges that all forecasts as likely to be correct as not and therefore always provides a probability judgment of 0.5 (a uniform judge) obtains a mean probability score of 0.25. Overconfidence is measured by the mean value of  $(f - d)$ .

Brier scores, together with bias scores, give me a way of assessing the overall accuracy of probability judgments. Brier scores are often referred to a measure of ‘calibration-in-the-large’. With a sufficient data points per forecaster, they can be usefully decomposed into

components. The commonly used Murphy decomposition divides the Brier score into incidence (a measure of the relative frequency of the target event), discrimination or resolution (a measure of ability to distinguish between target and non-target events), and calibration (a measure of ability to match labels for subjective probability to values of objective frequencies). The covariance decomposition (Yates, 1982; Yates & Curley, 1985) is useful for providing insights into the psychological processes underlying probability estimation. It divides the mean probability score into incidence, bias, separation (mean forecast when target event is present minus mean forecast when it is absent), and scatter (judgment noise).

Much research indicates that a) people are typically overconfident in their decisions, and b) overconfidence is greater with harder decisions, with under-confidence present with very easy ones (the hard-easy effect). However, much of the original research supporting these conclusions derived from studies that required people to give a probability (50-100%) that each of their answers to a set of two-alternative general knowledge questions was correct (e.g., Lichtenstein et al., 1982). Later work (e.g., Gigerenzer et al., 1991; Juslin, 1994) implies that these findings arose because experimenters had included too many misleading questions (i.e., those with counterintuitive answers) in their experiments. When questions were made more representative of the natural ecology by randomly selecting them from a population of all possible questions of a particular type, overconfidence greatly diminished. However, overconfidence and the hard-easy effect have been found in many other tasks where the ecological critique is harder to sustain (Harvey, 1994). These include perceptual tasks (e.g., Baranski & Petrusic, 1994) and motor tasks (e.g., Cohen et al., 1956; Cohen and Dearnley, 1962). However, in these reports, the effects do not appear to be as large as those reported in the original studies of confidence in answers to general knowledge questions.

There has been some debate about whether people are overconfident in their forecasts. Fischhoff and MacGregor (1982) found that confidence in forecasts had similar characteristics to confidence in answers to general knowledge questions (i.e., general overconfidence together with a hard-easy effect), though there were fewer 100% certain responses in the former case. However, Wright (1982) found that people were less overconfident in their forecasts about future events than in their answers to equally difficult questions about whether similar events had already occurred. Also, Ronis and Yates (1987) found that were less overconfident in their forecasts of basketball games than in their answers to general knowledge questions. In summary, it appears that forecasters are still biased towards overconfidence but not to the same degree as those answering general knowledge questions. This difference is likely to reflect the validity of the ecological critique of studies using general knowledge questions (Gigerenzer et al., 1991; Juslin, 1994).

A few studies have examined whether OFB reduces overconfidence or otherwise improves the quality of confidence judgments. For example, over four weekly sessions, Benson and Önkal (1992) asked people to forecast which team would win in each of 55 major college

football games and then to express their confidence in their forecast as a probability between 0.5 and 1.0. In one condition, people received OFB at the start of the last three sessions. This listed their predictions and probability assessments from the previous week, along with the actual scores in the 55 games. Results showed that the mean probability score, calibration, and scatter deteriorated over the four weeks and that there was no change in resolution or slope. However, bias in confidence judgments was reduced. This is certainly suggestive of an effect of OFB. Unfortunately, there was no control group in which OFB was not given: the effects reported could have occurred as a function of experience over time in the absence of OFB.<sup>17</sup>

Subbotin (1996) asked people to go through a list of 100 pairs of European capitals or 100 pairs of European countries and, for each one, a) decide which was larger and then b) estimate the probability (50%-100%) that their answer was correct. When decisions were easy, people were under-confident and simple OFB reduced this bias. However, when decisions were difficult, people were overconfident and simple OFB had no effect. Furthermore, when Subbotin (1996) repeated his experiment using unrelated questions that did not all refer to the same variable (city or country size), OFB had no effect irrespective of question difficulty.

In Winman and Juslin's (1993) experiment, people in one condition answered a series of related questions about which of two causes of death was the more common among Swedes and those in another condition decided which was the heavier of two weights. After each answer, people in both groups estimated the likelihood it was correct (50-100%). Feedback was given after each of the 40 trials in the middle four of six trial blocks. (Participants were told whether their decision had been right or wrong and, so, strictly speaking, they were given performance feedback rather than OFB.) There was also a control condition in which people performed the causes of death task without any feedback. An initial bias towards overconfidence in the causes of death task disappeared by the fourth block when feedback was given but remained throughout the experiment when it was not supplied. However, feedback had no effect on the under-confidence bias observed in the weight discrimination task.

Keren (1988) required people to carry out two tasks with or without OFB. In one, they answered unrelated general knowledge questions and, after each response, gave a confidence rating (50-100%). In the other, they decided whether a gap in a perceptual stimulus was on the left or right and gave the same type of confidence rating. This perceptual task was hard (small gaps with 67% correct and slight under-confidence) or easy (large gaps with 79% correct and high under-confidence). The difficulty of the general knowledge task was midway between that of the perceptual tasks (71% correct with overconfidence). OFB

---

<sup>17</sup> Various studies, including Benson and Önköl (1992), have examined effects of other types of feedback in which people were told of their accuracy, their calibration scores, their resolution scores, and other highly processed performance measures (e.g., Adams & Adams, 1958; Lichtenstein & Fischhoff, 1980; Sharp et al., 1988; Stone & Opel, 2000). Obtained effects have generally been modest (Keren, 1991).

had no effect on under-confidence in the perceptual tasks, whatever their difficulty, or on overconfidence in the general knowledge task.<sup>18</sup>

In their review of this issue, Russo and Schoemaker (1992, p. 11) state that: “We believe that timely feedback and accountability can gradually reduce the bias toward overconfidence in almost all professions. *Being ‘well-calibrated’ is a teachable, learnable skill*” (italics are theirs). However, in commenting on this statement, McClelland and Bolger (1994, p. 476) argue that “the evidence that outcome feedback alone is effective in reducing miscalibration is not encouraging”. Nevertheless, this evidence has some structure. OFB has no effect on biased confidence in answers to unrelated general knowledge questions (Keren, 1988; Subbotin, 1996) or on biased confidence in perceptual decisions (Keren, 1988; Winman & Juslin, 1993). However, OFB has been found to decrease biased confidence in answers to related questions (Benson & Önköl, 1992; Subbotin, 1996; Winman & Juslin, 1993).

There is one report that does not fit this neat picture. Zakay (1992) examined groups of people who answered general knowledge questions and who, after each answer, estimated the probability (0.5-1.0) that it was correct. Half carried out the experiment as a pencil-and-paper test and, for the other half, the experiment was run on computers. Within each of these groups, half the people received OFB after each answer and the other half did not. Feedback reduced overconfidence in the computerized version of the task but not in the paper-and-pencil version. The result in the former case is an outlier in that all other reports of experiments with unrelated items failed to show an effect of feedback on overconfidence.

### 5.1.3. Identifying the nature of outcome feedback effects

OFB may have two main effects (Annett, 1969). First, it may incentivise people to perform better: they may put more effort into tasks when they know they will find out how well they have performed. This effect disappears when feedback is removed. Second, feedback may produce learning that improves people’s forecasting ability: this effect is maintained when feedback is removed. To distinguish between these effects, two groups of participants must be used in experiments. One group should receive OFB in a first session but not in a second session; another group should receive feedback in neither session. A difference between the two groups in the first session shows a beneficial effect of feedback but does not identify the nature of this effect. Disappearance of the beneficial effect of feedback in the second session indicates that it was due to incentivisation. Evidence that it is fully maintained implies that it

---

<sup>18</sup> In addition, Önköl and Muradođlu (1995) and Fischer (1982) found no effects of OFB in full-range probabilistic forecasting tasks. In such tasks, people simply assess the probability (0-100%) that an event will happen; no separate confidence judgment is made. Keren (1991) recommends avoiding use of these tasks in the study of confidence because they suggest to participants that they are to consider aleatory uncertainty inherent in the system being assessed rather than the epistemic uncertainty present in their own minds.

was due to learning. Some partial maintenance of the effect would suggest that it arose because both incentivisation and learning had a role in producing it.

Incentivisation increases speed or accuracy in those tasks in which greater mental effort is effective in improving performance (Camerer & Hogarth, 1999; Lerner & Tetlock, 1999). The possibility that this explains OFB effects (e.g., by increasing the attention paid to critical aspects of the task) has been recognised (e.g., Adams & Adams, 1958; Zakay, 1992). However, only Winman and Juslin (1993) used an experimental design of the sort described above to determine whether incentivisation or learning produced the OFB effects that they obtained. They found that these effects were fully maintained after feedback was removed, thereby indicating that they were produced by learning. However, as Zakay (1992) himself pointed out, the anomalous finding that he reported may have arisen not through learning but because computer-based feedback incentivised people to pay attention to critical aspects of the task more than paper-based feedback did. If this is correct, there would have been no permanent effects of OFB in the computerized version of his task if he had examined performance after that feedback had been removed.

## **5.2. Experiment 6**

The experiment was aimed to determine whether OFB reduces laypeople's overestimation of inflation and whether it reduces their overconfidence in their inflation estimates. It used the design outlined above to determine whether any obtained effects arise from training-induced learning rather than from incentivising effects of OFB.

On the basis of the previous work reviewed above, my first hypothesis (H1) was that lay estimates of inflation will be too high. I am interested in whether OFB reduces this overestimation. I have seen that studies have shown that there is no evidence that OFB improves memory-based forecasts. There is some evidence that it benefits cue-based forecasts but only when few cues are combined in a simple way. No studies of effects of OFB on time series forecasting have been reported. However, as this type of forecasting is closer to cue-based forecasting than to memory-based forecasting (because some external information is provided), I tentatively hypothesise that OFB will improve time series forecasting (H2).

We have seen that people tend to show overconfidence in their responses to related items (though this overconfidence is not as great as when they respond to unrelated items). My participants made forecasts for each member of a sequence of time series showing inflation rates in different countries. As these were related items, I expected a modest degree of overconfidence in these forecasts (H3). Furthermore, I saw above that evidence indicates that OFB reduces overconfidence in responses to related items. As my task involved responses to related items, I expected it to have that effect here. Specifically, I expected it to lower confidence and, as a result, reduce bias and improve calibration (H4). Finally, because of the

hard-easy effect, I considered that overconfidence would be greater for series that were more difficult to forecast (H5).

Laypeople were presented with a tabular record of ten years (2009-2018) of inflation figures (Consumer Price Index, CPI)<sup>19</sup> and required to make a one-step ahead forecast for 2019 and to express their confidence (0-100%) that this forecast was within 20% of the actual level of inflation. They carried out this task for a sequence of 20 anonymised countries. The Feedback group received OFB for the first set of ten countries (session 1) but not for the second set of ten (session 2). The No-feedback group received no feedback in either session. (I use the term 'session' rather than 'trial block' because the two sets of trials were separated by a new instruction page that informed participants that they would not receive feedback information and that they would now be given a bonus payment related to their performance.)

### 5.2.1. Method

*Participants* One hundred and two participants (34 females, 68 males) with a mean age of 27 years (SD = 9 years). They were recruited via the web platform Prolific.com between 24 and 27 October 2020. They were paid £ 1.10 each for their participation. Additionally, in the second session, they were given a bonus of £0.10 whenever their inflation estimate for a country was within 20% of the correct value. The study was approved by UCL Department of Experimental Psychology Ethics Committee.

*Design* This was a mixed design experiment. The between-participants factor was participant group (Feedback group versus No-feedback group) and the within-participant factors were Session (first versus second session) and Trial (1-10). OFB was provided only to the Feedback group in the first session. Within each session, participants made judgments for ten countries that were presented in a different random order for each participant.

*Stimulus materials* Eleven years (2009-2019) of historical data were extracted from the World Bank dataset of annual consumer price inflation (CPI) for 20 countries (Armenia, Australia, Belgium, Brazil, China, Colombia, Denmark, El Salvador, Fiji, Finland, France, Greece, India, Ireland, Japan, South Korea, Peru, Thailand, the United Kingdom and the United States). On each trial, the first ten of these figures (2009-2018) were shown in a table for one of the countries. Figure 5.1 shows an example of one of these tables. For each participant, ten out of 20 countries were randomly allocated to the first session and randomly ordered within that session. The remaining 10 countries were randomly ordered to form the second session.

Participants made inflation estimates by completing the 2019 cell at the right-hand end of the table. After doing this, they were asked to move a slider to provide their judgment of the likelihood (0-100%) that their inflation estimate was within 20% of the correct value. In the first session, participants in the experimental group then moved on to a feedback screen. As

---

<sup>19</sup> Surveys such as the Federal Reserve Bank of Philadelphia's SPF provide respondents with values of inflation prior to the year for which their forecast is required.

an example, consider a trial in which the participant had estimated inflation for 2019 to be 1.11% when it was actually 0.48%. Feedback was given in two lines of text. In this example, the upper one would have read “The actual consumer price inflation (2019) for this country is 0.48%” and the lower one would have read “You estimated consumer price inflation (2019) for this country to be 1.11%”. In the second session, participants in the experimental group did not receive feedback but moved directly on to the next trial. Participants in the No-feedback group did not receive feedback in either session.

*Procedure* After reading the information sheet and completing the consent form, participants read the task instructions. These instructions were also provided on each trial (Figure 5.1). Before starting the task, participants were provided with a simple definition of consumer price inflation, together with an example.

### Figure 5.1

*A table showing annual inflation figures between 2009 and 2018 and an empty cell for participants to enter their inflation estimate for 2019*

**First set of ten countries : Country 1**

---

**Task instructions**

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

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

---

Consumer price inflation (annual %)

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Country O	-1.35	-0.72	-0.27	-0.05	0.35	2.76	0.79	-0.12	0.47	0.98	

---

In the first session, participants in the Feedback group but not those in the No-feedback group received OFB in the format described above. In the second session, participants in neither group received OFB. At the end of the experiment, participants provided demographical details, including their age, gender, highest level of education qualification attained, economic related knowledge and work experience.

#### 5.2.2. Results

Fifteen outliers whose judgments were more than three standard deviations from the mean values of the judgments for the 20 countries were excluded. The sample analysed below therefore comprised 87 participants (59 men, 28 women) with a mean age of 27 years (SD= 8 years). Forty-six of them were in the Feedback condition and 41 of them were in the No-feedback condition.

Trial-by-trial data for inflation judgments and confidence judgments are shown in Table A5.1 of the Appendix 2. Means of these inflation judgments and their associated accuracy levels across all 20 countries in the experiment, together with corresponding values produced by two simple algorithmic models, are given in Table A5.4 of the Appendix 2.

**Table 5.1**

*Experiment 6: Means and standard deviations (in parentheses) of inflation judgments, their total errors, constant errors and variable errors*

	Feedback condition	No-feedback condition	mean
<b>Judgment</b>			
Session 1	2.00(0.42)	2.17(0.40)	2.08(0.41)
Session 2	1.98(0.34)	2.14(0.42)	2.06(0.38)
mean	1.99(0.22)	2.15(0.28)	2.07(1.38)
<b>Absolute error</b>			
Session 1	0.89(0.27)	0.88(0.31)	0.89(0.29)
Session 2	0.87(0.19)	0.88(0.22)	0.87(0.21)
mean	0.88(0.14)	0.88(0.19)	0.88(0.86)
<b>Constant error</b>			
Session 1	0.20(0.25)	0.39(0.43)	0.29(0.36)
Session 2	0.24(0.36)	0.37(0.35)	0.30(0.36)
mean	0.22(0.22)	0.38(0.28)	0.29(1.19)
<b>Variable error</b>			
Session 1	0.81(0.31)	0.69(0.25)	0.75(0.29)
Session 2	0.75(0.22)	0.66(0.21)	0.71(0.22)
mean	0.78(0.16)	0.68(0.12)	0.73(0.68)

*Forecasting performance* To obtain an overall measure of accuracy for each participant on each trial, I used the absolute error (AE): this is the absolute difference between the relevant value of inflation given by the World Bank for 2019 and the value estimated by the participant. The algebraic difference between the same two values provided my measure of constant error (CE), also known as directional error or bias. Constant error is one of two types of error that contributes to overall error. The other is variable error (VE), also known as scatter, noise, or inconsistency. This variable error adds random noise to the measure of CE that is obtained by taking the algebraic difference between the correct and judged value of inflation. To obtain a measure of the size of VE for each participant on each trial, I proceeded as follows. Separately for each participant in each session, I fitted a trend line with linear and quadratic components to the values of CE. I then extracted the residuals from these regressions and used the absolute value of the residual on each trial as my measure of VE on that trial. Mean values of raw inflation judgments and these three error measures are shown in Table 5.1 for the two sessions and the two conditions of the experiment.

To address the issue of whether people tend to over-estimate future inflation rates (H<sub>1</sub>), I used one-sample t-tests to determine whether the mean CE values in the first and second sessions completed by the Feedback and No-feedback groups were significantly different from zero. In each case, the test was significant: Feedback group, session one ( $t(45) = 5.51, p < 0.001$ ); Feedback group, session two ( $t(45) = 4.47, p < 0.001$ ); No-feedback group, session one ( $t(40) = 5.75, p < 0.001$ ); No feedback group, session two ( $t(40) = 6.79, p < 0.001$ ). These results confirm those of many previous studies: laypeople tend to overestimate inflation rates.

To examine the effects of OFB on inflation judgments (H<sub>2</sub>), I carried out three-way mixed ANOVAs on the raw inflation judgments, AE, CE, and VE with Condition (feedback versus no-feedback) as a between-participant factor and with Session (one versus two) and Trial within Sessions (1-10) as within-participant factors.

The analysis of the inflation judgments showed a main effect of Condition ( $F(1, 85) = 9.06, p = 0.003, ges = 0.0035$ )<sup>20</sup>: participants in the no-feedback group estimated inflation to be higher than those in the feedback group. There was also a main effect of Trial ( $F(8.10, 688.50) = 2.79, p = 0.005, ges = 0.0141$ ). As Table A5.1 shows, this arose because participants in both groups estimated inflation to be higher at the beginnings and ends of each session.

Analysis of AE showed no main or interactive effects of any of the three variables (Table 5.1). Analysis of CE showed only a main effect of Condition ( $F(1, 85) = 9.03, p = 0.003, ges = 0.0046$ ): participants in the no-feedback group overestimated inflation more than those in the feedback group (Table 5.1). Analysis of VE also showed only a main effect of Condition ( $F(1, 85) = 11.13, p = 0.001, ges = 0.0060$ ) but it was in the opposite direction to that observed for CE: participants in the feedback group made noisier judgments than those in the no-feedback group (Table 5.1).

Taken together, these analyses suggest that feedback had no effect on overall error (AE) because its effects on CE and VE were in opposite directions.

*Confidence in forecasts* Table 5.2 shows mean values of confidence judgments, bias scores, and calibration scores (i.e.,  $1 - \text{Brier score}$ ) for both conditions and both sessions. (I transformed the Brier score to produce the calibration score so that higher scores indicate better performance.)

To investigate whether people were overconfident in their inflation estimates (H<sub>3</sub>), I used one-sample t-tests to determine whether the mean bias scores in the first and second sessions completed by the Feedback and No-feedback groups were significantly different from zero.

---

<sup>20</sup> When Mauchy's test showed a deviation from sphericity, Greenhouse-Geisser corrections were used to adjust degrees of freedom. Generalised eta squared (*ges*) measured effect size (Olejnik and Algina, 2003).

**Table 5.2**

*Experiment 6: Means and standard deviations (in parentheses) of confidence judgments, their bias scores, and calibration scores*

	Feedback condition	No-feedback condition	mean
<b>Confidence judgment</b>			
Session 1	48.66(18.91)	56.20(18.91)	52.21(19.18)
Session 2	47.49(20.20)	59.99(19.15)	53.38(20.57)
mean	48.08(19.19)	58.09(18.50)	52.80(23.33)
<b>Bias</b>			
Session 1	0.24(0.24)	0.29(0.22)	0.26(0.23)
Session 2	0.24(0.24)	0.38(0.23)	0.30(0.24)
mean	0.24(0.22)	0.33(0.20)	0.28(0.49)
<b>Calibration</b>			
Session 1	0.70(0.10)	0.66(0.10)	0.68(0.10)
Session 2	0.72(0.12)	0.64(0.14)	0.68(0.13)
mean	0.71(0.10)	0.65(0.11)	0.68(0.24)

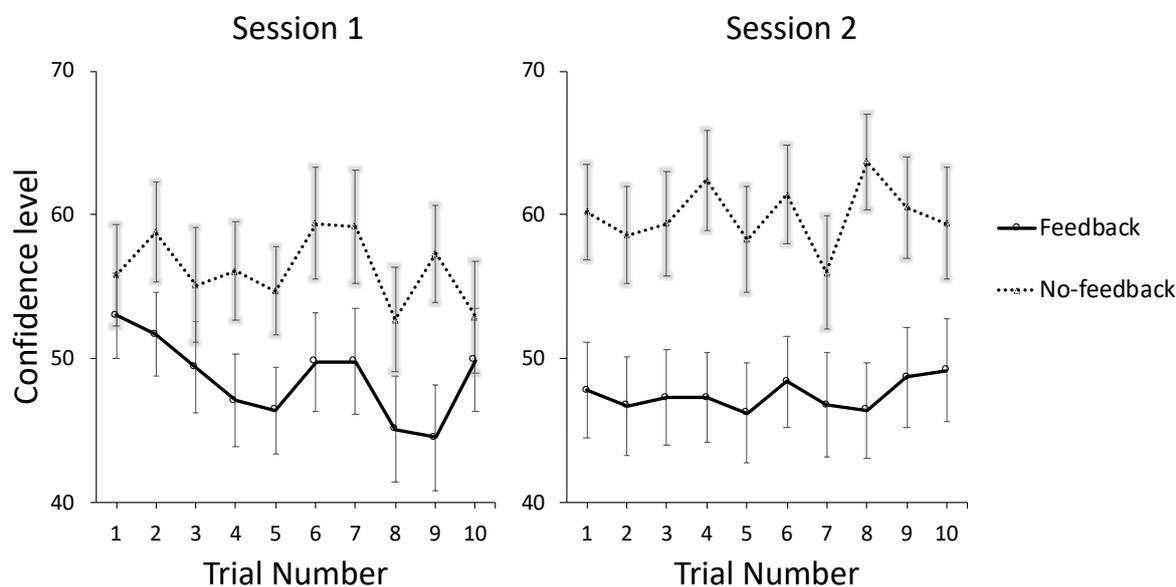
In each case, the test was significant: Feedback group, session one ( $t(45) = 6.60, p < 0.001$ ); Feedback group, session two ( $t(45) = 6.71, p < 0.001$ ); No-feedback group, session one ( $t(40) = 8.43, p < 0.001$ ); No-feedback group, session two ( $t(40) = 10.53, p < 0.001$ ). These results are consistent with the hypothesis that laypeople are overconfident in their estimates of inflation rates.

To examine the effects of OFB on confidence ( $H_4$ ), three-way mixed ANOVAs were carried out on the confidence variables using the same factors as those used for the analyses of forecasts. The analysis of confidence level showed an effect of Condition ( $F(1, 85) = 6.10, p = 0.015, ges = 0.0464$ ) and an interaction between Condition and Session ( $F(1, 85) = 7.85, p = 0.006, ges = 0.0030$ ). Further investigation showed that the simple effect of Condition was significant in Session 2 ( $F(1, 85) = 8.71, p = 0.004, ges = 0.0929$ ) but not in Session 1 and that the simple effect of Session was significant for the No-feedback condition ( $F(1, 40) = 7.41, p = 0.01, ges = 0.0101$ ) but not for the Feedback condition. As shown in Figure 5.2, these effects arose because participants in the No-feedback condition had higher confidence in their forecasts than those in the Feedback condition and because only participants in the No-feedback condition increased their confidence from session one to session two.

Finally, there was an interaction between Trial and Session ( $F(7.64, 649.49) = 2.15, p = 0.03, ges = 0.0035$ ). This arose because a simple effect of Trial was significant in Session 1 ( $F(6.89, 592.88) = 2.13, p = 0.04, ges = 0.0077$ ) but not in session 2. Confidence dropped over Session 1 but showed no further change over Session 2. These effects are shown in Figure 5.2.

**Figure 5.2**

*Experiment 6: Graph of mean confidence level (with standard error bars) in the two sessions for each group*



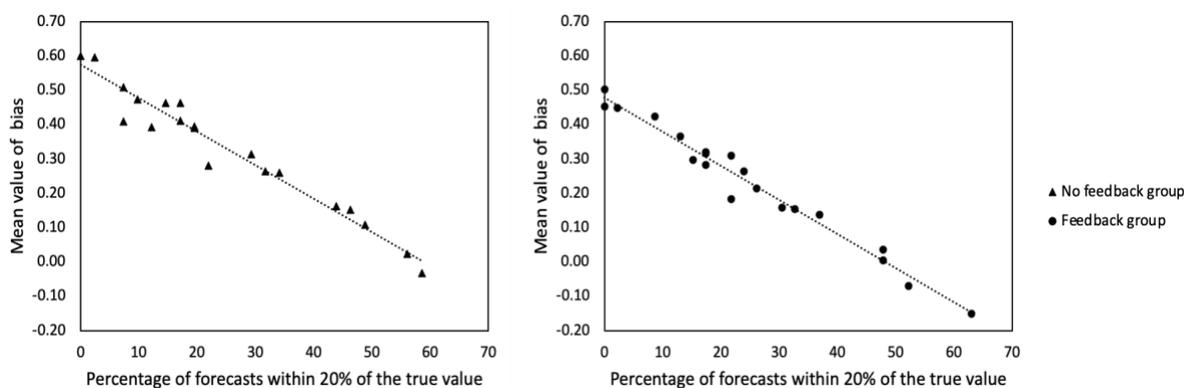
The analysis of Bias showed a main effect of Condition ( $F(1, 85) = 4.52, p = 0.036, ges = 0.0098$ ): overconfidence was greater in the No-feedback condition than in the Feedback condition. There also an interaction between Condition and Session ( $F(1, 85) = 4.07, p = 0.047, ges = 0.0020$ ). Tests of simple effects showed that the effect of Condition was significant only in session two ( $F(1, 85) = 7.74, p = 0.007, ges = 0.0835$ ) and that the effect of Session was significant only in the No-feedback condition ( $F(1, 40) = 7.79, p = 0.008, ges = 0.0370$ ). These effects show that, overall, participants were more overconfident in the No-feedback condition than in the Feedback condition and that this was because only participants in the No-feedback condition showed greater overconfidence in session two than in session one (Table 5.2).

Analysis of the Calibration scores showed a main effect of Condition ( $F(1, 85) = 6.50, p = 0.01, ges = 0.0139$ ) and a marginally significant interaction between Condition and Session ( $F(1, 85) = 3.80, p = 0.054, ges = 0.0022$ ). Further analyses showed that the simple effect of Condition was significant only in Session 2 ( $F(1, 85) = 8.16, p = 0.005, ges = 0.0876$ ) and that the simple effect of Session was not significant for either Condition. These effects are shown in Table 5.2.

Finally, I investigated whether there was a hard-easy effect ( $H_5$ ). For the Feedback group, a regression showed that  $Bias = 0.48 - 0.01 \text{ Difficulty}$  ( $Adj R^2 = 0.97; F(1, 18) = 655.80, p < 0.001$ ); for the No-feedback group, the corresponding regression showed  $Bias = 0.57 - 0.01 \text{ Difficulty}$  ( $Adj R^2 = 0.95, F(1, 18) = 346.67$ ). It is clear from these analyses that OFB affected the intercept (overall overconfidence) but had no effect whatsoever on the slope (hard-easy effect) as shown in Figure 5.3.

**Figure 5.3**

*Experiment 6: Bias plotted against forecast difficulty in the no-feedback group (left panel) and the feedback group (right panel)*



### 5.2.3. Discussion

People given the previous ten years of inflation levels for a set of countries overestimated the levels of inflation for the following year in those countries. This finding confirms reports of overestimation of inflation in a number of previous studies (e.g., Bruine de Bruin, Van der Klaauw, et al., 2011). However, in those studies, people were typically required to estimate future inflation without information about previous levels of inflation. I have shown that, even with that information, overestimation still occurs. However, provision of OFB over a set of 10 inflation judgments almost halved the degree of the overestimation. Furthermore, this effect continued once feedback had been removed. I conclude that it was produced via learning rather than incentivisation.

Despite this beneficial effect of OFB, overall error did not decline. This was because feedback had a detrimental effect in addition to its beneficial influence on bias: it increased the scatter or random error in judgments. This source of error in judgments has been recognised for some time (Thurstone, 1926) and has been examined thoroughly by judgment analysts working on multiple-cue probability learning (e.g., Brehmer, 1978). It explains why people make different judgments when presented with the same information on different occasions. However, I did not expect feedback to increase variable error. What could have produced this effect?

When people make a sequence of forecasts from a volatile series of independent points, their forecasts should lie on the mean value of, or on the trend line through, those points. Instead, they scatter their sequence forecasts around the mean or trend line. People making forecasts from time series tend to add noise to those forecasts; the noise they add is higher when the series noise is higher but still somewhat less than the series noise (Harvey, 1995; Harvey et al., 1997). It appears that forecasters aim to simulate the noise (as well as the pattern) in the series but do not fully succeed in doing so. It appears that providing people with feedback gives them the impression that the volatility in the series from which they had forecast was

greater than they had assumed when making their forecast. As a result, they attempt to simulate the noise in the series more accurately by increasing the scatter in their forecasts for later series.

People were overconfident in their forecasts. Without OFB in the first session, overconfidence became even higher in the second session. People expected experience at the task, even in the absence of feedback, to improve their performance (Harvey & Fischer, 2005). As it did not do so, experience at the task produced an increase in overconfidence. However, provision of OFB in the first session disabused them of this notion. As a result, their overconfidence did not increase across sessions. Again, OFB continued to have an effect even once it had been removed: its influence can therefore be characterised as one of learning rather than as a product of incentivisation. Consistently with H<sub>4</sub>, OFB lowered confidence, reduced bias, and improved calibration.

Analyses of inflation showed effects of OFB on both inflation judgments and confidence judgments but there were no significant interactions showing that these effects were greater later in the first session. In other words, there was no dose-response effect: more feedback was no more effective than less feedback. It appears that all that mattered was that participants received some feedback. I examine this issue in the next experiment.

### 5.3. Experiment 7

Experiment 6 showed beneficial effects of OFB but failed to show that more feedback was more beneficial. There are two possibilities. First, though the experiment had sufficient statistical power to reveal the main effect of feedback condition, it may not have had the power needed to pick up the three-way interaction. Second, it may be that a lot of feedback is indeed no more beneficial than just a little feedback. This somewhat counterintuitive notion receives support from the literature on skilled behaviour (Harvey, 2011). A number of studies using positioning tasks employed the same type of experimental design that I have adopted here: in a first (learning) session, groups performed the task under different feedback conditions; in a second (retention) session, all groups performed without feedback. I will briefly consider these studies.

During learning, Schmidt et al. (1989) gave different groups either OFB on every trial or summary OFB after every five, 10, or 15 trials; the summaries described the errors that they had made since they last received feedback. During learning, error was greater when feedback was less frequent but, at retention, this effect reversed: error was *lower* with less frequent feedback. During learning, Wulf and Schmidt (1989) provided one group with simple OFB on every trial and another group with simple OFB on two-thirds of the trials. Performance in the two groups was not significantly different during learning but those who had received *less* feedback performed better at retention. Winstein and Schmidt (1990) replicated this pattern of results in experiments in which feedback on all trials during learning was compared

with feedback on 50% and 33% of trials. Schmidt and Bjork (1992) argued that these counterintuitive effects occur because making the task more difficult by reducing feedback causes people to put more effort into their task. This additional effort pays off once feedback has been removed in the retention session.

In the experiment that I report next, participants were randomly allocated into two sets of four groups. The two sets differed according to the type of feedback they received in the first session. Those in the first set received either no feedback ('no-feedback') or else simple OFB after every trial (i.e., ten times - '10-feedback'), after every other trial (i.e., five times - '5-feedback'), or after every five trials (i.e., twice - '2-feedback'). Those in the second set received either no feedback or else summary OFB after every trial ('10-feedback'), after every other trial ('5-feedback'), or after every five trials ('2-feedback'). No group received feedback in the second session.

Following the findings reviewed above, I expected inflation judgments during session one to be better with feedback than without it but to be no worse when feedback is given less often ( $H_6$ ). Also, though studies in the literature on skill learning have not, as far as I am aware, directly compared effects of summary OFB (Schmidt et al., 1989) with simple OFB (Wulf & Schmidt, 1989), I expected judgments to be better in the former case ( $H_7$ ). This was because summary OFB averaged over a number of trials provides information about the quality of more past judgments and is less subject to random noise than simple OFB.

### 5.3.1. Method

*Participants* Four hundred and fifty-one participants (277 males) were recruited with the mean age of 26 years ( $SD = 13$  years) via the online platform, Prolific.com. Data from two participants were excluded because their responses were incomplete. Each person was paid £1.00 for their participation and an additional £0.10 for every inflation judgment in the second session that was within 20% of the correct value. Data were collected from 23 December 2020 to 8 March 2021.

*Design* This was a mixed design experiment with Session (first versus second) and Trial (1-10) as within-participant variables and Feedback Type (simple OFB versus summarised OFB) and Feedback Amount (no-feedback, 2-feedback, 5-feedback, 10-feedback) as between-participant variables. Participants were randomly allocated to one of these eight conditions.

*Stimulus materials* Inflation data for 20 countries were identical to those used in Experiment 6. Information on the feedback page varied according to feedback condition. The simple OFB page was identical to that used in the first experiment: it showed the actual inflation rate and the participant's estimated inflation rate for the immediately preceding trial. The summarized OFB page showed the actual inflation rate averaged over one, two, or five previous trials (depending on the condition), the *averaged* inflation judgment averaged over one, two, or five previous trials (depending on the condition), and the difference between these two

values. An example of a summarized OFB page is “Over the last two countries, the average inflation rate was 2.32% and your average forecast was 2.84%, so you overestimated by 0.52% on average”.

*Procedure* The procedure was identical to Experiment 6. The only differences were in the amount of OFB given (none, after every five trials, after every other trial, after every trial) and its type (simple versus summary).

### 5.3.2. Results

Participants' data were excluded when their inflation judgment for any of the 20 countries was beyond three standard deviations of the group mean judgment of that country. Implementation of this exclusion criterion resulted in data from 388 participants (241 males) with a mean age of 26 years ( $SD = 9$  years) being entered into the analyses. Trial-by-trial data of mean inflation judgments and mean confidence judgments for each of the eight groups are shown in Tables A5.2 and A5.3 of the Appendix 2, respectively. Means of these inflation judgments and their associated accuracy levels across all 20 countries in the experiment, together with corresponding values produced by two simple models, are given in Table A5.4 of the Appendix 2.

*Forecasting performance* Measures of forecasting performance (absolute error, constant error, variable error) were derived in the same way as in the first experiment.

Sixteen one-sample t-tests showed that the mean CE values in both sessions of all eight conditions were significantly different from zero ( $p < 0.001$ ). This provides further evidence consistent with  $H_1$ .

To examine the effects of OFB on inflation judgments ( $H_6$  and  $H_7$ ), I carried out four-way mixed ANOVAs on the raw inflation judgments, AE, CE, and VE with Feedback Type (simple OFB versus summarised OFB) and Feedback Amount (no-feedback, 2-feedback, 5-feedback, 10-feedback) as between-participant factors and with Session (one versus two) and Trial (1-10) as within-participant factors.

The analysis of inflation judgments (Figure 5.4, upper panel) showed only a main effect of Feedback Amount ( $F(3, 380) = 7.16, p < 0.001, ges = 0.0014$ ). One-tailed post-hoc paired comparison tests revealed significant differences only between the no-feedback and the 2-feedback condition ( $p = 0.02$ ), the 5-feedback condition ( $p = 0.002$ ), and the 10-feedback condition ( $p = 0.05$ ). A Scheffé test comparing inflation judgments in the no-feedback condition with those in all feedback conditions combined confirmed the beneficial effect of feedback ( $p = 0.024$ ). These findings are consistent with  $H_6$ : feedback decreased inflation judgments but the effect did not depend on how much feedback was provided.

Analysis of AE (Figure 5.4, second panel) showed only a main effect of Feedback Amount ( $F(3, 380) = 3.49, p = 0.02, ges = 0.0008$ ). However, no post-hoc comparisons reached significance. Analysis of CE (Figure 5.4, third panel) revealed a main effect of Feedback

Amount ( $F(3, 380) = 7.16, p < 0.001, ges = 0.0019$ ). One-tailed paired comparison tests showed significant differences only between the no-feedback condition and the 2-feedback condition ( $p = 0.01$ ), the 5-feedback condition ( $p = 0.001$ ), and the 10-feedback condition ( $p = 0.04$ ). A Scheffé test comparing the no-feedback condition with the three feedback conditions combined confirmed the beneficial effect of feedback ( $p = 0.005$ ). As Figure 5.4 shows, CE was, on average, 33% higher in the no-feedback group. These findings are consistent with  $H_6$ . However, there was no evidence that summarized OFB was more beneficial than simple OFB ( $H_7$ ). Finally, there a marginally significant interaction between Feedback Amount, Session, and Trial ( $F(26.46, 3351.60) = 1.51, p = 0.05, ges = 0.0054$ ): while there was a small but constant difference between the no-feedback condition and the feedback conditions in Session 1, that difference became larger over Session 2.

Analysis of VE (Figure 5.4, lower panel) revealed only a main effect of Trial ( $F(7.78, 2954.88) = 2.06, p = 0.04, ges = 0.0023$ ). This arose because VE showed an inverted U-shaped trend across the 10 trials (Table A5.2). Although the overall effect of feedback amount was not significant, I expected that my Experiment 6 finding that simple feedback on every trial produces a higher level of VE than no feedback would be replicated. A Bonferroni post-hoc test showed that this comparison was indeed significant ( $p = 0.03$ ), with VE in the 10-feedback simple OFB condition (0.76) higher than VE in the no-feedback groups (0.71).

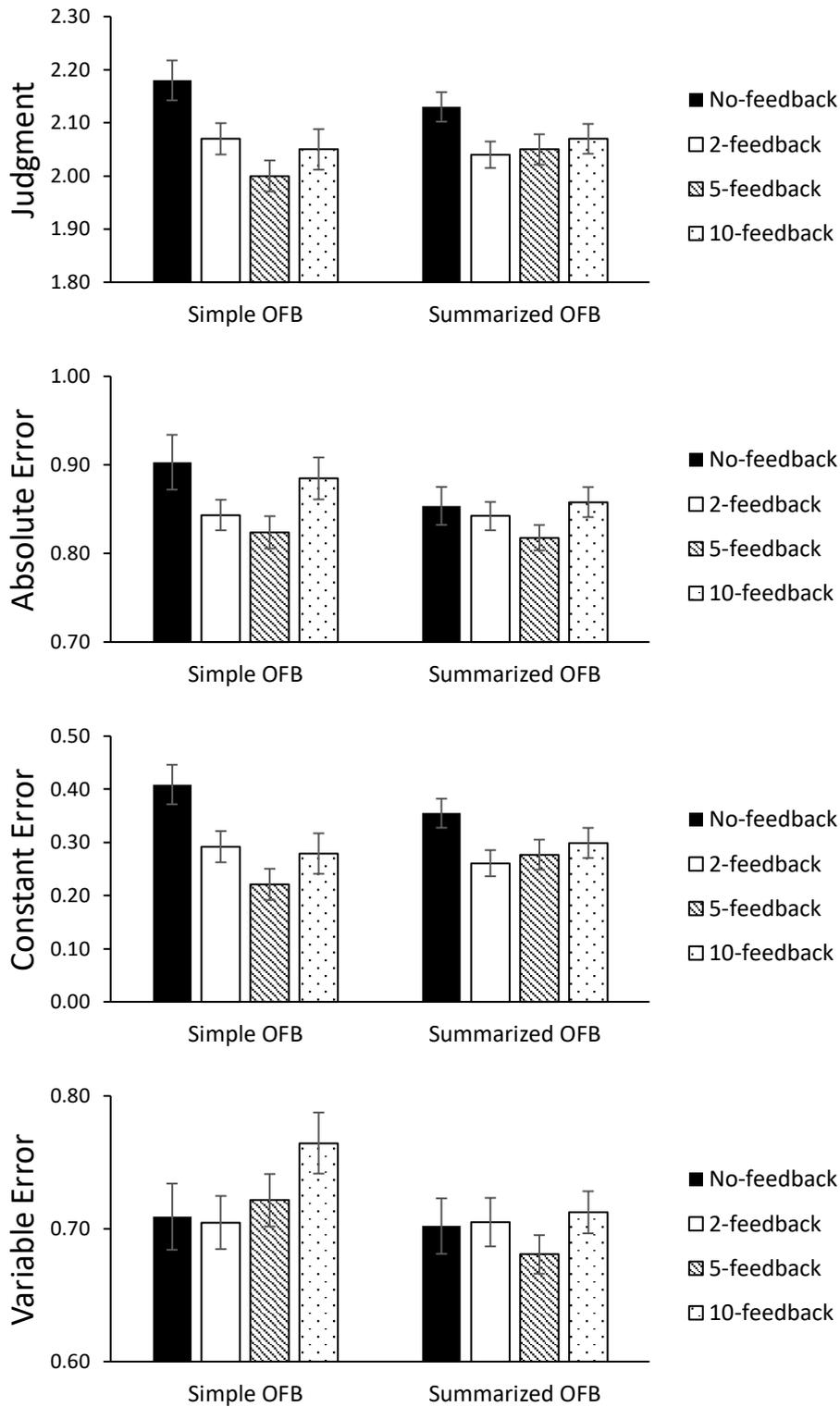
*Confidence in forecasts* Measures of confidence in forecasts (bias, calibration score) were derived in the same way as in the first experiment. Sixteen one-sample t-tests showed that the mean Bias scores in both sessions of all eight conditions were significantly different from zero ( $p < 0.001$ ). This provides further evidence consistent with  $H_3$ .

Analysis of confidence levels (Figure 5.5, upper panel) failed to show a significant main effect of Feedback Amount ( $p = 0.15$ ). However, as I had an *a priori* expectation that feedback would reduce levels of confidence ( $H_4$ ), I used a Scheffé test to compare confidence in the no-feedback condition with that in all feedback conditions combined. This showed that feedback did indeed reduce people's confidence in their inflation judgments ( $p < 0.001$ ). There was also a main effect of Trial ( $F(8.26, 3139.56) = 3.11, p = 0.001, ges = 0.0011$ ). This arose because high initial confidence dropped over the early part of each session (Table A5.3).

Analysis of bias (Figure 5.5, middle panel) showed a marginally significant main effect of Feedback Amount ( $F(3, 380) = 2.21, p = 0.087, ges = 0.0028$ ). However, as I had an *a priori* expectation that feedback would reduce Bias ( $H_4$ ), I used a Scheffé test to compare confidence in the no-feedback condition with that in all feedback conditions combined. This showed that feedback did indeed reduce Bias in judgments of inflation ( $p = 0.007$ ): on average, bias scores in the no-feedback groups were 4% greater than those in the feedback groups. There was also

**Figure 5.4**

*Experiment 7: Mean Judgment (upper panel), Absolute Error (second panel), Constant Error (third panel), and Variable Error (lower panel) in each group, all shown with standard error bars*



a main effect of Feedback Type ( $F(1, 380) = 5.72, p = 0.02, ges = 0.0024$ ). This arose because bias scores were 0.05 greater when participants received summarised OFB than when they received simple OFB.

Analysis of calibration scores (Figure 5.5, lower panel) showed a marginal main effect of Feedback Amount ( $F(3, 380) = 2.21, p = 0.086, ges = 0.0043$ ). Given that I expected feedback to improve calibration ( $H_4$ ), I used a Scheffé test to compare calibration score in the no-feedback groups with that in the feedback groups. This confirmed that calibration was better with feedback ( $p < 0.001$ ): on average, calibration scores in the no-feedback groups were 3% lower than those in the feedback groups.

The ANOVA showed a main effect of Trial ( $F(8.52, 3238.74) = 2.92, p = 0.002, ges = 0.0027$ ): and interactions between Feedback Type, Feedback Amount, and Session ( $F(3, 380) = 3.68, p = 0.01, ges = 0.0014$ ) and between Trial, Feedback Type, Feedback Amount, and Session ( $F(25.46, 3225.06) = 1.72, p = 0.01, ges = 0.0047$ ). To understand these effects in more detail, I carried out separate three-way ANOVAs on summarised OFB and simple OFB groups.

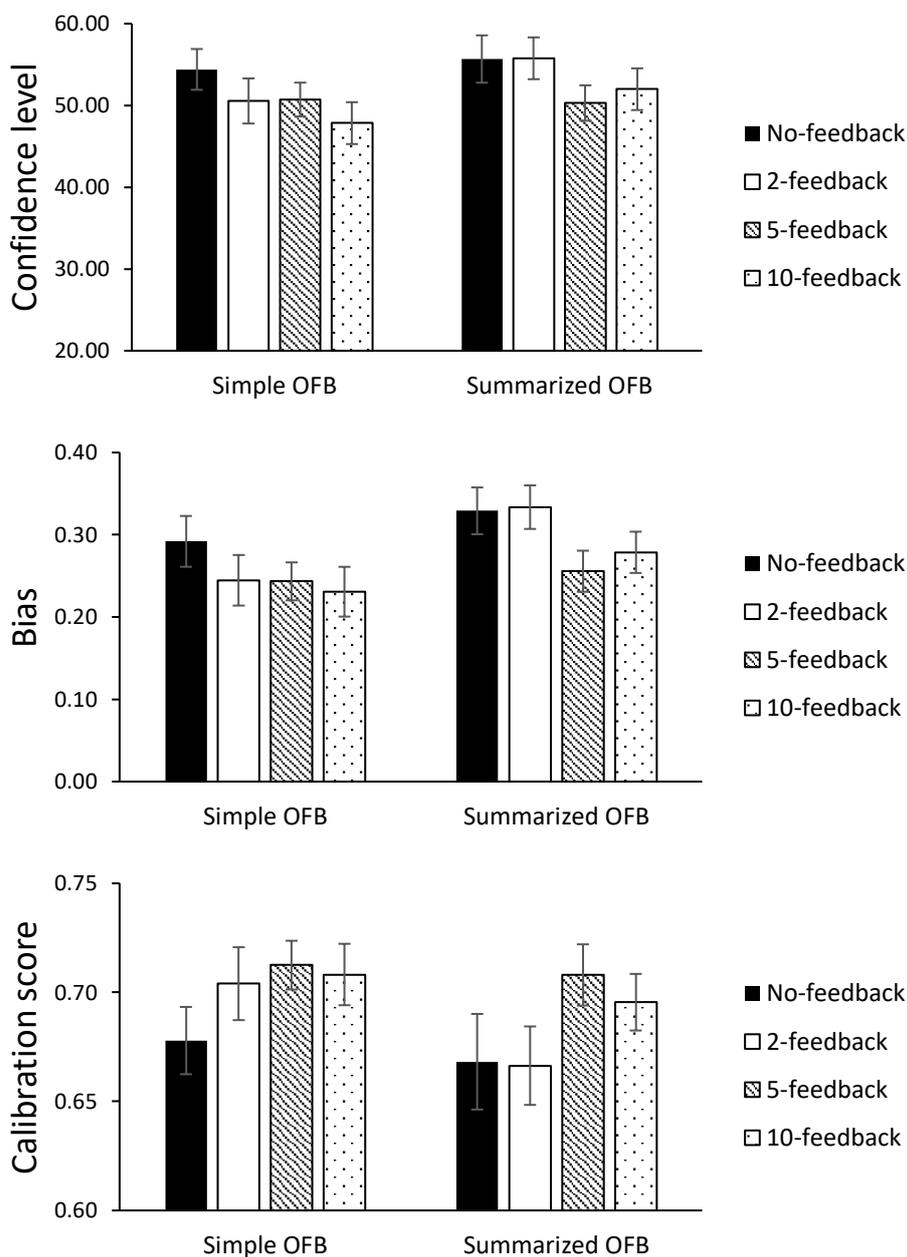
The ANOVA on the summarised OFB groups showed two significant effects. First, there was a main effect of Trial ( $F(8.16, 1632.60) = 2.40, p = 0.01, ges = 0.0040$ ): post-hoc comparisons revealed only that calibration on Trial 3 (0.71) was better than on Trial 1 (0.66) and regression failed to show a significant linear or quadratic trend over the 10 trials within a session. The other significant effect was an interaction between Feedback Amount and Session ( $F(3, 200) = 3.31, p = 0.02, ges = 0.0022$ ). This arose because the simple effect of Session was significant only in the No-feedback group ( $F(1, 46) = 5.88, p = 0.02, ges = 0.0195$ ). Without feedback, mean calibration score dropped between the first session (0.69) and second one (0.65). Thus, as in the first experiment, calibration deteriorated across sessions unless feedback was provided in the first session. People expect their performance to improve with more experience at the task even when it does not: feedback provision disabuses them of this erroneous notion.

The ANOVA on the simple OFB groups revealed only a significant three-way interaction between Feedback Amount, Session, and Trial ( $F(25.08, 1504.98) = 1.66, p = 0.02, ges = 0.0097$ ). This arose because the two-way interaction between Feedback Amount and Session, which was significant across all trials in the summarised OFB groups, was now significant only for trial 4 ( $F(3, 180) = 3.22, p = 0.02, ges = 0.0207$ ) and trial 7 ( $F(3, 180) = 3.44, p = 0.02, ges = 0.0184$ ). Furthermore, these interactions arose for different reasons. In the case of Trial 4, the simple effect of Session was significant only in the 2-feedback condition ( $F(1, 45) = 5.58, p = 0.02, ges = 0.0393$ ) where the calibration score was *higher* in Session 2 (0.75) than in Session 1 (0.66). In the case of Trial 7, the simple effect of Session was significant only in the 5-feedback condition ( $F(1, 46) = 6.01, p = 0.02, ges = 0.0368$ ) where the calibration score was *lower* in Session 2 (0.65) than in Session 1 (0.72). Thus, it appears that the effect of feedback in disabusing people of the notion that their forecasting performance improves with

experience at the task is less robust when simple OFB is given than when summarized OFB is provided.

**Figure 5.5**

*Experiment 7: Mean Confidence level (upper panel), Bias (middle panel), and Calibration score (lower panel) in each group, all shown with standard error bars*

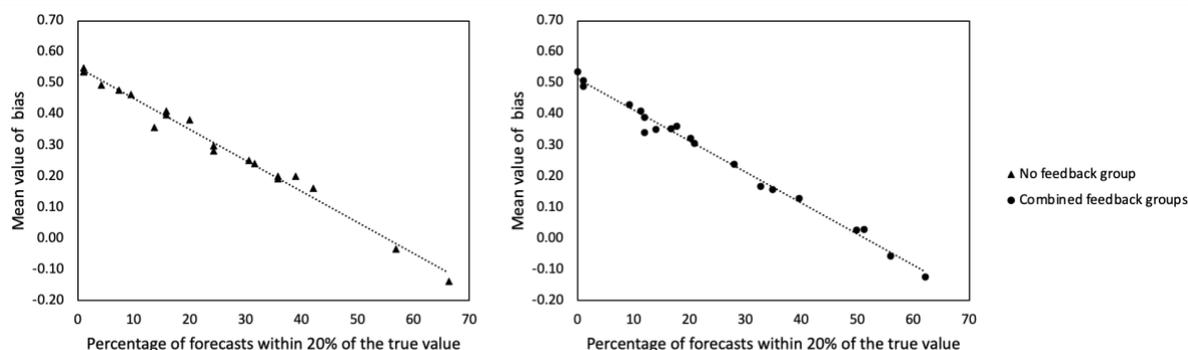


I again examined the evidence for a hard-easy effect ( $H_5$ ). Figure 5.6 shows separate graphs for the hard-easy effect in the no-feedback group and in the three feedback groups combined. For the no-feedback group, the regression revealed that  $\text{Bias} = 0.55 - 0.01 \text{ Difficulty}$  ( $\text{Adj } R^2 = 0.99$ ;  $F(1, 18) = 1274.18$ ,  $p < 0.001$ ). And for the feedback groups, the regression showed that  $\text{Bias} = 0.51 - 0.01 \text{ Difficulty}$  ( $\text{Adj } R^2 = 0.99$ ;  $F(1, 18) = 1807.85$ ,  $p < 0.001$ ). Thus, as in the first

experiment, it is clear that the effect of feedback was to lower the intercept while leaving the slope unchanged.

**Figure 5.6**

*Experiment 7: Bias plotted against forecast difficulty in the no-feedback group (left panel) and the feedback groups (right panel)*



### 5.3.3. Discussion

The results confirmed that inflation judgments are better with feedback provision but they were no worse when feedback was given less often ( $H_6$ ). Specifically, these judgments were lower and, hence, less biased when feedback was given in the first session than when it was not given. However, giving feedback more frequently did not produce any greater reduction in constant error. Furthermore, against my expectations ( $H_7$ ), simple OFB was just as beneficial as summarized OFB. There was no sign that this beneficial effect of feedback was reduced after feedback had been removed in the second session: this implies that its effect was again due to learning rather than incentivisation.

In contrast to the first experiment, provision of OFB had no *overall* effect on variable error. However, the comparison of the no-feedback group with the 10-feedback group revealed a significant effect in the simple OFB conditions but not in the summarized OFB conditions. Thus I replicated my finding from the first experiment but the effect that I found did not generalise to the summarized OFB conditions. Why was this? In the first experiment, feedback merely specified the actual inflation rate along with the participant's judged inflation rate. In the current experiment, this was true only in the simple OFB conditions. In the summarized OFB conditions, participants were also explicitly given the difference between the actual and judged inflation rate (e.g., "you overestimated by 0.52% on average"). If this made feedback easier to process and thereby reduced judgment noise for participants in the summarized feedback conditions, VE would have been lowered such that it became close to that observed in the no-feedback group.

Absolute error reflects both constant and variable error. In my first experiment, feedback had beneficial effects on the former but detrimental effects on the latter; the effects of OFB on

these two types of error cancelled each other out and, as a result, feedback had no overall effect on AE. In the current experiment, there was a main effect of the amount of feedback but post-hoc comparisons failed to reach significance. However, the opposite effects of providing feedback on every trial on CE and VE in the simple OFB conditions were still statistically significant and their contributions to AE cancelled each other out just as they did in the first experiment (Figure 5.4). The overall effect of feedback amount may reflect lower AE values in the 2-feedback and 5-feedback conditions than in the no-feedback and 10-feedback conditions. However, I have no statistical evidence to support this interpretation.

I expected feedback to lower confidence, reduce bias, and improve calibration ( $H_4$ ) and I found that it did have these effects. However, as with inflation judgments, providing more feedback did not have a more beneficial effect. In addition, overconfidence was greater when participants received summarized OFB than when they received simple OFB. I assume that this occurred because people given summarised OFB erroneously believed that such feedback would benefit their performance more than people given simple OFB expected that that type of feedback would benefit their performance. This implies that people have an imperfect internal model of the factors that affect their performance (c.f., Harvey & Fischer, 2005).

There were some main and interactive effects of Trial. These were generally small and, though not specifically anticipated, most of them make sense in the context of the other results. The difference in CE between feedback and no-feedback groups increased over the experiment: provision of feedback reduced CE and kept it relatively low whereas, in the absence of feedback, CE drifted higher over time. With experience at the task, confidence dropped after the first trial. Analysis of calibration scores showed high level interactions which, while not easy to interpret, appear to reflect the changeable quality of both inflation judgments and confidence judgments in the absence of feedback but their relative stability in its presence. Finally, in all groups, VE scores were relatively low at the beginning and end of each session but somewhat higher middle. People may start sessions by making relatively cautious judgments close to the mean of the presented series, become more adventurous as they suspect that some series must show trends, and then revert to caution when experience of more stimulus series convinces them that none do.

#### **5.4. General discussion**

I first consider inflation forecasts and the effect that OFB had on them. I then discuss confidence in those forecasts and the effect that OFB had on it. Next, I assess implications of my findings for consumers and for the way in which surveys of inflation expectations are used by central banks and other agencies. I then outline some limitations of my work and, finally, provide a brief concluding section.

##### *5.4.1. Inflation forecasts*

In both experiments, inflation forecasts were too high. This is consistent with previous research on lay people's expectations of inflation and may arise because they find it easier to

bring to mind past examples of high price rises and use these as a basis for their forecasts (Bruine de Bruin, Van der Klaauw, et al., 2011). However, in my experiments, forecasts were not memory-based; there was no need to recall past price rises. Instead, people were given past values of inflation from previous years to base their forecasts on. (These are typically available for respondents in experts' surveys, such as the SPF, but not for respondents in lay surveys, such as the MSC.) Furthermore, in my experiments, people made forecasts for countries other than their own: their recall of particular price rises in their own countries should not have been relevant (though it may still have had an influence). Another explanation for over-forecasting of inflation, at least in experts, has been that forecasters may be penalised less for over-forecasting than for under-forecasting (e.g., Capistrán & Timmermann, 2009). However, in my experiments, performance-related incentives were not asymmetrical. Thus it is more likely that, people expect inflation to be more likely to rise than fall (Niu & Harvey, 2022a, Chapter 4). This may be a risk-amplification effect of using the availability heuristic: media coverage of potential and actual rises in inflation is more extensive than corresponding coverage of falls in inflation.

Outcome feedback reduced over-forecasting. As far as I am aware, this is the first time that such an effect has been demonstrated for time series forecasting. In past studies, successive trials have required people to make forecasts from the same time series and so provision of outcome feedback has been unavoidable (in order to provide the latest point for the next forecast). As a result, previous studies were only able to use an OFB condition as a baseline against which to assess the effectiveness of more sophisticated types of feedback (e.g., Goodwin & Fildes, 1999; Remus et al., 1996). In contrast, by presenting forecasters with a different time series on each trial, I was able to compare the OFB condition with a no-feedback condition. Furthermore, the lack of any significant reduction in the effect of feedback after it had been removed in the second session is not consistent with the feedback having its effect by providing an incentive for better performance; instead, it is best explained by a learning effect (Annett, 1969).

However, this learning does not appear to have been a slow incremental process of the sort that is characteristic of much skill acquisition. In the first experiment, there was no evidence that the advantage of the feedback group over the no-feedback group increased over trials. In the second experiment, groups receiving feedback produced less biased inflation judgments than those not receiving feedback but there was no evidence that groups receiving more feedback were less biased than those receiving less feedback. This suggests that provision of just some information was sufficient for bias to be reduced. It appears that OFB transmitted knowledge to forecasters and that this knowledge needed to be transmitted just once or twice to be effective.

Recently, there has been greater appreciation that the overall error in judgments depends not just on biases but also on the scatter or noise that they contain (Kahneman et al., 2021). Previous studies of judgmental forecasting have shown that different factors influence these

two error components and that the same factor can affect them in different ways (e.g., Harvey & Bolger, 1996). The present study provides further evidence of this. In the first experiment, OFB reduced bias (CE) but increased scatter (VE); as a result, it had no effect on overall error (AE). This effect was replicated in the simple OFB conditions of the second experiment: VE was again higher and CE was lower when feedback was provided on every trial than when no feedback was provided and, consequently, there was no resultant effect of feedback on AE. In the summarized OFB conditions, feedback did not increase VE: I attributed this to the fact feedback given in these conditions included an explicit comparison of the participant's judgment and the outcome, thereby reducing one contributor to judgment noise.

#### *5.4.2. Confidence in forecasts*

On average, people were 28% overconfident in their inflation forecasts in the first experiment and 28% overconfident in them in the second experiment. This degree of overconfidence in forecasts is higher than that obtained by Wright (1982) and Ronis and Yates (1987). It is unlikely that this difference arose because my participants were asked to assess the likelihood that an outcome would appear within an interval whereas theirs were asked to estimate the likelihood that one of two possible answers was correct. This is because Hansson, Juslin, et al. (2008) also found low overconfidence, comparable to that reported by Ronis and Yates (1987), when they asked people to estimate the probability that an outcome would fall within a specified interval. It is more likely that the difference relates to the fact that Wright (1982), Ronis and Yates (1987) and Hansson, Juslin, et al. (2008) all studied memory-based judgments whereas I examined time-series forecasting.

Outcome feedback reduced overconfidence by 9% in the first experiment and by 5% in the second one. As my task involved a sequence of judgments to related items, these findings are consistent with the conclusions of my earlier review: OFB reduces biases in confidence judgments when those judgments are made to related items (Benson & Önköl, 1992; Subbotin, 1996; Winman & Juslin, 1993) but not when they are made to unrelated items (Keren, 1988; Subbotin, 1996). I attributed the only exception to this generalisation, as indeed did its author (Zakay, 1992), to an incentive effect of OFB that increased people's attention to critical aspects of the task. However, the beneficial influence of OFB on confidence judgments in my task is not attributable to such an effect: this is because there was no evidence of a significant reduction in that influence after feedback was withdrawn in the second session. Instead the benefits of OFB must have arisen from some type of learning effect.

What was the nature of this learning? Figures 5.3 and 5.6 indicate that feedback had no effect whatsoever on the hard-easy effect: the intercept of the regression of bias on to task difficulty was reduced by feedback but its slope remained the same. A finding such as this can be interpreted within Ferrell's (1994) model of calibration. According to his approach, the skill of good calibration depends on two separate abilities that combine to produce accurate probability judgments: base-rate identification and discriminability. The intercepts of the regressions shown in Figures 5.3 and 5.6 depend on an ability to identify the base-rate of the

occurrence of an event: in my task, the ability to determine the average likelihood of a forecast being within 20% of the correct value. The slopes of the regressions depend on an ability to discriminate between more likely and less likely events: in my task, the ability to discriminate between series for which producing a forecast is so difficult that it is unlikely that it will be within 20% of the outcome and series for which producing a forecast is much easier so that it is much more likely that it will be within 20% of the outcome. Without this second ability, people will, on average, judge all series close to the identified base rate and, as a result, they will be more overconfident when judging more difficult series. My findings show that OFB improved base-rate identification but had no effect on discriminability.

Provision of OFB significantly improved calibration but, even with this feedback, the mean calibration score was only 0.70. This is not especially impressive: as we have seen, a uniform judge who always considers that the forecast is as likely as not to be within 20% of the outcome would obtain a calibration score of 0.75. This poor calibration after provision of OFB is not especially surprising given that, as we have just seen, feedback improved people's ability to judge the mean difficulty of forecasting but did not improve their ability to discriminate between series that were harder to forecast and those that were easier to forecast.

#### 5.4.3. Implications

Lay people tend to expect inflation to be higher than it turns out to be. This is assumed to have a number of effects on their economic behaviour. For example, they are likely to bring forward their purchasing of durable goods; this will, in turn, produce higher demand for those goods and increase their prices. In other words, expectations of higher inflation increase inflation. This is why central banks use surveys of lay expectations of inflation to help them forecast inflation. It follows that more realistic inflation expectations should decrease inflation. I have shown that training with OFB can produce more realistic expectations. However, one would not want to provide such training only to survey respondents as their more realistic expectations would not then be representative of the population from whom they were drawn.

Central banks carry out surveys of lay and professional forecasters for different reasons. Whereas lay expectations *feed into* the forecasting process because those expectations are assumed to influence economic behaviour and hence inflation, professional expectations represent one way of *providing* forecasts. Thus, there are no downsides to improving inflation forecasts made by those responding to professional surveys; there are no concerns about altering the degree to which professional forecasters responding to surveys are representative of the population of professional forecasters as a whole. Training professional survey respondents to produce better forecasts could reduce their tendency to over-forecast inflation (Ang et al., 2007). It would be relatively easy to implement as professional surveys (unlike lay surveys) already use the approach that I adopted here and provide respondents with past values of inflation when asking them to forecast future ones.

Although survey respondents have been asked to assess aleatory uncertainty associated with future inflation (e.g., assess the likelihood that inflation will exceed 4% next year), they have not been asked to assess epistemic uncertainty (e.g., assess the likelihood that your estimate of next year's inflation is within 20% of what it turns out to be). However, there is an argument that central banks should put more weight on survey expectations that respondents produce with greater certainty. If those who expect inflation to be high are not as certain of their expectations as those who expect it to be low, more emphasis should be given to the latter view when data are aggregated. This approach would work better when respondents are better calibrated. Training with OFB could be helpful here.

#### *5.4.4. Limitations and suggestions for future work*

These experiments demonstrated the effectiveness of training with OFB. I showed that such training remains effective even after feedback has been withdrawn. However, a question remains about how long it would remain effective. If the interval between the first and second sessions of my experiments had been extended to hours or days, would the advantage of training still persist?

Training using guidance has been found to be more effective than training with OFB when people have little experience in performing a task and when tasks are complex (Holding & Macrae, 1964; Macrae & Holding, 1965). Also, within the MCPL tradition, guidance (task information) has been found to be more effective than feedback (Balzer et al., 1989). Guidance provides advance information about a task that can be incorporated into the knowledge used to perform it. My finding that people benefit from provision of OFB but are insensitive to the amount provided suggests that this feedback provides them with useful knowledge about their task; once they have obtained this knowledge, they receive no additional advantage from being given it again. In other words, just like guidance, OFB provides task information but, unlike guidance, it is provided after rather than before performance. This line of thinking suggests that it would be worth investigating whether guidance is as effective for improving inflation forecasting as training with OFB. It would be relatively simple to implement: even providing a warning that the average survey respondent overestimates inflation by some percentage could be effective.

I presented my participants with inflation rates for the years 2009-2018 and required them to forecast the inflation rate for 2019. In fact, 14 out of the 20 countries showed a fall in CPI inflation from 2018 to 2019; inflation rates in one country are not independent of those in other countries. Although participants were not informed of the base rates of different types of inflation change, those receiving OFB may, over the first experimental session, have slowly come to realize (correctly) that inflation was more likely to fall than rise. However, this would imply that inflation judgments would be lower at the end of the first session in the feedback groups but not in the no-feedback groups. In fact, as I have seen, inflation judgments showed a U-shaped function over the 10 trials in each session in both feedback and no-feedback groups in Experiment 6 (Table A5.1) and no trend over the 10 trials in any group in Experiment

7 (Table A5.2). These data are not consistent with cross-series dependencies informing participants' judgments. (To confirm this, simulated series could be used to manipulate the base rate systematically to determine whether it has any effect on participants' judgments.)

I have shown that OFB decreases over-forecasting of inflation. There may be other manipulations that have a similar effect. Inflation rates are typically small numbers ranging between -1.00 and 2.00. Laypeople, especially those with low numeracy, may not find it easy to process such numbers. Indeed, less educated people have been found to have higher expectations of future inflation rates, possibly because they round up their expectations in order to express them as whole numbers (Bryan & Venkatu, 2001a). This suggests that recoding inflation by, say, multiplying inflation rates by 100, might reduce over-forecasting by such people (c.f., Sedlmeier, 2000). However, such recoding would result in values that people would find difficult to relate to their everyday experience of price changes. An increase of one percent is easily interpretable as an extra cent in every dollar whereas an increase of 100 units in every 10,000 units is not so easy to process.

#### *5.4.5. Conclusions*

In the past, many approaches to improving decision making have been examined in a variety of domains have been found to facilitate performance without bringing it very close to optimal. Here I have shown for the first time that one of the simplest of these approaches, provision of OFB, is effective in reducing biases in time series forecasting and in the confidence that people have in their forecasts. However, there was little effect on overall error in forecasts because OFB increased the noise or scatter in forecasts. In practice, this should not detract from the advantages of using OFB because, unlike bias, noise or scatter can be reduced by averaging over a number of different judges (Surowiecki, 2004) or over a number of judgments from the same individual (Herzog & Hertwig, 2014).

## Chapter 6 How do elicitation methods improve measurement of inflation expectations

### 6.1. Introduction

There are three main ways in which people use judgment to make predictions about the future values of a variable. In point forecasting, they make a single point estimate of its expected value. In interval (or range) forecasting, they provide a range of values within which they judge there is some probability (e.g., 90%) of the outcome occurring. The mid-point of the bounds of the interval is taken to correspond to their expected value of the variable; it should be the same as the point forecast<sup>21</sup>. In probability density forecasting, they provide a probability that the outcome will be in each of a number of different ranges. The mean or median of the distribution of these probabilities should correspond to the expected value of the outcome (i.e., equal to the point forecast).

These types of judgmental forecast vary in two ways: they differ in how simple they are to elicit and in terms of how much information they provide to users. Unlike point forecasts, interval and density forecasts provide users with information about forecasters' uncertainty in their forecasts: in many applications, this is important for planning purposes. Furthermore, density forecasts provide more detail about this uncertainty than interval forecasts. In many domains, this additional information is useful for guiding future decisions. For example, those assessing financial risks may need to know more than that there is a difference in the width of the interval forecasts for the future returns on two investments; they may wish to know whether it reflects a difference in variance or kurtosis of the distribution of those returns.

There is a trade-off between the simplicity of elicitation of forecasts and the detail that those forecasts provide. In some domains, time pressure arising from the number of forecasts required within a short period of time limits forecasters to the provision of point forecasts: for example, demand forecasters may need to make forecasts for many stock-keeping units (SKUs) within a short period. In other areas, such as meteorology, agriculture, and the nuclear industry, users need density forecasts but time pressure is less of a consideration. Reviewers have tended to focus on either point and interval forecasting (e.g., Lawrence et al., 2006) or on density forecasting (e.g., O'Hagan et al., 2006), partly because the salient issues for research and practice depend on the type of forecast under consideration. Perhaps for this reason, there has been little concern with comparing the accuracy of the different types of forecast. It is this neglected issue that I focus on here.

---

<sup>21</sup> This assumes that interval bounds are placed symmetrically around where the point forecast would be placed. Any systematic difference between the mid-point of the interval and the position of the point forecast would produce a difference in the mean error (bias) for the two types of forecast. No such difference in bias was obtained in my experiments.

Estimates of expected value can be extracted from all three types of forecast: I should expect point forecasts, the mid-point of the bounds describing interval forecasts, and the mean or median of density forecasts to be the same when they are all based on the same data. Here I ask people to make a number of forecasts from the same set of data series. For each person, I measure the mean central (expected) values that they produce for each type of forecast (point, interval, density) to determine whether there is any difference between them. As I also have the true outcomes corresponding to each forecast, I also examine whether the overall accuracy, measured by the Root Mean Squared Error (RMSE), varies across the different types of forecast. This overall measure of error can be decomposed into Mean Error (ME), also known as bias, directional error, or constant error, and Variable Error (VE), also known as noise, or inconsistency. Hence, I also examine whether these components of overall error vary across different types of forecast.

My investigation is framed in terms of inflation forecasting. There were three reasons for this. First, this is one of the few domains in which all three types of forecast (point, interval, and density) are used in practice: surveys of both experts (economists and professional forecasters) and lay people (consumers and households) require them. Second, and related to the first point, inflation forecasting by lay people provides, to the best of our knowledge, the only previous studies that compare different types of forecasting (Bruine de Bruin, Manski, et al., 2011). Third, data series for inflation are available and are regularly updated. Thus, it is possible to provide people with inflation data series for various countries, ask them to make inflation forecasts for those countries, and then compare their forecasts with true outcomes when those are available.

Though my studies are framed within the inflation forecasting domain<sup>22</sup>, I anticipate that my conclusions about differences between different types of forecasting will generalize across all content areas. Indeed, calibration of inflation forecasts and of calibration of other types of forecast (e.g., Benson & Önköl, 1992) are affected in a similar way by specific factors such as feedback (Niu & Harvey, 2022b, Chapter 5).

## 6.2. Experiment 8

Bruine de Bruin, Manski, et al. (2011) found that, on average, point forecasts are the same as the mean (and median) value of density forecasts. They obtained this result using a within-participant design: people first made an interval or point forecast, then, if they had made an interval forecast, they made a point forecast, and, finally, they made a density forecast. Here, I seek to replicate this finding using a between-participants design: separate groups of participants made point forecasts, interval forecasts, and density forecasts.

The reason that I made this change is that context effects are known to influence responses in both traditional and online surveys (e.g., Reips, 2002; Smyth et al., 2009; Tourangeau et al.,

---

<sup>22</sup> The studies do not aim to simulate inflation surveys. What I ask participants to do is not the same as what respondents are asked to do in those surveys.

2000), including surveys of inflation expectations (Niu & Harvey, 2022a, Chapter 4). My concern here is that people's responses to survey questions eliciting density forecasts may be influenced by their earlier responses to survey questions eliciting point (or interval) forecasts. In particular, point forecasts may act as mental anchors for estimates of the means of density forecasts: because of under-adjustment (Tversky & Kahneman, 1974), these two values would then be more similar than they would otherwise be. Once the possibility of anchoring is eliminated by the use of a between-participants design, differences between point forecasts, the mid-point of interval forecasts, and the mean value of density forecasts may appear.

Although Bruine de Bruin, Manski, et al. (2011) did not report whether the mid-point of the range (i.e., interval) forecast matched the point forecast and the mean of the density forecast, it is reasonable to assume that it would do so given that they found that the latter two values were the same. Thus the first hypothesis that I test is the following one.

*H<sub>1</sub>: Point forecast = Mid-point of the interval forecast = Mean of the density forecast*

This is a null hypothesis. I use a well-powered experiment to examine whether I can obtain evidence inconsistent with it in a between-participants design that excludes the possibility of anchoring effects.

To test my other hypotheses, I extracted from the data the three error measures that I mentioned above. Given that  $D$  is the judged rate of inflation for a particular country minus the actual rate of inflation for that country and given that each participant makes inflation judgments for  $n$  countries, each participant's Mean Error (ME) is given by  $\Sigma D/n$ . Their Variable Error (VE) is given by  $\sqrt{\Sigma (D - ME)^2/n}$ . Their Root Mean Squared Error (RMSE) is given by  $\sqrt{\Sigma (D^2)/n}$ . Equivalently, RMSE can be expressed via its decomposition into ME and VE as  $\sqrt{((ME)^2 + (VE)^2)}$ .<sup>23</sup>

Many previous studies (e.g., Bruine de Bruin, Van der Klaauw, et al., 2011; Bryan & Venkatu, 2001a, b; Georganas et al., 2014) have shown that people tend to over-forecast inflation. It appears that people have a general expectation that inflation will be higher than it turns out to be. Thus I test whether the Mean Error (ME) is positive.

*H<sub>2</sub>: ME > 0*

If  $H_1$  is true, it also means that the three different types of forecast will be equally biased. Thus I also seek to obtain evidence against the following hypothesis.

*H<sub>3</sub>: ME<sub>Point</sub> = ME<sub>Interval</sub> = ME<sub>Density</sub>*

---

<sup>23</sup> Mean Absolute Error (MAE) provides an alternative measure of overall error. I did not use it here because, unlike RMSE, it does not decompose neatly into ME and VE. However, in almost all cases reported here, analysis of it leads to the same conclusions as those arising from the analysis of RMSE.

People's judgments are noisy (Kahneman et al., 2021): they are subject not only to bias but to moment-to-moment random variation. It is not surprising, therefore, that taking the average of a number of judgments from a single person produces a more accurate estimate than using a single judgment (Herzog & Hertwig, 2009, 2014; Vul & Pashler, 2008). When someone makes a point forecast, they make a single judgment ( $f$ ). When they make an interval forecast, they make two judgments ( $f + \delta f$ ;  $f - \delta f$ ). Thus they estimate  $f$  twice; I therefore expect the average of the bounds of the interval used to express the range forecast to be more accurate than the point forecast. An analogous argument leads me to expect the central tendency of a density forecast to be more accurate than the mean of the bounds of a range forecast. To examine the validity of these arguments, I test the following hypothesis.

$$H_4: VE_{Point} > VE_{Interval} > VE_{Density}$$

If  $H_3$  and  $H_4$  are true, then I should also expect:

$$H_5: RMSE_{Point} > RMSE_{Interval} > RMSE_{Density}$$

### 6.2.1. Method

Separate groups of participants made point forecasts, interval forecasts, and density forecasts.

*Participants* One hundred and thirty-nine participants (75 males, 64 females) with a mean age of 22 years ( $SD = 5$  years) were recruited for the online study. They were divided into three groups: a point forecasting group ( $N = 56$ ); an interval forecasting group ( $N = 42$ ); a density forecasting group ( $N = 41$ ). Forty of these participants were recruited from the participant pool at University College London (UCL) and given 0.25 credits for their participation. The remaining 99 participants were recruited in UCL or China; the former received £3.00 and the latter received 3RMB for taking part. Data were collected between 1 July 2019 and 30 September 2019.

*Stimulus materials* Participants in each group were shown 10 graphs of real inflation rate data from 10 different countries. Each one displayed a time series representing 20 years of annual historical inflation data from 1998 to 2017.<sup>24</sup> The last displayed data point was for the period immediately before the one to be forecast. The identities of the 10 countries were not specified; instead, they were labelled with numbers. Seven of the series showed no trends; three contained shallow trends.

The way that data are graphed can affect the forecasts that people make (Lawrence & O'Connor, 1992). For example, people are less likely to follow an upward trend in the data when the last data point is already close to the top of the vertical axis. To avoid such problems, the inflation data series were displayed in the central part of the y-axis scale, which ranged

---

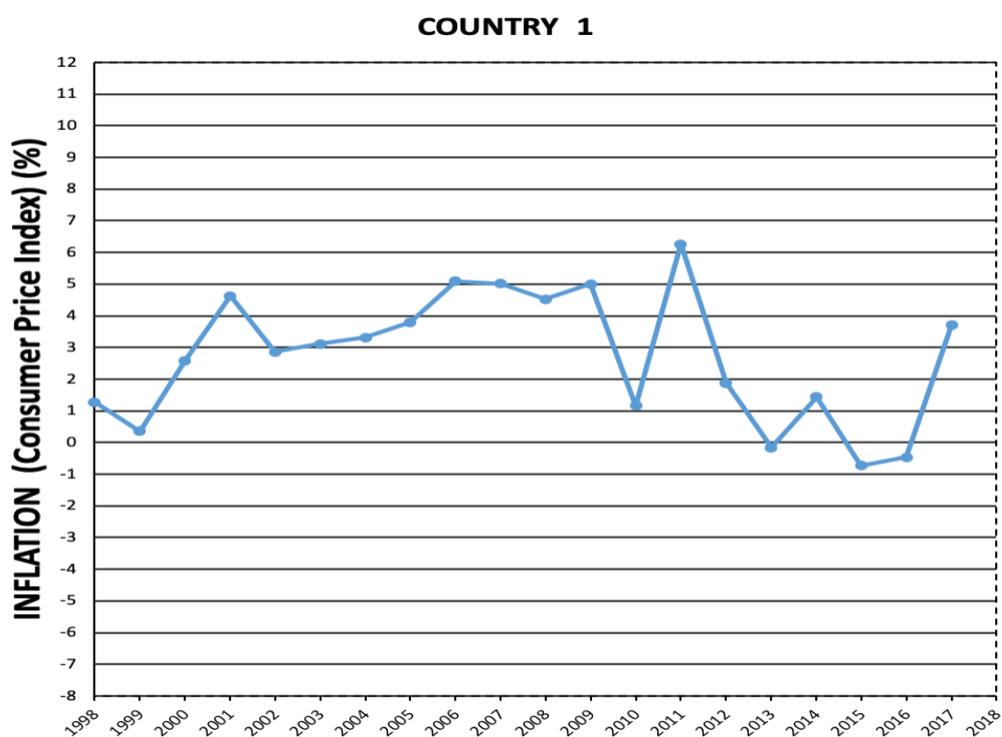
<sup>24</sup> Some but not all surveys of inflation expectations provide respondents with information about inflation in preceding periods.

from 12% to -8% and the final points for all 10 countries were in the middle of that scale, ranging from -0.31% to 7.55%. Series were broadly comparable with a mean inflation level at 2.61% ( $SD = 2.90\%$ ) across the 10 series. A typical series is shown in Figure 6.1.

Two versions of the experiment were programmed, one in English for English speakers and one in Chinese Mandarin for Chinese speakers. To ensure these were comparable, the English version was initially translated into Chinese and then back-translated into English. The back-translation was then compared to the original version to ensure that they matched.

**Figure 6.1**

*Experiment 8: Time series showing 20 years of annual inflation figures from Country 1*



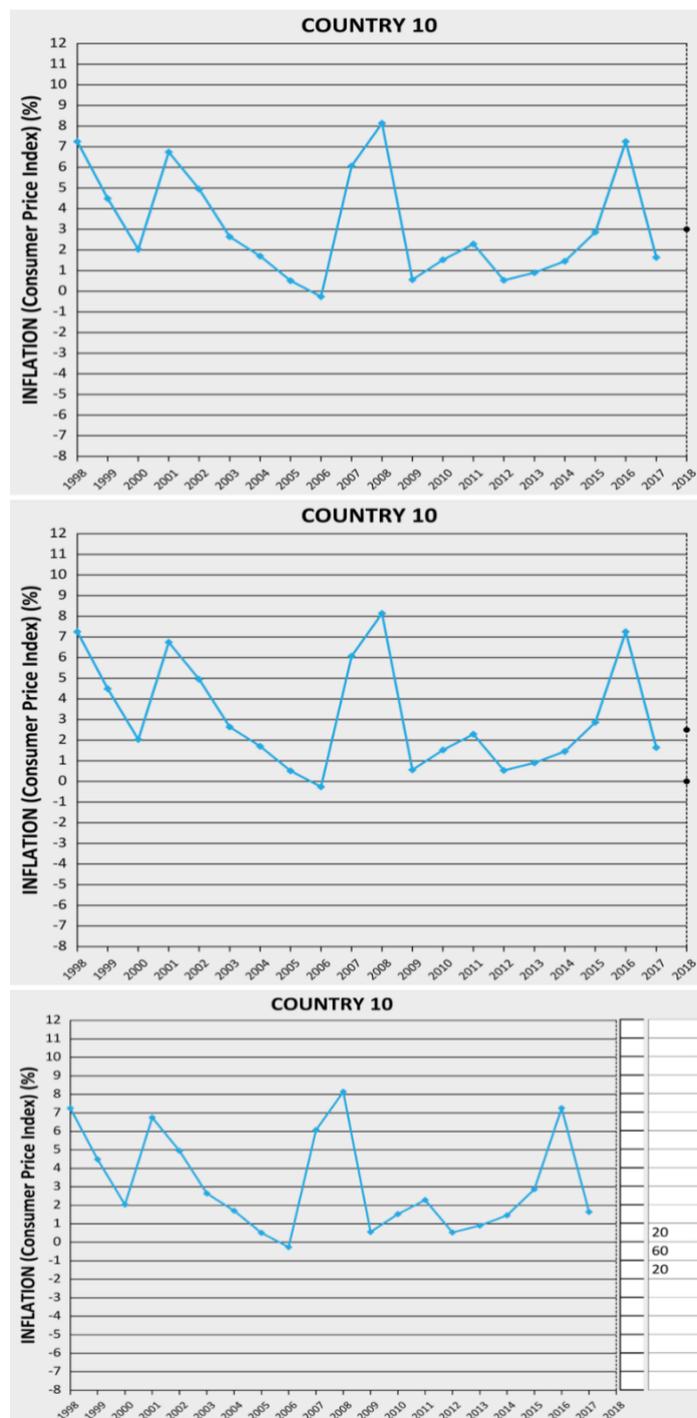
*Design* The experiment used a between-participants design. Participants from each language group were randomly assigned to one of three groups: point forecasting, interval forecasting, density forecasting.

*Procedure* Participants first saw an information screen that outlined the nature of the study and a consent screen that detailed the ethical permission that had been provided and that elicited their consent for participating. They were then asked basic demographical questions that required them to specify their age, gender, level of education, main academic discipline that they had studied, the country that they had lived in for most of their life, and any economics-related work experience. A brief explanation of the nature of inflation was then provided. After that, participants in each group completed their 10 forecasts. The 10 countries

for which inflation had to be forecast were presented in a different random order for each participant<sup>25</sup>.

**Figure 6.2**

*Experiment 8: Examples of point forecasting (upper panel), interval forecasting (middle panel), and density forecasting (lower panel)*



<sup>25</sup> At the end of the experiment, participants also answered three open-ended questions about how they made their forecasts. I do not report details of their responses here.

Instructions given to those in the point forecasting group were: “Below is a series of inflation rates for one country. WHAT WILL HAPPEN NEXT? Please estimate the actual value of the inflation the next year by clicking once on the punctuated line”. Instructions for those in the interval forecasting group were: “Below is a series of inflation rates for one country. WHAT WILL HAPPEN NEXT? Please make your 90% prediction interval. (90% prediction intervals correspond to the interval in which future observations will fall, with a 90% probability.) Click twice on the punctuated line at the end of the graph to show the upper and lower boundary of this 90% interval”. Finally, those in the density forecasting group were instructed as follows: “Below is a series of inflation rates for one country. WHAT WILL HAPPEN NEXT? Please allocate £100 into the 20 bins appearing on the screen. Money allocation should be higher in the bins where you believe there is a greater probability for the actual inflation next year. To allocate all £100, please enter your bets to each of the bins at the end of the graph”. These instructions appeared above each of the 10 graphs for which participants had to make forecasts. Examples of the screens in the three conditions are shown in Figure 6.2.

### 6.2.2. Results

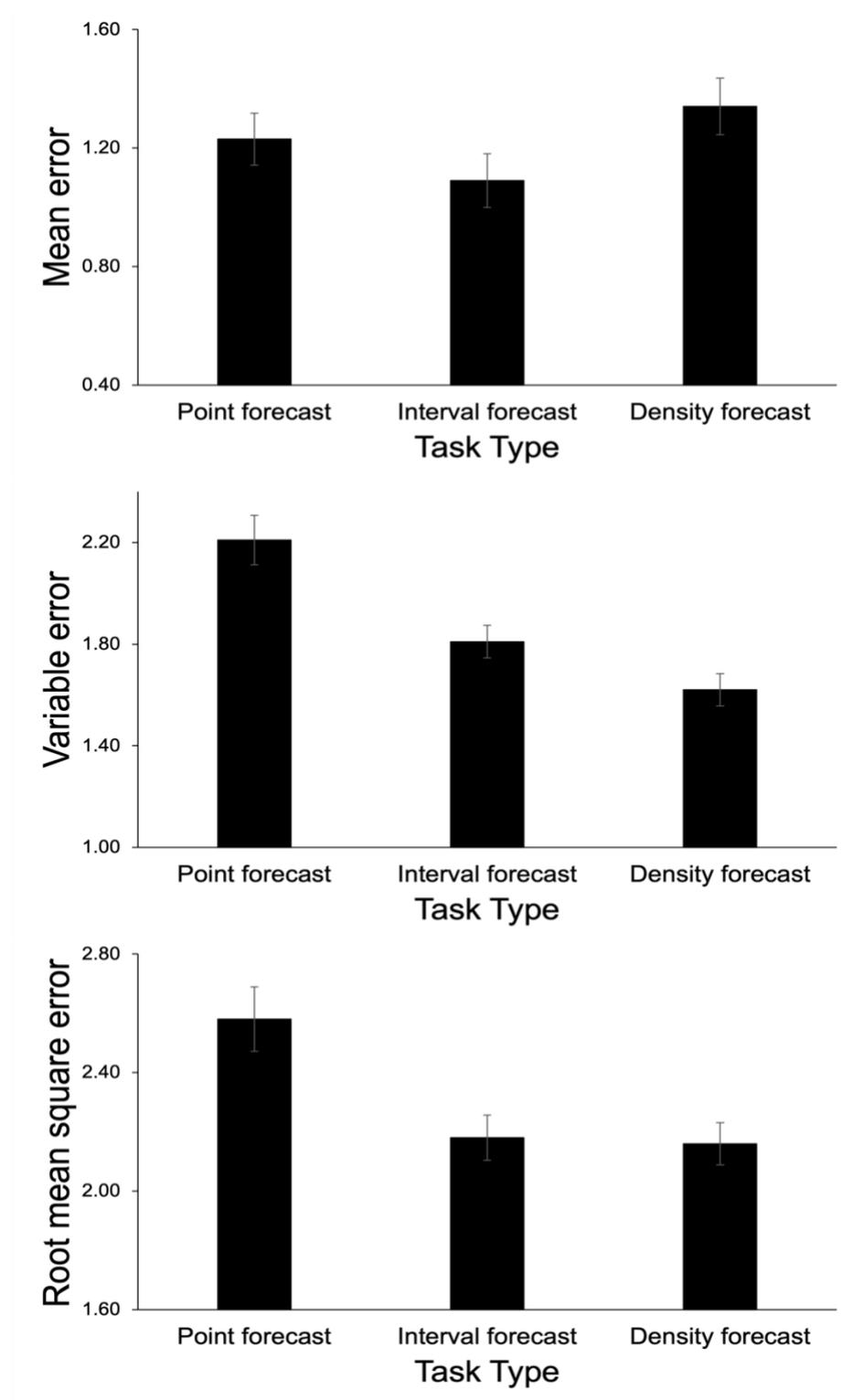
The central tendency of forecasts that each person made in the three conditions was first extracted. For those in the point forecasting group, this was simply the point forecasts that they made. For those in the interval forecasting group, it was the mid-point of the interval bounds that they provided. For those in the density forecasting group, I used the reported bin probabilities to fit an underlying parametric density and then extracted the underlying forecast density mean from this. This approach, developed by Engelberg et al. (2009), assumes that probabilistic beliefs are unimodal and that a participant’s distribution can be specified as a member of the generalised Beta family. However, when a forecaster fills in values for only two of the 20 bins, it is only possible to specify the mean of distribution when the two bins are adjacent. In my experiment, 10 participants failed to do this. As a result, my sample for the density forecasting group was reduced.

*Once I had extracted the mean value of forecasts on each trial for each participant in each condition, I carried out a two-way mixed analyses of variance (ANOVA) on these values using forecast type (point forecasting, interval forecasting, density forecasting) as a between participant factor and trial number (1-10) as a within-participant factor. This analysis showed no significant main effects or interactions. Thus, I obtained no evidence inconsistent with H<sub>1</sub>: even with a between-participants design that excluded the possibility of anchoring effects, there was no suggestion that different forecasting methods produced different estimates of the central value of inflation. This replicates and reinforces Bruine de Bruin, Manski, et al.’s (2011) conclusions.*

It is clear from upper panel of Figure 6.3 that people overestimated inflation: a one sample t-test showed that ME was significantly positive ( $t(128) = 22.67, p < 0.001$ ). This finding is consistent with H<sub>2</sub>. A one-way ANOVA on ME using forecast type as a between-participant factor revealed no significant main or interactive effects. Thus I failed to obtain evidence inconsistent with H<sub>3</sub>.

**Figure 6.3**

Experiment 8: Bar chart showing ME, VE, and RMSE error scores (with standard error bars) in each forecasting task



A one-way ANOVA on VE revealed an effect of forecast type ( $F(2, 126) = 12.43, p < 0.001, ges = 0.1648$ )<sup>26</sup>. Post-hoc analyses revealed significant differences between the point and density forecast ( $p < 0.001$ ) and between the point forecast and the interval forecast ( $p = 0.002$ ) but no significant difference between the interval forecast and the density forecast (Figure 6.3, Middle panel).

Finally, a one-way ANOVA on RMSE revealed an effect of forecast type ( $F(2, 126) = 6.56, p = 0.002, ges = 0.0943$ ). Post-hoc analyses revealed significant differences between the point and density forecast ( $p = 0.01$ ) and between the point forecast and the interval forecast ( $p = 0.007$ ) but no significant difference between the interval forecast and the density forecast. This is shown in the lower panel of Figure 6.3.

My hypotheses do not concern the relative quality of uncertainty estimation in interval and density forecasting. However, in selecting between those two types of forecasting, users may wish to take this issue into account. Hence, I present an analysis of it in Appendix 3.

### 6.2.3. Discussion

Despite using a well-powered between-participant design to eliminate anchoring and other context effects, I obtained no evidence against  $H_1$ . If there is any effect of type of forecast on estimates of central tendency produced by different types of forecast, it must be small. Thus, data from my between-participant design replicates the finding that Bruine de Bruin, Manski, et al. (2011) obtained in their within-participant design. This implies that context effects did not influence their result.

My finding that people overestimated inflation rate is consistent with  $H_2$  and replicates findings from previous studies (e.g., Bruine de Bruin, Van der Klaauw, et al., 2011; Bryan & Venkatu, 2001a, b; Georganas et al., 2014). It appears that people expect inflation to be higher than it turns out to be.

Given the lack of evidence against  $H_1$ , it is not surprising that there was also no evidence suggesting that bias in inflation forecasts depends on the type of forecast made ( $H_3$ ).

I expected that estimates of central tendency derived from point forecasts would be noisier and less accurate than those derived from interval and density forecasts and that those derived from interval forecasts would be noisier and less accurate than those derived from density forecasts ( $H_4$ ). This received partial support: point forecasts were indeed noisier and less accurate than estimates of central tendency derived from interval and density forecasts but estimates of central tendency derived from interval forecasts were not noisier and less accurate than those derived from density forecasts. My hypothesis was based on the ‘wisdom of the inner crowd’ effect (Herzog & Hertwig, 2009, 2014; Van Dolder & Van den Assem, 2018; Vul & Pashler, 2008): the average of a number of judgments from a single person produces a

---

<sup>26</sup> I use generalised eta squared (*ges*) to measure effect size (Olejnik & Algina, 2003).

more accurate estimate than a single judgment from that person. I argued that the mean value of the forecast distribution is estimated just once in point forecasting, twice (albeit implicitly) in interval forecasting, and three or more times (albeit implicitly and depending on the number of bins filled) in density forecasting.

Why did I fail to obtain a difference between the VE and RMSE values associated with interval and density forecasting? First, the benefit gained from averaging more judgments decreases with the number of judgments already averaged. Referring to the increases in accuracy obtained by averaging judgments from more advisors, Budescu and Yu (2007, p 154) point out that “The accuracy of the average opinion increases monotonically as a function of the number of advisors, but at a diminishing rate that depends on the inter-judge correlation”. Thus, it could be that most of the gain to be obtained by aggregating judgments is associated with increasing the number of judgments from one (point forecasting) to two (interval forecasting) and that little extra benefit is obtained by increasing the number of judgments beyond two (density forecasting).

Second, the difficulties that people are reported to have in making density forecasts (O’Hagan et al., 2006) may have increased the random noise in their responses and thereby cancelled out any benefit derived from repeatedly making implicit judgments of the central tendency of the distribution. In contrast, providing interval forecasts is a simple task and so the benefit derived from making an estimate of the central tendency twice would not be diluted by judgment noise associated with performing a difficult task.

### 6.3. Experiment 9

My first experiment showed that, relative to point forecasting, the gain in accuracy from using interval forecasts was as great as the gain in accuracy from using density forecasts. Furthermore, unlike point forecasts, interval forecasts provide survey users with some information about respondents’ estimates of the aleatory uncertainty<sup>27</sup> in the system responsible for generating inflation. This should be important for predicting consumers’ behaviour: they are less likely to act on less certain inflation forecasts.

Thus, the mid-point of the bounds of an interval forecast provides a more accurate estimate of the expected value of inflation than a point forecast and interval forecasts also provide uncertainty information. Furthermore, they are no less accurate than density forecasts and much simpler to produce. Their only drawback is that the uncertainty information that they provide is not as detailed as and not as accurate as that produced by density forecasts. However, if survey users do not require uncertainty information that is as detailed and

---

<sup>27</sup> Aleatory uncertainty refers to uncertainty arising from fundamentally random factors in the environment and is contrasted with epistemic uncertainty that can, in principle, be eliminated by provision of additional information (Tannenbaum et al., 2017).

accurate as that obtained from density forecasts, interval forecasts would provide advantages over point forecasts without incurring the disadvantages of density forecasts.

Surveys ask participants many questions: responses to the earlier questions may influence how later ones are answered (Niu & Harvey, 2022a, Chapter 4). Although the last experiment indicated that such context effects did not influence Bruine de Bruin, Manski, et al.'s (2011) finding that the mean value of central forecast is unaffected by the type of forecast made, it is possible that such effects may differentially influence the noisiness (VE) and accuracy (RMSE) of different types of forecast. For example, the accuracy advantage of interval forecasts may vanish when they are made after point forecasts. Thus, in this experiment I ask whether the accuracy advantage of interval forecasts is preserved in a within-participant design. Using this design, I address the same hypotheses as before. I also examine two other issues.

If context effects do influence the accuracy of different types of forecast (without influencing their mean value), the strength of that influence may be affected by the order in which the different types of forecast are made. Explicit point forecasts may provide stronger anchors for interval forecasts than the (implicit) central values of interval forecasts provide for point forecasts. If they do, point forecasts would reduce judgment noise (VE) in interval forecasts that follow them more than interval forecasts would reduce judgment noise in point forecasts that follow them. Thus, the order in which the two types of forecast are made may affect the noisiness of interval forecasts more than that of point forecasts. With this in mind, I test the following hypothesis.

$$H_6: |VE_{Interval\ second} - VE_{Interval\ first}| > |VE_{Point\ second} - VE_{Point\ first}|$$

In this experiment, I also measure people's confidence in the judgments. I asked them to assess the likelihood (0 -100%) that their point forecast or their interval bounds were within 10% either way of their true values. In other words, I asked them to assess their epistemic uncertainty in their own judgments. As their interval forecast (but not their point forecast) provided their estimate of the aleatory uncertainty in the inflation figures, this allowed me to examine how aleatory and epistemic uncertainty are related (Tannenbaum et al., 2017). People may be more confident in judgments in which they have been allowed to express their uncertainty (interval forecasts) than in those in which they have not (point forecasts).

$$H_7: Confidence_{Point\ Forecasts} < Confidence_{Interval\ Forecasts}$$

### 6.3.1. Method

Two groups of participants made both point forecasts and interval forecasts: one of those groups made point forecasts followed by interval forecasts and the other group made them in the reverse order.

*Participants* One hundred and one participants (65 males, 39 females) with a mean age of 29 years ( $SD = 11$  years) took part in the web-based study. They were recruited from the online

participant recruitment platform, [www.Prolific.com](http://www.Prolific.com), between 16 July 2020 and 13 August 2020 and paid £1.10 for their participation.

*Stimulus materials* Participants made forecasts from 10 graphs, each showing 20 years (2000 to 2019) of real inflation data from 10 countries that were extracted from the World Bank website. For each one, they made both a point forecast and an interval forecast for the 2020 inflation rate. The order of these judgments varied between participants. Both types of forecast were made in the same way as in Experiment 8.

After each forecast, participants expressed their confidence in it by moving a slider that ranged between 0% and 100%. For the point forecast, participants were first asked “How confident are you about the point forecast you just made?” and then moved their slider along a scale that was labelled at the left end “0% - My estimate is **definitely not accurate** to within 10% either way of the correct value”, in the middle “50% - My estimate is **as likely to be** within 10% of the true value as it is not to be within that range”, and at the right end “100% - My estimate is **definitely accurate** to within 10% either way of the true value”. The chosen position on the slider was made numerically explicit with a message posted below it: for example, “You are 60% confident about your forecast”. To reduce anchoring effects, the starting position of the slider was randomised from the 21 possibilities that were 5% apart (0%, 5%, 10%, ... 100%).

For the interval forecast, participants were first asked “How confident are you about the interval boundaries that you set?” and then moved their slider along a scale that was labelled at the left end “0% - My interval boundaries are **definitely not accurate** to within 10% either way of the correct value”, in the middle “50% - My interval boundaries are **as likely to be** within 10% of the true value as it is not to be within that range”, and at the right end “100% - My interval boundaries are **definitely accurate** to within 10% either way of the true value”. The chosen position on the slider was made numerically explicit with a message posted below it: for example, “You are 60% confident about your interval forecast”. To reduce anchoring effects, the starting position of the slider was randomised from the 21 possibilities that were 5% apart (0%, 5%, 10%, ... 100%).

For the same reasons as before, inflation data series were displayed in the central part of the y-axis scale, which ranged from 10% to -6%. The final points for all 10 countries were in the middle of that scale, ranging between 0.08% and 2.90%. The 10 series were broadly comparable, with a mean inflation of 2.05% (SD = 1.75%).

*Design* Forecast type (point forecast, interval forecast) was varied within-participants. The order of these tasks was varied between participants, who were randomly allocated to a point-then-interval group ( $N = 43$ ) or to an interval-then-point group ( $N = 44$ ). Presentation order of the 10 graphs in each task was individually randomised for each participant. The

identities of the 10 real countries were anonymised by labelling them with numbers: for example, “Country 3 of 10”.

*Procedure* The procedure and task instructions for point and interval forecasts were the same as those described for Experiment 8.

### 6.3.2. Results

Forecasts for each country were compared with the actual 2020 inflation rates that were extracted from the World Bank website. Respondents were excluded from the analysis if any of their forecasts (point forecasts or mid-point of the interval forecasts) were beyond three standard deviations of the mean forecast for a particular country and forecast type. This led a sample for analysis of 87 people (56 males, 31 females) with a mean age of 28 years ( $SD = 9$  years).

*Forecasts* After extracting the mean value of forecasts on each trial for each participant in each condition, I carried out a three-way ANOVA on these values using task order (point forecasting first, interval forecasting first) as a between-participant factor and forecast type (point forecasting, interval forecasting) and trial number (1-10) as within-participant factors. This analysis showed no significant main effects or interactions. Thus, I again failed to obtain evidence inconsistent with  $H_1$  but, this time, in a within-participants design of the sort used by Bruine de Bruin, Manski, et al. (2011).

Again, people systematically overestimated inflation (Figure 6.4, Upper panel): a one sample t-test showed that ME was significantly positive ( $t(86) = 19.14, p < 0.001$ ), consistent with  $H_2$ . However, a two-way mixed ANOVA on ME with task order as a between-participant variable and forecast type as a within-participant one revealed no significant effects. As in Experiment 8, we failed to obtain evidence inconsistent with  $H_3$ .

A two-way mixed ANOVA on VE (Figure 6.4, Middle panel) using the same factors as before revealed only a main effect of forecast type ( $F(1, 85) = 4.09, p = 0.046, ges = 0.0058$ ). This provides further evidence consistent with  $H_4$ . However, there was no evidence for the interaction predicted by  $H_6$ : the relative noisiness of point and interval forecasts was not affected by the order in which they were made.

Finally, a two-way mixed ANOVA on RMSE using the same factors as before yielded no significant effects. Despite forecast type significantly affecting VE, this did not feed through to producing a correspondingly significant effect on RMSE (Figure 6.4, Lower panel). Presumably, this was because any such effect was overwhelmed by the influence of ME on RMSE<sup>28</sup>.

---

<sup>28</sup> There was a significant effect of forecast type on MAE, the alternative measure of overall error ( $F(1, 85) = 6.15, p = 0.015, ges = 0.0033$ ). See footnote 23.

Figure 6.4

Experiment 9: Bar charts showing ME, VE, and RMSE scores (with standard error bars) in each condition

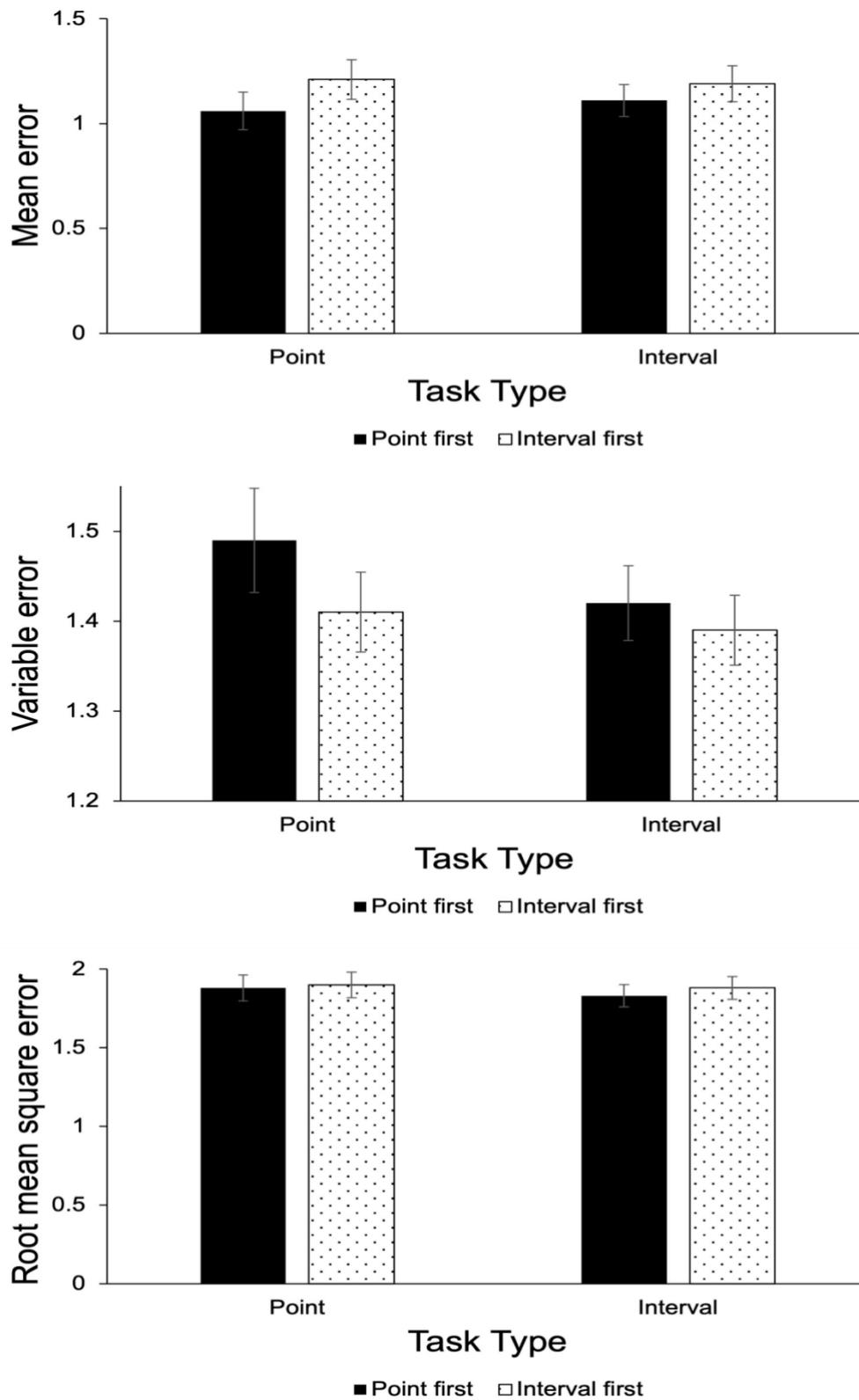
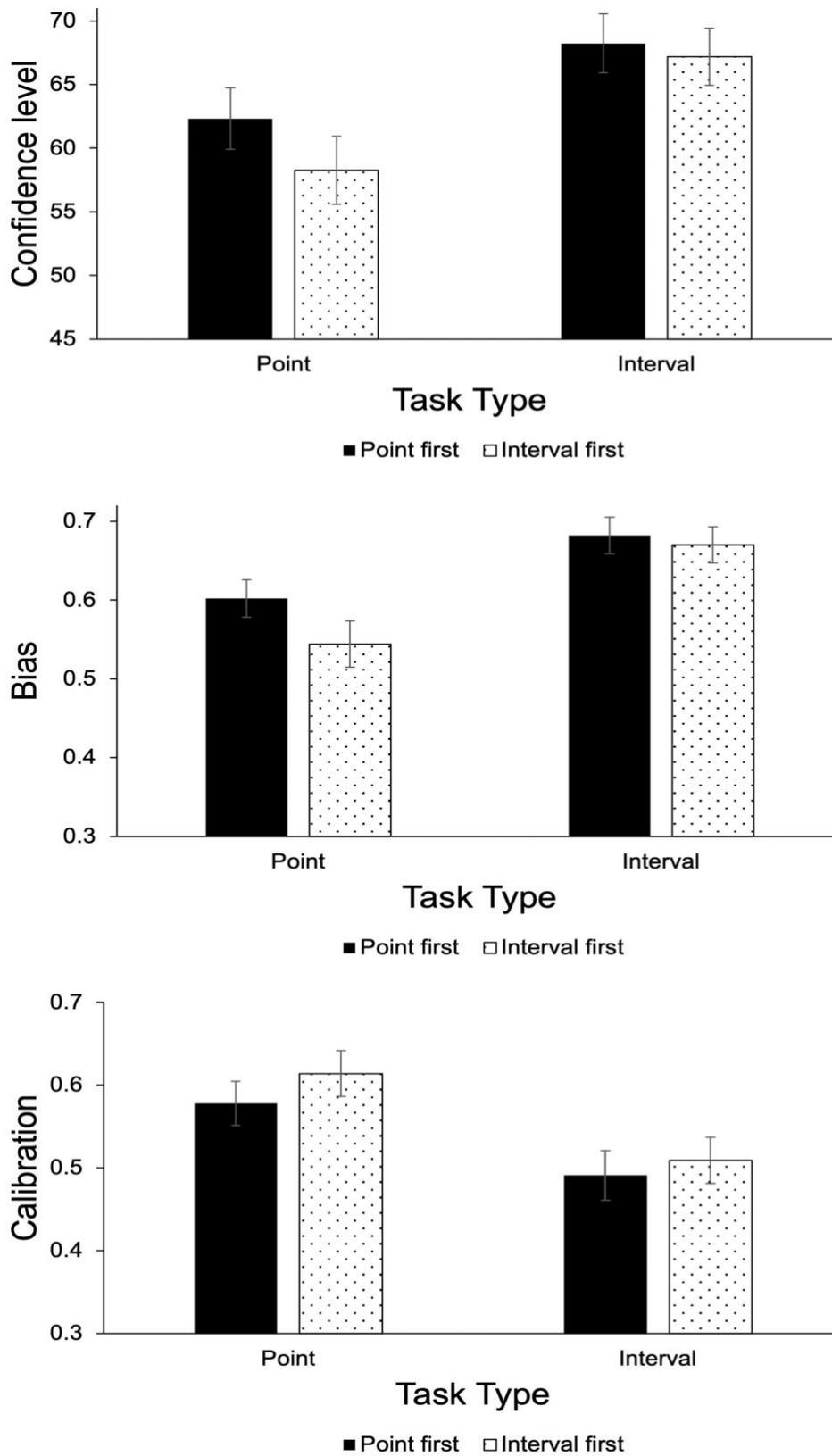


Figure 6.5

Experiment 9: Bar charts showing mean confidence levels, bias levels and calibration scores (with standard error bars) in each condition



*Confidence in forecasts* For each confidence judgment (e.g., there is a 60% chance of my point forecast is within 10% of the true value), I set an outcome index,  $d$ , at 1.00 when the event occurred (the forecast was within 10% of the true value) and at 0.00 when the event did not occur (the forecast was not within 10% of the true value). On each trial, the difference  $(j - d)$  between the judgment,  $j$ , expressed as a probability rather than as a percentage, and the outcome index,  $d$ , then provides a measure of the *bias* in the judgment: higher mean values of this difference indicate greater overconfidence. The square of the bias,  $(j - d)^2$ , is known as the probability score: lower mean values of the probability score indicate a greater ability to assign appropriate probabilities (Yates, 1990, 1994). Here I subtract the probability score from 1.00 so that higher values indicate better calibration: I term this the calibration score.

ANOVAs using the same three factors used for the analysis of forecasts indicated that forecast type had an effect on level of confidence ( $F(1,85) = 27.83, p < 0.0001, ges = 0.0299$ ), on bias ( $F(1,85) = 44.78, p < 0.001, ges = 0.0433$ ), and on the calibration score ( $F(1,85) = 36.10, p < 0.001, ges = 0.0354$ ): people were more confident in their interval forecasts than in their point forecasts, more overconfident in them, and less able to judge how likely they were to be accurate (Figure 6.5).

### 6.3.3. Discussion

The switch from a between-participants to a within-participants design had little effect on the nature of the findings that I obtained. Again, there was no significant difference between mean values of point forecasts and the mean values of the mid-point of interval forecasts ( $H_1$ ). Furthermore, though there was significant over-forecasting of inflation ( $H_2$ ), the size of this bias did not depend on the type of forecast made.

It would, however, be wrong to assume that the change in experimental design had no effect. In my previous between-participant experiment, the mean value of RMSE was 2.58 in the point forecasting condition and 2.18 in the interval forecasting condition: the former was 18% higher than the latter. In contrast, in this experiment, the corresponding RMSE values for point and interval forecasting were 1.89 and 1.85, respectively: the former was only 2% higher than the latter. In other words, the use of a within-participant rather than a between-participant design reduced the percentage difference in RMSE between point and interval forecasting by nine tenths. This is consistent with anchoring having an effect in reducing the difference in accuracy of central forecasts derived from different methods of elicitation. However, in contrast to claim of  $H_6$ , this anchoring effect was not asymmetrical: initial point forecasts acted as mental anchors for the central value of later interval forecasts to the same extent that the central value of initial interval forecasts acted as mental anchors for later point forecasts. As a result, there was no effect of the order in which the two types of forecast were made.

Although the mean of the bounds of interval forecasts provided less variable estimates of future inflation than point forecasts, people were more overconfident in their ability to

estimate those bounds than in their ability to make point forecasts. It is possible that allowing people to express uncertainty in the forecasts (i.e., making interval forecasts) reduces their concern about being wrong and that this, in turn, raises their confidence in their accuracy. Whatever the mechanism, higher confidence in interval judgments could have implications for survey design and the use of survey data.

I suggested that the mean of the bounds of an interval forecast provides a more accurate central forecast than a point forecast because people make two separate (though implicit) point forecasts when estimating intervals. Because of the ‘wisdom of the inner crowd’ effect (Herzog & Hertwig, 2009, 2014; Vul & Pashler, 2008), this acts to cancel out judgment noise (Kahneman et al., 2021) and so increases accuracy. This can explain the difference between these two types of forecasting in both a between-participants design (Figure 6.3) and a within-participants design (Figure 6.4). In the next experiment, I test predictions arising from this account.

#### 6.4. Experiment 10

If the ‘wisdom of the inner crowd effect’ is responsible for the mid-point of the bounds of an interval forecast providing a more accurate estimate of the expected value of inflation than a point forecast, then asking people to make a point forecast from the same data on two separate occasions and taking the average of those judgments should result in less noisy and, hence, more accurate estimates of inflation than either one of those judgments separately. In other words, the VE and RMSE of the average of the two point forecasts should be less than the average VE and average RMSE of the two separate forecasts.

*H<sub>8</sub>: VE of average of two point forecasts < Average VE of two point forecasts*

*H<sub>9</sub>: RMSE of average of two point forecasts < Average RMSE of two point forecasts*

The ‘wisdom of the inner crowd’ effect may or may not be sufficient to explain the difference in accuracy of point and interval forecasts. If it is sufficient,

*H<sub>10</sub>: VE of average of two point forecasts = VE of the mid-point of interval forecast bounds*

*H<sub>11</sub>: RMSE of average of two point forecasts = RMSE of the mid-point of interval forecast bounds*

*H<sub>12</sub>: Average VE of two point forecasts > VE of the mid-point of interval forecast bounds*

*H<sub>13</sub>: Average RMSE of two point forecasts > RMSE of the mid-point of interval forecast bounds*

The ‘wisdom of the inner crowd’ effect is assumed to arise because averaging two or more judgments cancels out some of the noise in the judgments (Kahneman et al., 2021). Thus the effect should be greater when single judgments contain more noise. If separate judgments were noise-free, I would expect no reduction in RMSE after averaging them; if separate

judgments were very noisy, I would expect averaging them to produce a large reduction in both VE and RMSE. People's time-series forecasts contain more noise when the data series on which they are based contain more noise (Harvey, 1995, Harvey et al., 1997). Thus, I expect the 'wisdom of the inner crowd' effect to be greater when people make forecasts from more volatile series.

*H<sub>14</sub>: The effect identified in H<sub>8</sub> and H<sub>9</sub> will be greater with noisier inflation series*

*H<sub>15</sub>: The effect identified in H<sub>12</sub> and H<sub>13</sub> will be greater with noisier inflation series*

#### 6.4.1. Method

To test these hypotheses, I ran two groups of participants. The first group made interval forecasts of inflation for 20 countries. The second group made point forecasts of inflation for those 20 countries and then made point forecasts for those same countries again (in the same order).

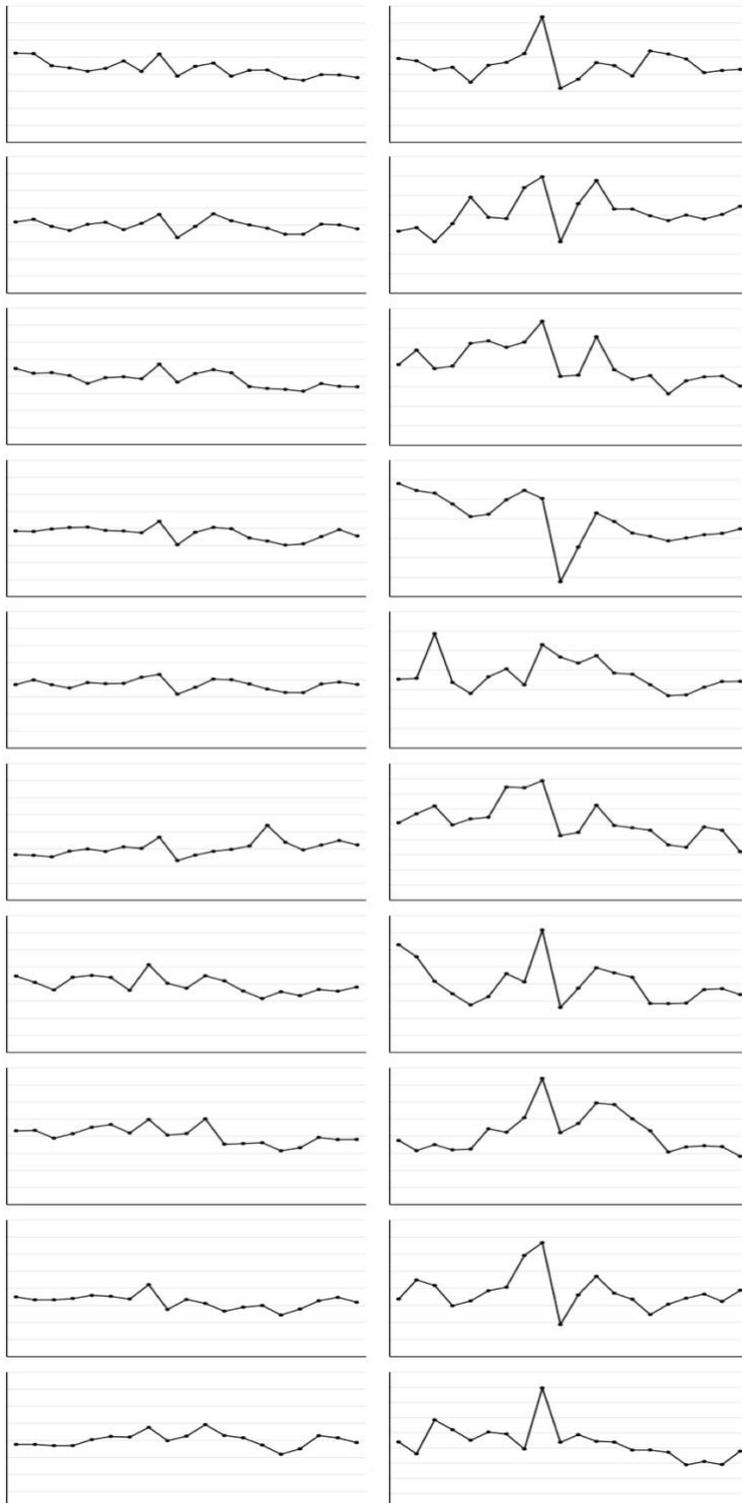
*Participants* One hundred and one participants (66 males, 35 females) with a mean age of 26 years ( $SD = 10$  years) took part in the web-based study. They were recruited from the online participant recruitment platform, [www.Prolific.com](http://www.Prolific.com), between 28 November 2020 and 19 January 2021 and paid £1.00 for their participation.

*Stimulus materials* Stimulus graphs depicting 20 years (2000-2019) of real inflation data from 20 countries extracted from the annual CPI dataset provided by the World Bank. The series were displayed in the central part of the y-axis scale and ranged from 10% to -6%. The final points for all 10 countries were in the middle of that scale, ranging between -0.36% and 2.90%. For the 10 low volatility series, the mean inflation was 1.58% ( $SD = 1.11%$ ); for the 10 high volatility series, the mean inflation was 2.45% ( $SD = 2.28%$ ). Mean levels of variance were 0.70 ( $SD = 0.18$ ) for the 10 low volatility series and 4.50 ( $SD = 1.20$ ) for the 10 high volatility series. These were significantly different ( $t(18) = 8.94, p < 0.001$ ). The two sets of series are shown in Figure 6.6. It is clear that the greater volatility in the second set of countries occurred mainly around the time of the world-wide financial crisis (2008-2011). As before, countries were not explicitly named in the experiment: they were referred to by number.

*Design* Forecast type was a between-participant variable: one group made interval forecasts from the 20 series; the other group made point forecasts from those series and then made another set of point forecasts from those same 20 series in the same order as before. Series volatility was a within-participant variable: forecasts were made for 10 low volatility and 10 high volatility series. Order of the 20 series was randomized separately for each participant. In the point forecast condition, each participant received the series in the same order in the first and second blocks of forecasts. There was no explicit separation of these blocks: as far as participants were concerned, they made 40 forecasts in a single block of 40 series.

**Figure 6.6**

*Experiment 10: The 20 inflation series each showing 20 years of historical inflation data between 1998 and 2017 and ranging between -6% and 10%*



*Note.* Low volatility series are shown on the left and high volatility series on the right. In this figure, axis labels and numbering have been excluded for clarity but they were included in the experimental materials.

*Procedure* As before, participants saw an information screen and responded to a consent screen before receiving a simple definition of inflation (Consumer Price Index) and being given their instructions. Instructions for the point forecasting and interval forecasting groups were as described for the previous experiments. An example picture of an interval forecast or a point forecast (upper two panels of Figure 6.2) was provided before the start of the formal task. At the end of the experiment, participants answered the same demographical questions as before.

#### 6.4.2. Results

Two participants were excluded because of missing values in their data. Others were excluded from the analyses using the same criteria as specified for Experiment 9. As a result, data were analysed from 85 participants (54 males, 31 females) with a mean age of 27 years ( $SD = 10$  years). There were 45 participants in point forecast condition and 40 participants in interval forecast condition.

I calculated ME, VE, and RMSE scores derived a) from the average value of the two point forecasts, b) separately for each point forecast and then averaged, and c) from the mid-point (average) of the bounds of the interval forecast. These different types of error scores are shown in Figure 6.7 for low volatility series (left bar of each pair) and high volatility series (right bar of each pair). Inspection suggests that volatility affects the ME for point but not interval forecasts, VE is comparable for the average value of the two point forecasts and the mid-point of the interval forecasts but that both of these are lower than the average VE of the two point forecasts, and the pattern of RMSE scores closely reflects the combined values of the ME and VE scores.

To test  $H_8$ ,  $H_9$ , and  $H_{14}$ , I carried out ANOVAs on VE and RMSE values derived from the point forecasting group data with type of point forecast error (average error of the two separate forecasts versus error of the average of the two forecasts) and volatility (High/Low) as within-participant variables. The ANOVA on VE scores revealed main effects of type of forecast error ( $F(1, 44) = 52.06$ ,  $p < 0.001$ ,  $ges = 0.0831$ ) and volatility ( $F(1, 44) = 447.95$ ,  $p < 0.001$ ,  $ges = 0.8038$ ) and a significant interaction between these factors ( $F(1, 44) = 4.12$ ,  $p < 0.05$ ,  $ges = 0.0024$ ). Follow-up analyses revealed that the simple effect of forecast type error was significant for both high volatility series ( $F(1, 44) = 35.39$ ,  $p < 0.001$ ,  $ges = 0.0719$ ) and low volatility series ( $F(1, 44) = 52.82$ ,  $p < 0.001$ ,  $ges = 0.1332$ ) and that the simple effect of volatility was significant for both the average VE of the two point forecasts ( $F(1, 44) = 401.36$ ,  $p < 0.001$ ,  $ges = 0.7786$ ) and the VE of the average of the two point forecasts ( $F(1, 44) = 444.70$ ,  $p < 0.001$ ,  $ges = 0.8338$ ).

The ANOVA on RMSE scores revealed main effects of type of forecast error ( $F(1, 44) = 63.48$ ,  $p < 0.001$ ,  $ges = 0.0101$ ) and volatility ( $F(1, 44) = 294.62$ ,  $p < 0.001$ ,  $ges = 0.5163$ ) and a significant interaction between these factors ( $F(1, 44) = 11.54$ ,  $p = 0.001$ ,  $ges = 0.0007$ ). Follow-up analyses revealed that the simple effect of forecast type error was significant for

both high volatility series ( $F(1, 44) = 48.76, p < 0.001, ges = 0.0122$ ) and low volatility series ( $F(1, 44) = 51.07, p < 0.001, ges = 0.0081$ ) and that the simple effect of volatility was significant for both the average RMSE of the two point forecasts ( $F(1, 44) = 268.88, p < 0.001, ges = 0.5064$ ) and the RMSE of the average of the two point forecasts ( $F(1, 44) = 317.58, p < 0.001, ges = 0.5280$ ). In summary, these results are consistent with  $H_8, H_9,$  and  $H_{14}$ .<sup>29</sup>

To test  $H_{10}$  and  $H_{11}$ , I carried out mixed ANOVAs on the VE and RMSE values with type of forecast error (derived from the average of two point forecasts versus derived from the mid-point of the bounds of interval forecasts) as a between-participants factor and with volatility (High/Low) as a within-participants factor. The ANOVA on VE revealed only a main effect of volatility ( $F(1, 83) = 786.17, p < 0.001, ges = 0.8097$ ). Thus I obtained no evidence inconsistent with  $H_{10}$ .

However, the ANOVA on RMSE revealed a main effect of volatility ( $F(1, 83) = 545.50, p < 0.001, ges = 0.5708$ ) and an interaction between volatility and type of forecast error ( $F(1, 83) = 4.30, p < 0.04, ges = 0.0104$ ). Further analysis showed that the simple effect of volatility was significant both for the RMSE derived from the point forecasts ( $F(1, 44) = 317.58, p < 0.001, ges = 0.5280$ ) and for the RMSE derived from the interval forecasts ( $F(1, 39) = 235.62, p < 0.001, ges = 0.6749$ ) but that the simple effect of type of forecast error was significant only for the high volatility series ( $F(1, 83) = 5.54, p = 0.02, ges = 0.0625$ ).

The significance of this interaction for RMSE but not for VE strongly suggests that it arose because the ME for the average of the two point forecasts was affected much more by higher volatility than the ME associated with the mid-point of the bounds of the interval forecast. An ANOVA on the ME scores using the same factors as before confirmed this. It revealed a main effect of type of forecast error ( $F(1, 83) = 4.74, p = 0.03, ges = 0.0459$ ) and an interaction between that variable and volatility ( $F(1, 83) = 5.70, p = 0.02, ges = 0.0107$ ). ME values for point forecasts but not for interval forecasts were higher when series were more volatile. Thus the simple effect of volatility was significant only for the point forecasts ( $F(1, 44) = 5.82, p = 0.02, ges = 0.0204$ ) and the simple effect of type of forecast error was significant only for high volatile series ( $F(1, 83) = 5.66, p = 0.02, ges = 0.0638$ ). In summary, I found no evidence inconsistent with  $H_{10}$  but I did obtain evidence inconsistent with  $H_{11}$  because of this unexpected effect of forecast type on ME, one of the contributors to RMSE.

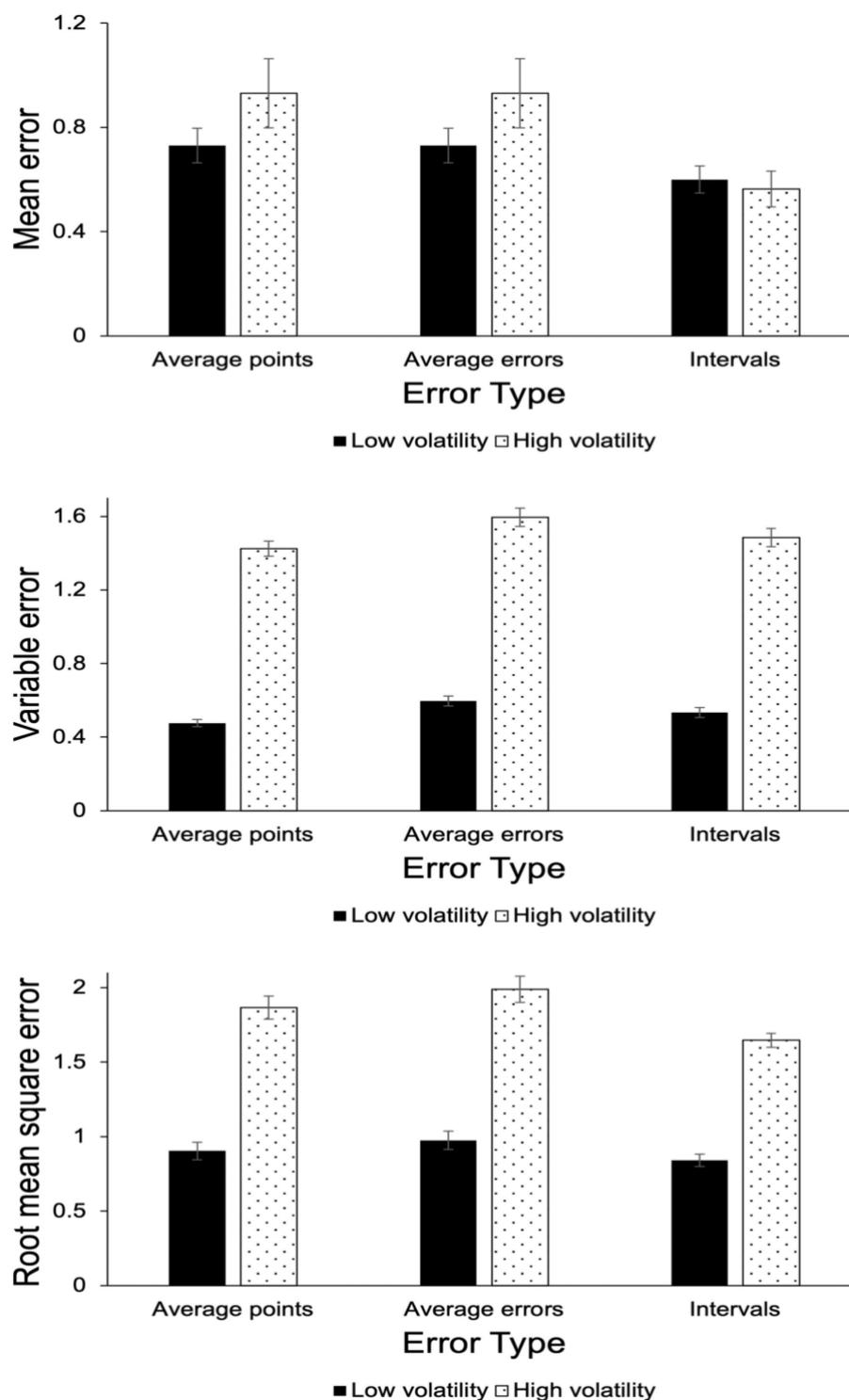
Finally, to test  $H_{12}, H_{13},$  and  $H_{15}$ , I carried out mixed ANOVAs on the VE and RMSE values with type of forecast error (the average of the errors calculated separately for each point forecasts versus the error derived from the mid-point of the bounds of interval forecasts) as a between-participants factor and with volatility (high/low) as a within-participants factor. The analysis

---

<sup>29</sup> ME derived from the average of the two point forecasts is the same as that obtained by taking the average of the ME scores calculated for each forecast separately. But, however they were calculated, ME scores were found to be higher with more volatile series ( $F(1, 44) = 5.82, p = 0.02, ges = 0.0204$ ).

Figure 6.7

Experiment 10: ME, VE, and RMSE scores for inflation forecasts made from low and high volatility series.



Note. These scores are based on the average value of the two point forecasts (Left), the average error of the two point forecasts (Middle), and the mid-point of the interval forecasts (Right).

of VE revealed main effects of type of forecast error ( $F(1, 83) = 3.87, p = 0.05, ges = 0.0274$ ) and volatility ( $F(1, 83) = 744.03, p < 0.001, ges = 0.7805$ ) but no interaction between these factors.

The analysis of RMSE revealed main effects of type of forecast error ( $F(1, 83) = 8.75, p = 0.004, ges = 0.0772$ ) and volatility ( $F(1, 83) = 490.41, p < 0.001, ges = 0.5491$ ), together with an interaction between these variables ( $F(1, 83) = 6.38, p = 0.01, ges = 0.0156$ ). Further analysis revealed a simple effect of type of forecast error only for high volatility series ( $F(1, 83) = 10.97, p = 0.001, ges = 0.1168$ ) and a simple effect of volatility for both point forecast error ( $F(1, 44) = 268.88, p < 0.001, ges = 0.5064$ ) and interval forecast error ( $F(1, 39) = 235.62, p < 0.001, ges = 0.6749$ ). Again, this interaction was observed for RMSE but not for VE because of the unexpected effect of volatility on the ME of point forecasts but not on the ME of interval forecasts. (Because ME is the same whether it is calculated as the average ME of the two separate point forecasts or as the ME of the average of the two point forecasts, the analysis of ME reported in the last paragraph applies here too.) In summary, these results are consistent with  $H_{12}$ ,  $H_{13}$ , but only partially consistent with  $H_{15}$ .

#### 6.4.3. Discussion

Comparison of VE and RMSE scores for the two ways of averaging point forecasts revealed results that I expected. Error scores based on the average of the two forecasts were significantly lower than averages of the error scores calculated for each forecast separately. This was expected on the basis of a 'wisdom of the inner crowd' effect. Furthermore, if this effect reduces VE by some proportion (e.g., 20%), then the absolute size of the reduction should be greater when VE is higher. VE is higher when forecasts are made from noisier series (Harvey, 1995). Hence, I expected the 'wisdom of the inner crowd' effect to be greater with the noisier series; the interactions between series volatility and forecast type for VE and RMSE provide evidence that it was indeed greater for noisier series.

Turning now to the comparisons between VE based on the mid-point of the bounds of the interval forecasts and the two values of VE based on the different methods of obtaining VE from point forecasts, I found that there was no evidence of a difference between the VE values derived from the interval forecasts and the VE values derived from the average of the two point forecasts. There was, however, a significant difference between the VE values derived from the interval forecasts and the average of the VE values calculated separately for each of the two point forecasts. Taken together, these results are consistent with the 'wisdom of the inner crowd' effect being sufficient to explain why VE values derived from the mid-point of the bounds of interval forecasts are lower than those derived from single point forecasts (Experiments 1 and 2).

Consistent with this, there was a main effect of volatility but no interaction between volatility and type of forecast when the VE values derived from the mid-points of the bounds of interval forecasts were compared with the VE values derived from the average of the two point

forecasts: no interaction was expected because there was no difference in the size of those VE values. However, I had expected the corresponding interaction to be significant when VE values derived from the mid-points of the bounds of interval forecasts were compared with the average of the VE scores calculated separately for each of the two point forecasts. In fact, this interaction did not attain significance.

There was one other way in which the results did not turn out in the manner I had predicted. In developing my hypotheses, I had expected that the results from the analyses of RMSE values would broadly reflect those obtained from the analyses of VE scores. My expectations were based on the assumption that ME scores would not be influenced by volatility or by whether they were derived from point forecasts or from interval forecasts. In fact, volatility increased not only VE (which I had expected) but also ME (which I had not expected). Furthermore, this effect of volatility on ME was restricted to ME values derived from point forecasts; it was not present for ME values derived from interval forecasts. Because of these differential effects of volatility on different error sources, results from analyses of RMSE did not reflect those from my analyses of VE in the way I had expected. In particular, analysis of RMSE scores derived from the mid-points of the bounds of interval forecasts and those derived from the averages of the two point forecasts revealed an interaction between volatility and error source (whereas analysis of the corresponding VE scores did not). Similarly, analysis of RMSE scores derived from the mid-points of the bounds of interval forecasts and the average RMSE scores calculated for each point forecast separately also revealed an interaction between volatility and error source (whereas the corresponding VE scores did not show that interaction between even though it was expected).

Why did higher volatility increase ME scores that are derived from point forecasts? There have been many demonstrations that lay people tend to expect inflation to be higher than it turns out to be (e.g., Bruine de Bruin, Van der Klaauw, et al., 2011; Bryan & Venkatu, 2001a, b; Georganas et al., 2014). When series contain little noise, they are constrained in how much they can over-forecast inflation while still producing a plausible prediction. However, when series are noisy, greater over-forecasting is possible because forecasts well above the statistical expectation may still be plausible if they are within the envelope provided by previous outliers. Why did higher volatility not increase ME scores that are derived from the mid-points of interval forecasts? People producing interval forecasts are likely to respond to higher volatility not by taking the opportunity of raising the mid-points of their interval forecasts but by simply widening the interval that they provide (without changing its mid-point).

## 6.5. General Discussion<sup>30</sup>

Domains in which judgmental forecasts are required differ in terms of the type of forecast that is seen as most appropriate. When many forecasts are needed within a limited period of

---

<sup>30</sup> Hypotheses and findings from the three experiments are summarised in Table A6.1 of the Appendix 4.

time, point forecasts are likely to be preferred even though they provide users with no information about the uncertainty associated with the forecasts. Demand forecasting frequently provides an example of such a domain. However, when users need information from forecasters about the uncertainty associated with their forecasts and are willing to sacrifice speed of forecasting to obtain this information, interval and density forecasting are more appropriate. Which of these is selected depends on how detailed the information about the uncertainty associated with the forecasts needs to be: interval forecasts provide only basic uncertainty information whereas density forecasting provides the distributional information needed in certain domains, such as finance or meteorology.

Within this overall framework, users make decisions about the type of forecast they require by identifying where their needs lie on a trade-off between speed (or convenience) of providing forecasts and the detail those forecasts contain about the uncertainty inherent in them. The work that I have reported here demonstrates that there are additional factors that should enter into their decisions. First, estimates of the expected value for the period being forecast are subject to more judgment noise in point forecasts than in interval or density forecasts: as a result, point forecasts are more inaccurate. Second, people are more overconfident in interval forecasts than in point forecasts: they are less able to assess the likelihood that their judgments are correct. Third, in domains in which forecasts are typically biased (inflation forecasting, sales forecasting), the level of bias increases with the volatility of the data on which forecasts are based when point forecasts are made but not when interval forecasts are made. Arguably, all these factors should be taken into account when users specify the type of forecast they require.

My results are consistent with the lower variable and overall error in interval and density forecasts than in point forecasts arising from a 'wisdom of the inner crowd' effect (Herzog & Hertwig, 2009, 2014; Van Dolder & Van den Assem, 2018; Vul & Pashler, 2008). I suggest that the expected value or central tendency of the variable for the period being forecast is estimated just once for point forecasts but more than once for interval and density forecasts. Consistent with this, my third experiment demonstrated that the relative disadvantage of point forecasting can be eliminated by eliciting point forecasts twice and using the average of the resulting estimates. Of course, for this to be effective, an effort must be made to ensure that forecasters do not remember the exact value of their first forecast when making their second one. I accomplished this by eliciting a set of 20 forecasts and then eliciting that set again; this procedure was sufficient to produce a significant reduction in VE, an indication that people were unable to fully remember their first forecasts when making their second ones.

### *6.5.1. Limitations*

I examined point, interval, and density forecasting. Other types of forecasting are sometimes used. For example, rather than asking forecasters to set an interval such that there is a 90% likelihood that the outcome will lie within it, it is possible to provide forecasters with a fixed interval and ask them to estimate the likelihood that the outcome will fall within it. Though

this latter approach is rarely used in practice, it reduces biases in forecasts (Hansson, Juslin, et al., 2008). In addition, the above types of forecast are sometimes used in combination: for example, forecasters may specify a point forecast and then place an interval around it or, alternatively, they may set an interval and then place a point within it to signify the most likely outcome within the interval. These procedural variations could influence ME or VE and might lead to modification of my conclusions. However, a ‘wisdom of the inner crowd’ approach could be used to predict how they would do so. For example, I would expect that a point forecast made before an interval forecast would contain a greater variable error than one made after it.

I discussed the trade-off between the simplicity of producing a particular type of forecast and the usefulness of the forecast. For example, density forecasts are relatively difficult to produce but they provide more detailed and more accurate information to users about the uncertainty associated with the forecast. However, while sophisticated users will easily absorb this additional detail (Armstrong, 2001; Ramos et al., 2013; Roulston & Kaplan, 2009), others may find it difficult to interpret these more complex forecasts (Fischhoff, 1994; Ramos et al., 2013; Yaniv & Foster, 1995; Yates et al., 1996). For example, Bruine de Bruin, Manski, et al. (2011) reported that respondents found density forecasts for price inflation were more difficult to appreciate and less clear than point forecasts. Here I have drawn attention to one previously ignored factor (accuracy of estimation of expected values) that should be taken into account when selecting the type of forecast that will be made. However, I have not studied how forecasters and users weigh the importance of different factors when making their choice.

### *6.5.2. Implications*

Lay surveys have consistently shown that people overestimate inflation. One explanation of this is that people are more likely to recall large price increases for specific goods (e.g., rice) because those are more salient and memorable than many smaller price increases (Bruine de Bruin, Van der Klaauw, et al., 2011). For this to be a factor, the respondents must have experienced those increases in their own country. However, I obtained consistent overestimation of inflation by people who did not even know the countries for which they were making inflation forecasts. It is not easy to reconcile this finding with a model based on personal and selective recall of large price rises. It is more consistent with people having a generally biased view of inflation, perhaps because the media devotes more coverage to possible inflation increases than inflation decreases (even when those are equally likely). This ‘risk amplification’ via the media could cause a difference in the availability of different possible levels of future inflation that influences expectations (Tversky & Kahneman, 1974).

### *6.5.3. Conclusions*

Selecting between point, interval, and density forecasting should not be just a matter of trading-off simplicity for potential usefulness. There are other relevant factors that need to

be taken into account. In particular, different types of forecasting vary in how well they are able to produce estimates of the expected values of future periods of variables of interest: these estimates differ in bias, judgment noise, overall accuracy, and the degree of overconfidence associated with them.

## Chapter 7 Conclusions

This research, using a series of experiments, is one of the first studies to explore the formation of inflation expectations from the point of view of psychology: price change information held in memory, descriptive historical data provided externally, learning effects arising from feedback, and response elicitation approaches. Here I summarize the main results of this research.

### 7.1. Summary of findings

In Chapter 3, I investigated how lay people form inflation expectations and I concentrated on when the role of personal purchasing information on inflation expectations matters. I proposed three models: a price-free model, a price-recall model and a price-salience model. Specifically, through task manipulations, participants were asked to complete the price assessment task and inflation expectation task separately in order to examine differences and correlations between those two sets of judgments. Findings from three experiments that were conducted in different years (2019 vs. 2022) revealed that whether lay people use the information on price changes to form inflation expectations depends on the actual inflation environment. Lay people put more effort into monitoring price change information when the actual inflation is generally high and full of uncertainty (a price-recall model, Experiment 2a) compared with a low and steady inflation situation (a price-free model, Experiment 1). Furthermore, the price-salience model is appropriate when the overall inflation rate is generally high and fairly stable, but a few product categories exhibit dramatically high price increases (Experiment 2b). Also, the priming effect of “asking for the largest price changes” did not affect people’s later judgments of inflation when a price-recall model was used (Experiment 2a).

Chapter 4 considered survey measurement of lay people’s inflation expectations and showed that providing external information (both within series and across series historical data) has a beneficial effect by improving the accuracy of inflation expectations. When the historical data information associated with inflation (inflation rate, unemployment rate, or interest rate) is available to lay people, their inflation judgments were significantly more accurate than when those judgments were made without the provision of external information. This can partially explain why expert judgements in professional surveys are superior to the judgments of lay people in consumer surveys (Experiment 3); expert surveys incorporate historical information but lay surveys do not. Also, as surveys elicit either separate evaluations of current and future inflation rates or joint evaluations of those rates, my experimental design manipulated the number of questions asked. People made inflation judgements either for the current or for the upcoming year (a between-participant design, Experiment 1), or else they made joint inflation judgements both for the current and for the upcoming year (a within-participant design, Experiment 4). Inflation judgments for the upcoming year were significantly higher than those for the current year only in the within-participants design experiment. This implies that lay people had an implicit assumption that inflation increased

over time. Experiment 5 also showed that providing more *useful* information (past values of inflation) is better than providing *more* information (past values of the unemployment rate and interest rate).

After demonstrating that lay inflation expectations are inaccurate, my focus was on finding ways of reducing this inaccuracy. Chapter 5 showed that this can be achieved by provision of simple outcome feedback (i.e., the actual inflation rates). On the one hand, lay people learnt to reduce their over-estimation by receiving information about the difference between their responses and actual values. On the other hand, they also learned to make better confidence judgments by reducing their bias towards overconfidence bias and improving their calibration scores. However, the overall error of inflation forecasts was not reduced because simple OFB increased the judgment noise (inconsistency) at the same time as reducing overconfidence (Experiments 6 and 7). In this chapter, I also showed that the beneficial effects of providing simple OFB (i.e., decreasing overestimation and overconfidence) did not depend on how much feedback was provided to lay people. This suggests that once participants have received some information about their bias, no additional advantage is to be obtained by receiving even more OFB (Experiment 7). Furthermore, provision of summarized OFB (the average performance outcome of previous forecasts) rather than simple OFB did not increase judgment noise but produced higher overconfidence than simple OFB (Experiment 7).

Chapter 6 examined the effects of using different survey response formats to obtain inflation expectations: point forecasts, interval forecasts, and density forecasts. The central tendency of each elicitation method was extracted and compared. Point inflation forecasts resulted in more judgment noise and overall error but less overconfidence than those participants who made interval forecasts and density forecasts of inflation (Experiment 8). These results were replicated in a within-participant design (Experiment 9). Furthermore, the accuracy advantage of interval forecasts persists regardless of the volatility of data series (Experiment 10). I also showed that a “wisdom of the inner crowd effect” can explain why the mid-point of the bounds of an interval forecast provides a more accurate estimate of the expected value of inflation than a point forecast (Experiment 10); the disadvantages of point forecasting could be eliminated by eliciting point forecast twice and using the average of the resulting estimates. The overall findings also suggest that the choice of elicitation method should not be determined solely by the trade-off between accuracy and simplicity but should also take the characteristics of the task and confidence measure into account.

## **7.2. Implications**

Although algorithms are widely used in the forecasting domain, pure judgmental forecasting still plays an essential role in certain contexts. Human cognition in forecasting is an extremely important but often an overlooked domain both in economic psychology research and in forecasting research. In this thesis, I demonstrated experimentally how human factors interact with the information provided to inflation forecasters and with the measures used to assess forecasting quality. Understanding the judgmental forecasting process underlying

formation of inflation expectations has important implications in both theoretical and applied settings. In particular, findings from this research may have applications for improving judgmental forecasting accuracy, for media guidance and for government policymaking.

### *7.2.1. Implications for improving judgmental forecasting accuracy*

Forming accurate inflation judgments is crucial for making reasonable financial decisions and establishing efficient communication between the government and the public. Though increasing financial literacy and education have been proposed as ways of achieving these goals (Lusardi & Mitchell, 2011b; Goyal & Kumar, 2021), they cost time, money and effort and there is heterogeneity in the effectiveness of financial literacy and education interventions (Hastings et al., 2013).

Findings from this thesis suggest ways in which the accuracy of lay inflation expectations could be increased. First, provide forecasters with useful information in advance. People who were provided with useful information when making inflation forecasts made more accurate judgments than those who did not receive that information. It is also critical to offer useful information, such as the historical data of inflation rates rather than trying to offer a large amount of information. Thus, even without provision of an extensive education program, identifying the most useful information and giving it to lay forecasters could help them to reduce forecast errors.

Second, it is appropriate to give forecasters simple outcome feedback. Here experimental results showed that receiving this feedback is an efficient way to reduce forecasting bias in inflation forecasting. I would expect that this finding would be generalizable to time-series forecasting tasks in other contexts because it shows that people are able to use OFB to recognize their judgment errors and so learn to reduce them. Also, OFB reveals to individuals their limitations and so produces reductions in their overconfidence. This implies that it diminishes any illusions of learning from practice: people intuitively assume, often wrongly, that their abilities improve with having more practice/experience on a task (Harvey et al., 1987; Harvey, 1994). Also, though I showed that providing feedback increases noise in people's forecasts, this could be reduced or even eliminated by use of certain approaches, such as the "wisdom of the inner crowd" effect (Vul & Pashler, 2008; Van Dolder & Van den Assem, 2018; Herzog & Hertwig, 2014).

Third, asking the correct question is a precondition to obtaining correct answers. When framing questions in surveys, there are two aspects that need to be taken into consideration: the wording of questions, and the format of answers. Regarding question wording, it has been shown that even though two different question wordings ought to result in the same forecasts (e.g., "price changes in general" and "inflation rate"), they, in fact, produce different inflation expectation responses (Bruine de Bruin et al., 2012; Bruine de Bruin, Van der Klaauw, et al., 2011). In this research, I additionally showed that different mental processes can occur when the same wording is used in different environments. For example, a task using the wording

“inflation” may trigger the same information processing procedure as “price changes in general” wording or it may induce direct recall of other information sources, such as media reports about inflation. This signals the need to understand the fundamental psychological mechanisms of inflation forecasting.

The choice of elicitation method is also critical for obtaining the best responses that people can produce. It is well known that probability density forecasts of inflation are more difficult to elicit than point forecasts and interval forecasts of inflation. However, point forecasts tend to be more biased than forecasts using interval and density formats. Elicitation difficulty level (and, hence, any resulting errors) may be amplified when people need to put more effort into converting price information from one format to another (Ranyard et al., 2008). Vanhuele and Drèze (2002) argued that people tend to favour a particular form of encoded prices (and this may be different across individuals). It is reasonable to argue that the best way of eliciting accurate responses is for questions to use a price format that matches the mental representations of prices as they are encoded in memory. However, there may be other factors should be taken into account when deciding what response format should be used; for example, finer grained price formats (e.g., cents rather than dollars) may affect perceived informativeness or data volatility and, hence, confidence in judgments.

### *7.2.2. Implications for media guidance*

The research findings indicate that people employ information from both inside memory and outside in the environment when forming their inflation expectations. However, different people are likely to put different weights on different sources when forming their expectations.

Media, as a set of advanced and ubiquitous tools, could be encouraged to provide information that would help to people to make accurate forecasts (of inflation and other economic variables). For example, if information that gives professional forecasters an advantage over lay forecasters (Binder & Rodrigue, 2018; Cavallo et al., 2017; Armantier et al., 2016) could be made accessible to lay people via the media, that advantage might disappear. However, provision of information may not be enough because lay people are likely to pay attention to information selectively and occasionally and so update their judgments only rarely whereas professionals are likely to attend to information more frequently and update their judgments more often. In summary, more media exposure may help lay people to improve the accuracy of their inflation expectations but it may not be sufficient to raise their performance up to the level of professionals.

Media managers could be encouraged to avoid over-exposure of certain types of events that catch the public’s attention because that would be expected to bias information acquisition and, hence, distort subsequent cognitive processing underlying judgment formation. This could occur in various ways. For example, because people use the availability heuristic to judge the likelihood of events having occurred or being true (Kahneman & Tversky, 1973;

Tversky & Kahneman, 1973; Tversky & Kahneman, 1974), distorted reporting (e.g., greater reporting of high price rises in just a few products than of much lower price rises in all other products) is likely to result in biased expectations for inflation overall. If only the high price rises are reported but people know that they are not representative of overall inflation, they are likely to use the reported high rises as a judgment anchor and adjust downward to allow for their non-representativeness. However, with the anchor-and-adjustment heuristic, adjustment is typically insufficient. As a result, overall inflation will be over-estimated. This is the process underlying the price-salience model that I discussed in Chapter 3. Media reporting could have other effects too. For example, it could strengthen base-rate neglect, thereby pushing people to overreact to news and to ignore the entire distribution of other possibilities (Denrell & Fang, 2010). The narrative nature of media reports may also be important: a good storyteller may easily introduce (incorrect) casual correlations to impress the public with a strong story line. This may exaggerate the audience's tendency to suffer from "the illusion of correlations" (Tversky & Kahneman, 1974) or to "see patterns" where there are none (Goodwin, Fildes, et al., 2011). All these potentially distorting influences of media reporting may bias inflation expectations. It is difficult to minimise these media amplification effects while respecting freedom of the press. However, by making editors and media managers aware of them, it may be possible to reduce them somewhat. If individual media consumers are also made aware of them, it may be possible to reduce their influence further.

### *7.2.3. Implications for governmental policymaking*

Techniques for improving forecasting accuracy may also have implications for relevant governmental organizations. For example, the work I have discussed can provide some guidance for the design of widely used consumer surveys in various ways: using the most appropriate question wordings, providing the most useful information, and choosing the ideal elicitation format for answers. Taking information provision as an example, information about historical data on inflation rates and the inflation target could improve inflation forecasting accuracy by providing relevant knowledge and by supplying a judgment anchor on which to base inflation expectations.

Another important characteristic of survey design is the structural design of questions. I found that lay people's expectations influence their judgments: they assumed inflation rises over years, so their estimates of current inflation and future inflation were different only when the two estimates were obtained from the same person. This result suggests that the context effect provided by prior questions could trigger different (additional) cognitive processing. Hence, it is essential for relevant government departments to identify the most appropriate designs for their surveys in order to acquire valid and precise answers from survey respondents.

Findings from this research also encourage communication between monetary policy-makers and the public. Lay people and professionals have the same access to price change information and the media, but evidence shows that professionals hardly use price changes

as indicators future inflation (Palardy & Ovaska, 2015) and the heterogeneity of professionals' responses does not depend on media coverage and reporting intensity (Lamla & Maag, 2012). Actually, professionals have more incentives to acquire the most recent and pertinent information, such as information from statistical offices and selected forecasting models irrespective of media activity. Similarly, information on the inflation target is assumed to be equally delivered to professionals and lay people, but the inflation target appears to have different effects on the formation of inflation expectations in professional forecasters and lay individuals. Only professionals' inflation expectations appear well-anchored on inflation targets in a short term and long term on average (Coibion, Gorodnichenko, Kumar, et al., 2020; Gábrriel et al., 2014). It is possible that the communication of public information on monetary policy and inflation targets is not clear to lay people. This suggests that further actions are needed from policymakers to make monetary policy announcements reach the general public more effectively (Rast, 2022).

Governments could also provide personalized and customised training schemes and economic knowledge support (Lusardi, 2019) to people, especially to those people who report extremely high values of inflation expectations. Improving basic and advanced knowledge of economy and finance can help them form more accurate inflation expectations and make better financial decisions. This could have long-term advantages for both individual wealth and the development of the economy.

### **7.3. Limitations and future directions**

This research suggests that there is a need for future investigations into the psychological mechanisms underlying judgmental forecasting of inflation. Some factors affecting inflation expectations have been identified here, but considerations might be different in real-world contexts where forecasting is carried out in a more complex environment.

One area of further research directly related to the current dissertation is the presentation format of time-series data provided to lay forecasters. This is likely to influence the forecast errors observed. For example, it has been shown that graphs and tables draw attention to different features of data series. Particular formats influence the cognitive process responsible for forecasting and therefore produce different types of biases (Levy et al., 1996; Lawrence & Makridakis, 1989; Godau et al., 2016; Glaser et al., 2019; Boduroglu, 2022). Additionally, the different graph formats such as line graphs, point graphs, and bar charts affect the perceived dependency level of successive data points (Zacks & Tversky, 1999; Theocharis et al., 2019).

In daily life, a large amount of information is described in text or framed as stories, just like the information disseminated by media. This suggests that one future area of research would be to examine the specific role of the media in the formation of inflation expectations. Here I concluded that lay people are able to monitor the price changes and/or media reports to make their inflation forecasts, but few experiments have shown direct effects of the media

on inflation expectations. However, research has shown that media coverage of inflation affects disagreement between consumers (Lamla & Maag, 2012). Specifically, the tone, the volume and the channel of the media reports on inflation affect inflation expectations at an aggregation level (Soroka, 2006; Lamla & Lein, 2014; Conrad et al., 2022). However, at the individual level, effects of how people react to inflation news has barely been investigated. It is not yet clear how they deal with information from a variety of media sources, especially when the weights given to those sources could change over time. In addition, the trustworthiness of the media sources may interact with the relative weights that people allocate to personal price change experiences and to information conveyed by media reports. Sources that are more trustworthy are likely to have stronger influences on judgment formation because it is known that the way in which information is combined depends on factors such as the likeableness, familiarity and credibility of information sources (Birnbaum et al., 1976; Yaniv, 2004). So measures such as the trust levels and likeability levels of different sources could be measured and used to predict the influence of those sources in the formation of inflation expectations.

There is a large corpus of forecasting research that has been used to develop principles of forecasting in practice (Lawrence et al., 2006; Armstrong, 2001; Armstrong et al., 2015). The current research has extended that work to provide suggestions for principles concerned with giving feedback and choosing an elicitation method. In the future, the usefulness of and cognitive underpinnings of other approaches could be investigated in attempts to improve judgmental forecasting. For example, use of scenarios (Wright & Goodwin, 2009) could mitigate or reduce effects of inappropriate framing, cognitive biases, motivational biases, and incorrect causal attributions. In addition, qualitative research may allow a deeper understanding of how lay people make sense of inflation and related economic variables.

In practice, both long-term (e.g. five-year ahead) and short-term (e.g. one-quarter ahead) forecasts are required for inflation forecasting. Here I focused only on the forecasts of one-year ahead inflation rates. The underlying models of short-term inflation expectation formation may be different from the models of long-term inflation expectation formation. Inflation further into future is forecast with higher uncertainty. I have shown that the provision of historical data is more beneficial for evaluation of the current year's inflation than for evaluation of next year's inflation. So, in future, it may be possible to study how people choose which information to utilize, the weights that they put on to that information, and how their use of mental heuristics varies with different forecast horizons.

An important feature of inflation expectations is that they may act as self-fulfilling prophecies. In other words, the level of judgmental forecasts of inflation (with high confidence in their judgments) influences economic decisions that people make, and these decisions, in turn, contribute to the actual level of inflation. Hence, two research questions need to be further disentangled: a) effects of confidence levels in inflation forecasts and b) the degree of consistency between judgments and actual decisions. The shift from forecasts to subsequent

behaviours is more likely to happen when people are pretty sure about their forecasts (the high subjective probability of one's own judgments being correct). In addition to measures of confidence, other variables related to confidence could be studied. These include the frequency of forecast revision, the size of forecast revision, and the consistency of behaviours over time (Galashin et al., 2020; Cornand & Hubert, 2022; Bruine de Bruin, Manski, et al., 2011).

Future research could adopt new methodological approaches. Most of the existing work on inflation expectations has used macro-econometric methods and experimental approaches. However, to make explicit the effects of particular variables and exclude other possible explanations, techniques of computational modelling in cognition may help to provide a novel source of insight into economic behaviour. It is also important that the quality of data is high (e.g., including a wide variety of demographic variables) as this will allow researchers to take account of the individual differences in inflation expectations that are known to be important (Armantier et al., 2016; Bruine de Bruin, Van der Klaauw, et al., 2010; D'Acunto, Malmendier, Weber, et al., 2021).

#### **7.4. Conclusions**

In this thesis, I investigated how lay judgmental forecasts of inflation are formed through a series of experiments. Inflation expectations are of great importance in the areas of macroeconomics, finance and psychology; pure judgmental forecasting of inflation is critical for both individuals and organizations. The findings from the current research revealed that lay people are able to employ price change information and other information like media reports of inflation to form their inflation expectations. Importantly, in order to compensate for the cost of information searching and the demands on cognitive resources, people use a strategy of switching the type of information used for forecasting on the basis of the inflation environment. Both inflation expectations of laypeople and inflation expectations of professionals are measured using surveys, but their surveys are framed in different ways. I found that additional information displayed in the professional surveys but not in the consumer surveys increased inflation forecast accuracy once it was provided to lay people. Having a prior question (current inflation estimation) also influenced their responses to future inflation expectations. Furthermore, the research demonstrated that there are ways of improving inflation forecasting by the public. Specifically, providing simple outcome feedback significantly reduced forecast bias (over-estimation) and confidence bias (over-confidence). Using the interval forecast rather than the point forecasts reduced judgment noise and the effect of volatility in past data. These findings revealed underlying psychological mechanisms that are likely to generalize to other types of judgmental forecasting task. Overall, the exploration of factors affecting inflation forecasting should contribute to a better understanding of how people's inflation judgments are affected by the information provided and the measures that are used to assess them. The research shed light on the research fields of forecasting, economic psychology and cognitive science. Future research efforts should be

aimed at exploring the role of the media and how it interacts at the cognitive level with other sources of information. Other approaches to enhancing inflation forecasting accuracy also require further research.

## References

- Abelson, R. P. (1963), Computer simulation of 'hot' cognition. In S. S. Tomkins & S. Messick (eds.), *Computer Simulation of Personality: frontier of psychological theory* (pp. 277-298). New York: Wiley.
- Abildgren, K., & Kuchler, A. (2021). Revisiting the inflation perception conundrum. *Journal of Macroeconomics*, 67, 103264. <https://doi.org/10.1016/j.jmacro.2020.103264>
- Ackerman, D., & Perner, L. (2004). Did you hear what my friend paid! Examining the consequences of social comparisons of prices. *Advances in Consumer Research*, 31, 586–592. <http://hdl.handle.net/10211.2/2156>
- Adam, K. (2007). Experimental evidence on the persistence of output and inflation. *The Economic Journal*, 117(520), 603-636. <https://doi.org/10.1111/j.1468-0297.2007.02043.x>
- Adams, P. A., & Adams, J. K. (1958). Training in confidence-judgments. *The American Journal of Psychology*, 71(4), 747-751. <https://doi.org/10.2307/1420334>.
- Affleck-Graves, J., Davis, L. R., & Mendenhall, R. R. (1990). Forecasts of earnings per share: Possible sources of analyst superiority and bias. *Contemporary Accounting Research*, 6(2), 501-517. <https://doi.org/10.1111/j.1911-3846.1990.tb00771.x>
- Ahn, Y. B., & Tsuchiya, Y. (2022). Consumers' perceived and expected inflation in South Korea. *Applied Economics Letters*, 1-5. <https://doi.org/10.1080/13504851.2022.2070593>
- Alicke, M. D., & Govorun, O. (2013). The better-than-average effect. In M. D. Alicke, D. A., Dunning, & J. Krueger (Eds.), *The self in social judgment* (pp.85-106). Psychology Press.
- Almenberg, J., & Dreber, A. (2012). *Gender, stock market participation and financial literacy* (SSE/EFI Working Paper Series No. 737). Stockholm School of Economics. <http://dx.doi.org/10.2139/ssrn.1880909>
- Alpert, M., & Raiffa, H. (1982). A progress report on the training of probability advisors. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 294–305). Cambridge: Cambridge University Press.
- Amir, E., & Ganzach, Y. (1998). Overreaction and underreaction in analysts' forecasts. *Journal of Economic Behavior & Organization*, 37(3), 333-347. [https://doi.org/10.1016/S0167-2681\(98\)00092-4](https://doi.org/10.1016/S0167-2681(98)00092-4)
- Andersson, M., Gärling, T., Hedesström, M., & Biel, A. (2012). Effects on stock investments of information about short versus long price series. *Review of Behavioural Finance*, 4(2), 81-97. <https://doi.org/10.1108/19405971211284871>
- Andersson, P., & Rakow, T. (2007). Now you see it now you don't: The effectiveness of the recognition heuristic for selecting stock. *Judgment and Decision Making*, 2(1), 29-39. <http://journal.sjdm.org/jdm06162.pdf>.
- Andrade, P., Crump, R. K., Eusepi, S., & Moench, E. (2016). Fundamental disagreement. *Journal of Monetary Economics*, 83, 106-128. <https://doi.org/10.1016/j.jmoneco.2016.08.007>

- Andreassen, P. B. (1990). Judgmental extrapolation and market overreaction: On the use and disuse of news. *Journal of Behavioral Decision Making*, 3(3), 153-174.  
<https://doi.org/10.1002/bdm.3960030302>
- Andreassen, P. B., & Kraus, S. J. (1990). Judgmental extrapolation and the salience of change. *Journal of forecasting*, 9(4), 347-372. <https://doi.org/10.1002/for.3980090405>
- Ang, A., Bekaert, G., & Wei, M. (2007). Do Macro variables, asset markets or surveys forecast inflation better? *Journal of Monetary Economics*, 54(4), 1163–1212.  
<https://doi.org/10.1016/j.jmoneco.2006.04.006>
- Ang, S., & O'Connor, M. (1991). The effect of group interaction processes on performance in time series extrapolation. *International Journal of Forecasting*, 7(2), 141-149.  
[https://doi.org/10.1016/0169-2070\(91\)90048-Z](https://doi.org/10.1016/0169-2070(91)90048-Z)
- Angus-Leppan, P., & Fatseas, V. (1986). The forecasting accuracy of trainee accountants using judgemental and statistical techniques. *Accounting and Business Research*, 16(63), 179-188.  
<https://doi.org/10.1080/00014788.1986.9729316>
- Annett, J. (1969). *Feedback and human behaviour: The effects of knowledge of results, incentives and reinforcement on learning and performance*. London: Penguin Books.
- Anufriev, M., & Hommes, C. (2012). Evolutionary selection of individual expectations and aggregate outcomes in asset pricing experiments. *American Economic Journal: Microeconomics*, 4(4), 35-64. <http://dx.doi.org/10.1257/mic.4.4.35>.
- Anufriev, M., Hommes, C., & Makarewicz, T. (2019). Simple forecasting heuristics that make us smart: Evidence from different market experiments. *Journal of the European Economic Association*, 17(5), 1538-1584. <https://doi.org/10.1093/jeea/jvy028>
- Arioli, R., Bates, C., Dieden, H., Duca, I., Friz, R., Gayer, C., ... & Pavlova, I. (2017). *EU consumers' quantitative inflation perceptions and expectations: an evaluation* (No. 186). European Central Bank Occasional Paper. <http://hdl.handle.net/10419/162203>
- Arkes, H. R. (2001). Overconfidence in judgmental forecasting. In J. S. Armstrong (Ed.), *Principles of forecasting: a handbook for researchers and practitioners* (pp. 495-515). Springer, Boston, MA.
- Arkes, H. R., Dawes, R. M., & Christensen, C. (1986). Factors influencing the use of a decision rule in a probabilistic task. *Organizational behavior and human decision processes*, 37(1), 93-110.  
[https://doi.org/10.1016/0749-5978\(86\)90046-4](https://doi.org/10.1016/0749-5978(86)90046-4)
- Armantier, O., Boumahdi, F., Koşar, G., Somerville, J., Topa, G., Van der Klaauw, W., & Williams, J. C. (2022). *What Do Consumers Think Will Happen to Inflation?*. Liberty Street Economics 20220526, Federal Reserve Bank of New York.  
<https://ideas.repec.org/p/fip/fednls/94274.html>
- Armantier, O., Bruine de Bruin, W., Potter, S., Topa, G., Van der Klaauw, W., & Zafar, B. (2013). Measuring Inflation Expectations. *Annual Review of Economics* 5(1), 273-301.  
<https://doi.org/10.1146/annurev-economics-081512-141510>

- 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. <https://doi.org/10.1111/iere.12113>
- Armantier, O., Goldman, L., Koşar, G., Topa, G., Van der Klaauw, W., & Williams, J. C. (2022). *What Are Consumers' Inflation Expectations Telling Us Today?* Liberty Street Economics 20220214, Federal Reserve Bank of New York. <https://ideas.repec.org/p/fip/fednls/93706.html>
- Armantier, O., Koşar, G., Pomerantz, R., Skandalis, D., Smith, K., Topa, G., & Van der Klaauw, W. (2020). *Inflation expectations in times of COVID-19*. Federal Reserve Bank of New York. No. 20200513. <https://libertystreeteconomics.newyorkfed.org/2020/05/inflation-expectations-in-times-of-covid-19/>
- Armantier, O., Koşar, G., Pomerantz, R., Skandalis, D., Smith, K., Topa, G., & Van der Klaauw, W. (2021). How economic crises affect inflation beliefs: Evidence from the Covid-19 pandemic. *Journal of Economic Behavior & Organization*, 189, 443-469. <https://doi.org/10.1016/j.jebo.2021.04.036>
- Armantier, O., Nelson, S., Topa, G., Van der Klaauw, W., & Zafar, B. (2016). The price is right: Updating inflation expectations in a randomized price information experiment. *Review of Economics and Statistics*, 98(3), 503-523. [https://doi.org/10.1162/REST\\_a\\_00499](https://doi.org/10.1162/REST_a_00499)
- Armantier, O., Topa, G., van Der Klaauw, W., & Zafar, B. (2017). An overview of the survey of consumer expectations. *Economic Policy Review - Federal Reserve Bank of New York*, 23(2), 51-72. [https://www.newyorkfed.org/medialibrary/media/research/epr/2017/epr\\_2017\\_survey-consumer-expectations\\_armantier.pdf](https://www.newyorkfed.org/medialibrary/media/research/epr/2017/epr_2017_survey-consumer-expectations_armantier.pdf)
- Armstrong, J. S. (1983). Relative accuracy of judgemental and extrapolative methods in forecasting annual earnings. *Journal of Forecasting*, 2(4), 437-447. <https://doi.org/10.1002/for.3980020411>
- Armstrong, J. S. (1985). *Long-Range Forecasting* (2<sup>nd</sup> ed.) Wiley. [https://repository.upenn.edu/marketing\\_papers/211](https://repository.upenn.edu/marketing_papers/211)
- Armstrong, J. S. (2001). *Principles of forecasting: a handbook for researchers and practitioners*. Boston, MA: Kluwer Academic.
- Armstrong, J. S., Green, K. C., & Graefe, A. (2015). Golden rule of forecasting: Be conservative. *Journal of Business Research*, 68(8), 1717-1731. <https://doi.org/10.1016/j.jbusres.2015.03.031>
- Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, 41(3), 258. <https://doi.org/10.1037/h0055756>
- Ashby, F. G., & Maddox, W. T. (1998). Stimulus categorization. In M. H. Birnbaum (ed.), *Measurement, judgment and decision making* (pp. 251-301). Academic Press. <https://doi.org/10.1016/B978-012099975-0.50006-3>
- Ashiya, M. (2009). Strategic bias and professional affiliations of macroeconomic forecasters. *Journal of Forecasting*, 28(2), 120-130. <https://doi.org/10.1002/for.1095>

- Asquith, P., Mikhail, M. B., & Au, A. S. (2005). Information content of equity analyst reports. *Journal of financial economics*, 75(2), 245-282. <https://doi.org/10.1016/j.jfineco.2004.01.002>
- Assenza, T., Heemeijer, P., Hommes, C. H., & Massaro, D. (2013, January). *Individual expectations and aggregate macro behavior*. [Discussion paper]. Tinbergen Institute Discussion Paper 13-016/II. SSRN. <http://dx.doi.org/10.2139/ssrn.2200424>
- Atir, S., Rosenzweig, E., & Dunning, D. (2015). When knowledge knows no bounds: Self-perceived expertise predicts claims of impossible knowledge. *Psychological Science*, 26(8), 1295-1303. <https://doi.org/10.1177/0956797615588195>
- Ayton, P. (1992). On the competence and incompetence of experts. In F., Bolger, & G. Wright (Eds.), *Expertise and decision support* (pp. 77-105). Springer, Boston, MA.
- Ayton, P., Önkal, D., & McReynolds, L. (2011). Effects of ignorance and information on judgments and decisions. *Judgment and Decision Making*, 6(5), 381-391. <http://journal.sjdm.org/11/rh7/rh7.html>.
- Babad, E. (1987). Wishful thinking and objectivity among sports fans. *Social Behaviour*. 4(4), 231-240. <https://psycnet.apa.org/record/1989-24476-001>
- Badarinza, C., & Buchmann, M. (2009). *Inflation perceptions and expectations in the euro area: the role of news* (ECB Working Paper No. 1088). European Central Bank. <http://dx.doi.org/10.2139/ssrn.1457631>
- Baginski, S. P., Conrad, E. J., & Hassell, J. M. (1993). The effects of management forecast precision on equity pricing and on the assessment of earnings uncertainty. *Accounting Review*, 68(4), 913-927. <https://www.jstor.org/stable/248513>
- Balzer, W. K., Doherty, M. E. and O'Connor, R. Jr. (1989). Effects of cognitive feedback on performance. *Psychological Bulletin*, 106(3), 410-33. <https://doi.org/10.1037/0033-2909.106.3.410>.
- Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin*, 74(2), 81-99. <https://doi.org/10.1037/h0029531>
- Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgments. *Perception & Psychophysics*, 55(4), 412-428. <https://doi.org/10.3758/BF03205299>.
- Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies*, 21(2), 785-818. <https://doi.org/10.1093/rfs/hhm079>
- Barro, R. (1999). Reagan Vs Clinton: Who's the economic champ? *Bloomberg*, 22 February. <https://www.bloomberg.com/news/articles/1999-02-21/reagan-vs-dot-clinton-whos-the-economic-champ>
- Bartlett, F. C. (1932). *Remembering*. Cambridge university press.

- Bartos, J.A. (1969) The Assessment of Probability Distributions for Future Security Prices. [Unpublished PhD Thesis]. Indianan University.
- Bastounis, M., Leiser, D., & Roland-Lévy, C. (2004). Psychosocial variables involved in the construction of lay thinking about the economy: Results of a cross-national survey. *Journal of economic psychology*, 25(2), 263-278. [https://doi.org/10.1016/S0167-4870\(02\)00198-8](https://doi.org/10.1016/S0167-4870(02)00198-8)
- Batchelor, R. A. (1986). The psychophysics of inflation. *Journal of Economic Psychology*, 7(3), 269-290. [https://doi.org/10.1016/0167-4870\(86\)90021-8](https://doi.org/10.1016/0167-4870(86)90021-8)
- Batchelor, R. A. (2007). Bias in macroeconomic forecasts. *International Journal of Forecasting*, 23(2), 189-203. <https://doi.org/10.1016/j.ijforecast.2007.01.004>
- Batchelor, R. A., & Dua, P. (1990). Product differentiation in the economic forecasting industry. *International Journal of Forecasting*, 6(3), 311-316. [https://doi.org/10.1016/0169-2070\(90\)90058-J](https://doi.org/10.1016/0169-2070(90)90058-J)
- Batchelor, R. A., & Dua, P. (1995). Forecaster diversity and the benefits of combining forecasts. *Management Science*, 41(1), 68-75. <https://doi.org/10.1287/mnsc.41.1.68>
- Bates, J. M., & Gabor, A. (1986). Price perception in creeping inflation: Report on an enquiry. *Journal of Economic Psychology*, 7(3), 291-314. [https://doi.org/10.1016/0167-4870\(86\)90022-X](https://doi.org/10.1016/0167-4870(86)90022-X)
- Bayar, Y., Sezgin, H.F., Öztürk, Ö.F. & Şaşmaz, M. (2017). Impact of Financial Literacy on Personal Savings: A Research On Usak University Staff. *Journal of Knowledge Management Economics and Information Technology*, 6(6), 1-19. <https://www.researchgate.net/publication/323074914>
- Beach, L. R., & Solak, F. (1969). Subjective judgments of acceptable error. *Organizational behavior and human performance*, 4(3), 242-251. [https://doi.org/10.1016/0030-5073\(69\)90008-7](https://doi.org/10.1016/0030-5073(69)90008-7)
- Beer, J. S., & Hughes, B. L. (2010). Neural systems of social comparison and the “above-average” effect. *Neuroimage*, 49(3), 2671-2679. <https://doi.org/10.1016/j.neuroimage.2009.10.075>
- Behrend, H. (1977). Research into inflation and conceptions of earnings. *Journal of Occupational Psychology*, 50(3), 169-176. <https://doi.org/10.1111/j.2044-8325.1977.tb00373.x>
- Benson, P. G., & Önkal, D. (1992). The effects of feedback and training on the performance of probability forecasters. *International Journal of Forecasting*, 8(4), 559-573. [https://doi.org/10.1016/0169-2070\(92\)90066-I](https://doi.org/10.1016/0169-2070(92)90066-I)
- Berardi, M. (2009). Monetary policy with heterogeneous and misspecified expectations. *Journal of Money, Credit and Banking*, 41(1), 79-100. <https://doi.org/10.1111/j.1538-4616.2008.00188.x>
- Bergman, O., Ellingsen, T., Johannesson, M., & Svensson, C. (2010). Anchoring and cognitive ability. *Economics Letters*, 107(1), 66-68. <https://doi.org/10.1016/j.econlet.2009.12.028>
- Bernanke, B.S. (2004, April). *The economic outlook and monetary policy*. The Bond Market Association Annual Meeting at the *World Economy Laboratory Spring Conference* [Speech].

The Federal Reserve Board, New York, Washington. Retrieved from [www.federalreserve.gov/boarddocs/speeches/2004/200404222/default.htm](http://www.federalreserve.gov/boarddocs/speeches/2004/200404222/default.htm)

- Bernanke, B. S. (2007, July). *Inflation Expectations and Inflation Forecasting*. National Bureau of Economic Research Monetary Economics Workshop [Speech]. The Federal Reserve System, Cambridge, Massachusetts. Retrieved from <http://www.federalreserve.gov/newsevents/speech/Bernanke20070710a.htm>
- Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2018). Behavioral household finance. In *Handbook of Behavioral Economics: Applications and Foundations 1* (Vol. 1, pp. 177-276). North-Holland. <https://doi.org/10.1016/bs.hesbe.2018.07.004>
- Binder, C. (2017). Measuring uncertainty based on rounding: New method and application to inflation expectations. *Journal of Monetary Economics*, *90*, 1-12. <https://doi.org/10.1016/j.jmoneco.2017.06.001>
- Binder, C. (2020). Coronavirus fears and macroeconomic expectations. *Review of Economics and Statistics*, *102*(4), 721-730. [https://doi.org/10.1162/rest\\_a\\_00931](https://doi.org/10.1162/rest_a_00931)
- Binder, C., & Rodrigue, A. (2018). Household informedness and long-run inflation expectations: Experimental evidence. *Southern Economic Journal*, *85*(2), 580-598. <https://doi.org/10.1002/soej.12306>
- Birnbaum, M. H., Wong, R., & Wong, L. K. (1976). Combining information from sources that vary in credibility. *Memory & cognition*, *4*(3), 330-336. <https://doi.org/10.3758/BF03213185>
- Birnberg, J. G., & Slevin, D. P. (1976). A note on the use of confidence interval statements in financial reporting. *Journal of Accounting Research*, *14*(1), 153-157. <https://doi.org/10.2307/2490461>
- Blanchflower, D. G., & Mac Coille, C. (2009). *The formation of inflation expectations: an empirical model for the UK* (NBER Working Paper No. 15388). National Bureau of Economic Research. [https://www.nber.org/system/files/working\\_papers/w15388/w15388.pdf](https://www.nber.org/system/files/working_papers/w15388/w15388.pdf)
- Block, R. A., & Harper, D. R. (1991). Overconfidence in estimation: Testing the anchoring-and-adjustment hypothesis. *Organizational behavior and human decision processes*, *49*(2), 188-207. [https://doi.org/10.1016/0749-5978\(91\)90048-X](https://doi.org/10.1016/0749-5978(91)90048-X)
- Blomqvist, H. C. (1983). On the formation of inflationary expectations: Some empirical evidence from Finland, 1979–1980. *Journal of Economic Psychology*, *4*(4), 319-334. [https://doi.org/10.1016/0167-4870\(83\)90037-5](https://doi.org/10.1016/0167-4870(83)90037-5)
- Blomqvist, H. C. (1988). Uncertainty and predictive accuracy: an empirical study. *Journal of economic psychology*, *9*(4), 525-532. [https://doi.org/10.1016/0167-4870\(88\)90017-7](https://doi.org/10.1016/0167-4870(88)90017-7)
- Boduroglu, A. (2022). Judgmental Time Series Forecasting: A systematic analysis of graph format and trend type. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*. <https://escholarship.org/uc/item/0233x259>
- Boero, G., Smith, J., & Wallis, K. F. (2008). Uncertainty and disagreement in economic prediction: the Bank of England Survey of External Forecasters. *The Economic Journal*, *118*(530), 1107-1127. <https://doi.org/10.1111/j.1468-0297.2008.02162.x>

- Boje, D. M., & Murnighan, J. K. (1982). Group confidence pressures in iterative decisions. *Management Science*, 28(10), 1187-1196.  
<https://doi.org/10.1287/mnsc.28.10.1187>
- Bolger, F., & Harvey, N. (1993). Context-sensitive heuristics in statistical reasoning. *The Quarterly Journal of Experimental Psychology*, 46(4), 779-811.  
<https://doi.org/10.1080/14640749308401039>
- Bolger, F., & Harvey, N. (1995). Judging the probability that the next point in an observed time-series will be below, or above, a given value. *Journal of Forecasting*, 14(7), 597-607.  
<https://doi.org/10.1002/for.3980140705>
- Bolger, F., & Harvey, N. (1998). Heuristics and biases in judgmental forecasting. In G. Wright, and P. Goodwin (Eds.). *Fore-casting with judgment* (pp.113-137). Wiley, Chichester.
- Bolger, F., & Önkal-Atay, D. (2004). The effects of feedback on judgmental interval predictions. *international Journal of forecasting*, 20(1), 29-39.  
[https://doi.org/10.1016/S0169-2070\(03\)00009-8](https://doi.org/10.1016/S0169-2070(03)00009-8)
- Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational behavior and human decision processes*, 101(2), 127-151. <https://doi.org/10.1016/j.obhdp.2006.07.001>
- Borges, B., Goldstein, D. G., Ortmann, A., & Gigerenzer, G. (1999). Can ignorance beat the stock market? In G. Gigerenzer, P. M. Todd and the ABC Research Group (Eds.) *Simple heuristics that make us smart*, Oxford: Oxford University Press. (pp 59-72).
- Bornstein, B. H., & Zickafoose, D. J. (1999). "I know I know it, I know I saw it": The stability of the confidence-accuracy relationship across domains. *Journal of Experimental Psychology: Applied*, 5(1), 76-88. <https://doi.org/10.1037/1076-898X.5.1.76>
- Bottan, N. L., & Perez-Truglia, R. (2020). *Betting on the house: Subjective expectations and market choices* (Working paper No. w27412). National Bureau of Economic Research.  
<https://www.nber.org/papers/w27412>
- Brachinger, H. W. (2008). A new index of perceived inflation: Assumptions, method, and application to Germany. *Journal of Economic Psychology*, 29(4), 433-457.  
<https://doi.org/10.1016/j.joep.2008.04.004>
- Bradshaw, M. T., Drake, M. S., Myers, J. N., & Myers, L. A. (2012). A re-examination of analysts' superiority over time-series forecasts of annual earnings. *Review of Accounting Studies*, 17(4), 944-968. <https://doi.org/10.1007/s11142-012-9185-8>
- Branch, W. A. (2004). The theory of rationally heterogeneous expectations: evidence from survey data on inflation expectations. *The Economic Journal*, 114(497), 592-621.  
<https://doi.org/10.1111/j.1468-0297.2004.00233.x>
- Branch, W. A., & Evans, G. W. (2006). A simple recursive forecasting model. *Economics Letters*, 91(2), 158-166. <https://doi.org/10.1016/j.econlet.2005.09.005>

- Branch, W. A., & McGough, B. (2011). Business cycle amplification with heterogeneous expectations. *Economic Theory*, 47(2), 395–421. <https://doi.org/10.1007/s00199-010-0541-2>
- Brehmer, B. (1978). Response consistency in probabilistic inference tasks. *Organizational Behavior and Human Performance*, 22(1), 103-115. [https://doi.org/10.1016/0030-5073\(78\)90008-9](https://doi.org/10.1016/0030-5073(78)90008-9).
- Brehmer, B. (1980). In one word: Not from experience. *Acta Psychologica*, 45(1-3), 223-241. [https://doi.org/10.1016/0001-6918\(80\)90034-7](https://doi.org/10.1016/0001-6918(80)90034-7)
- Brehmer, A., & Brehmer, B. (1988). What have we learned about human judgment from thirty years of policy capturing?. In *Advances in psychology* (Vol. 54, pp. 75-114). North-Holland. [https://doi.org/10.1016/S0166-4115\(08\)62171-8](https://doi.org/10.1016/S0166-4115(08)62171-8)
- Brenner, L. A. (1995). *A stochastic model of the calibration of subjective probabilities*. [Unpublished dissertation]. Stanford University. <https://www.proquest.com/docview/304220153?pq-origsite=gscholar&fromopenview=true>
- Brenner, L. A. (2003). A random support model of the calibration of subjective probabilities. *Organizational Behavior and Human Decision Processes*, 90(1), 87-110. [https://doi.org/10.1016/S0749-5978\(03\)00004-9](https://doi.org/10.1016/S0749-5978(03)00004-9)
- Brenner, L., Griffin, D., & Koehler, D. J. (2005). Modeling patterns of probability calibration with random support theory: Diagnosing case-based judgment. *Organizational Behavior and Human Decision Processes*, 97(1), 64-81. <https://doi.org/10.1016/j.obhdp.2005.02.002>
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1), 1-3. [https://doi.org/10.1175/1520-0493\(1950\)078<0001:VOFEIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2)
- Brischetto, A., & De Brouwer, G. (1999, March). Householders' Inflation Expectations [Discussion Papers]. *Reserve Bank of Australia Research*. <https://www.rba.gov.au/publications/rdp/1999/1999-03/introduction.html>
- Brown, L. D. (1993). Earnings forecasting research: its implications for capital markets research. *International journal of forecasting*, 9(3), 295-320. [https://doi.org/10.1016/0169-2070\(93\)90023-G](https://doi.org/10.1016/0169-2070(93)90023-G)
- Brown, L. D. (1996). Analyst forecasting errors and their implications for security analysis: An alternative perspective. *Financial Analysts Journal*, 52(1), 40-47. <https://doi.org/10.2469/faj.v52.n1.1965>
- Brown, L. D., Hagerman, R. L., Griffin, P. A., & Zmijewski, M. E. (1987). Security analyst superiority relative to univariate time-series models in forecasting quarterly earnings. *Journal of accounting and Economics*, 9(1), 61-87. [https://doi.org/10.1016/0165-4101\(87\)90017-6](https://doi.org/10.1016/0165-4101(87)90017-6)
- Brown, L. D., & Zmijewski, M. E. (1987). The effect of labor strikes on security analysts' forecast superiority and on the association between risk-adjusted stock returns and unexpected earnings. *Contemporary Accounting Research*, 4(1), 61-75. <https://doi.org/10.1111/j.1911-3846.1987.tb00655.x>
- Brown, S. D., & Steyvers, M. (2009). Detecting and predicting changes. *Cognitive Psychology*, 58(1), 49 – 67. <https://doi.org/10.1016/j.cogpsych.2008.09.002>.

- Bruine de Bruin, W., Manski, C. F., Topa, G., & Van der Klaauw, W. (2011). Measuring consumer uncertainty about future inflation. *Journal of Applied Econometrics*, 26(3), 454-478. <https://doi.org/10.1002/jae.1239>
- Bruine de Bruin, W., Potter, S., Rich, R. W., Topa, G., & Van der Klaauw, W. (2010). Improving Survey Measures of Household Inflation Expectations. *Economics and Finance* 16(7). [https://www.newyorkfed.org/research/current\\_issues/ci16-7.html](https://www.newyorkfed.org/research/current_issues/ci16-7.html)
- Bruine de Bruin, W., Van der Klaauw, W., Downs, J. S., Fischhoff, B., Topa, G., & Armantier, O. (2010). Expectations of inflation: The role of demographic variables, expectation formation, and financial literacy. *Journal of Consumer Affairs*, 44(2), 381-402. <https://doi.org/10.1111/j.1745-6606.2010.01174.x>
- Bruine de Bruin, W., Van der Klaauw, W., & Topa, G. (2011). Expectations of inflation: The biasing effect of thoughts about specific prices. *Journal of Economic Psychology*, 32(5), 834-845. <https://doi.org/10.1016/j.joep.2011.07.002>
- Bruine de Bruin, W. B., Van der Klaauw, W., Topa, G., Downs, J. S., Fischhoff, B., & Armantier, O. (2012). The effect of question wording on consumers' reported inflation expectations. *Journal of Economic Psychology*, 33(4), 749-757. <https://doi.org/10.1016/j.joep.2012.02.001>
- Bruine de Bruin, W. B., Van der Klaauw, W., van Rooij, M., Teppa, F., & de Vos, K. (2017). Measuring expectations of inflation: Effects of survey mode, wording, and opportunities to revise. *Journal of Economic Psychology*, 59, 45-58. <https://doi.org/10.1016/j.joep.2017.01.011>
- Bryan, M. F., & Venkatu, G. (2001a). *The demographics of inflation opinion surveys*. Federal Reserve Bank of Cleveland, *Economic Commentary*, 10.15.2001. <https://core.ac.uk/download/pdf/6888515.pdf>
- Bryan, M. F., & Venkatu, G. (2001b). *The curiously different inflation perspectives of men and women*. Federal Reserve Bank of Cleveland, *Economic Commentary*, 11.01.2001. <https://core.ac.uk/download/pdf/6893382.pdf>
- Buchheit, S., Dalton, D., Downen, T., & Pippin, S. (2012). Outcome feedback, incentives, and performance: Evidence from a relatively complex forecasting task. *Behavioral Research in Accounting*, 24(2), 1-20. <https://doi.org/10.2308/bria-50151>
- Budescu, D. V., & Du, N. (2007). Coherence and consistency of investors' probability judgments. *Management science*, 53(11), 1731-1744. <https://doi.org/10.1287/mnsc.1070.0727>
- Budescu, D. V., & Yu, H. T. (2007). Aggregation of opinions based on correlated cues and advisors. *Journal of Behavioral Decision Making*, 20(2), 153-177. <https://doi.org/10.1002/bdm.547>
- 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. <https://doi.org/10.1111/jmcb.12144>

- Bunn, D. (1992). Synthesis of expert judgment and statistical forecasting models for decision support. In *Expertise and decision support* (pp. 251-268). Springer, Boston, MA. [https://doi.org/10.1007/978-0-585-34290-0\\_12](https://doi.org/10.1007/978-0-585-34290-0_12)
- Bunn, D. W., & Salo, A. A. (1996). Adjustment of forecasts with model consistent expectations. *International Journal of Forecasting*, *12*(1), 163-170. [https://doi.org/10.1016/0169-2070\(95\)00660-5](https://doi.org/10.1016/0169-2070(95)00660-5)
- Bureau of Labor Statistics (2010a). *Consumer Price Index News Release August 2010*. Washington CD: Department of Labor. [https://www.bls.gov/news.release/archives/cpi\\_09172010.htm](https://www.bls.gov/news.release/archives/cpi_09172010.htm)
- Bureau of Labor Statistics (2010b). *Consumer Price Index News Release March 2010*. Washington CD: Department of Labor. [https://www.bls.gov/news.release/archives/cpi\\_04142010.htm](https://www.bls.gov/news.release/archives/cpi_04142010.htm)
- Camerer, C. F. (1992). The rationality of prices and volume in experimental markets. *Organizational Behavior and Human Decision Processes*, *51*(2), 237-272. [https://doi.org/10.1016/0749-5978\(92\)90013-W](https://doi.org/10.1016/0749-5978(92)90013-W)
- Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, *19*(1), 7-42. <https://doi.org/10.1023/A:1007850605129>.
- Campbell, S. D., & Sharpe, S. A. (2009). Anchoring bias in consensus forecasts and its effect on market prices. *Journal of Financial and Quantitative Analysis*, *44*(2), 369-390. <https://doi.org/10.1017/S0022109009090127>
- Capistrán, C. & Timmermann, A. (2009). Disagreement and biases in inflation expectations. *Journal of Money, Credit and Banking*, *41*(2-3), 365-396. <https://doi.org/10.1111/j.1538-4616.2009.00209.x>.
- Carbone, R., Andersen, A., Corriveau, Y., & Corson, P. P. (1983). Comparing for different time series methods the value of technical expertise individualized analysis, and judgmental adjustment. *Management Science*, *29*(5), 559-566. <https://doi.org/10.1287/mnsc.29.5.559>
- Carbone, R., & Gorr, W. L. (1985). Accuracy of judgmental forecasting of time series. *Decision Sciences*, *16*(2), 153-160. <https://doi.org/10.1111/j.1540-5915.1985.tb01480.x>
- Carlson, J. A., & Parkin, M. (1975). Inflation expectations. *Economica*, *42*(166), 123-138. <https://doi.org/10.2307/2553588>
- Carrillo, P. E., & Emran, S. M. (2012). Public information and inflation expectations: Microeconomic evidence from a natural experiment. *Review of Economics and Statistics*, *94*(4), 860-877. [https://doi.org/10.1162/REST\\_a\\_00213](https://doi.org/10.1162/REST_a_00213)
- Carroll, C. (2003). Macroeconomic Expectations of Households and Professional Forecasters. *The Quarterly Journal of Economics*, *118*(1), 269-298. <https://doi.org/10.1162/00335530360535207>
- 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. <https://doi.org/10.1257/mac.20150147>

- Cesarini, D., Sandewall, Ö., & Johannesson, M. (2006). Confidence interval estimation tasks and the economics of overconfidence. *Journal of Economic Behavior & Organization*, *61*(3), 453-470. <https://doi.org/10.1016/j.jebo.2004.10.010>
- Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of personality and social psychology*, *39*(5), 752. <https://doi.org/10.1037/0022-3514.39.5.752>
- Chang, W., Chen, E., Mellers, B., & Tetlock, P. (2016). Developing expert political judgment: The impact of training and practice on judgmental accuracy in geopolitical forecasting tournaments. *Judgment & Decision Making*, *11*(5), 509-526. <https://www.sas.upenn.edu/~baron/journal/16/16511/jdm16511.pdf>
- Cheek, N. N., & Norem, J. K. (2017). Holistic thinkers anchor less: Exploring the roles of self-construal and thinking styles in anchoring susceptibility. *Personality and Individual Differences*, *115*, 174-176. <https://doi.org/10.1016/j.paid.2016.01.034>
- Chen, K. H., & Summers, E. L. (1981). A study of reporting probabilistic accounting figures. *Accounting, Organizations and Society*, *6*(1), 1-15. [https://doi.org/10.1016/0361-3682\(81\)90018-0](https://doi.org/10.1016/0361-3682(81)90018-0)
- Chen, Q., Costantini, M., & Deschamps, B. (2016). How accurate are professional forecasts in Asia? Evidence from ten countries. *International Journal of Forecasting*, *32*(1), 154-167. <https://doi.org/10.1016/j.ijforecast.2015.05.004>
- Chiarella, C., He, X. Z., & Zwinkels, R. C. (2014). Heterogeneous expectations in asset pricing: Empirical evidence from the S&P500. *Journal of Economic Behavior & Organization*, *105*, 1-16. <https://doi.org/10.1016/j.jebo.2014.03.003>
- Ciccione, L., & Dehaene, S. (2021). Can humans perform mental regression on a graph? Accuracy and bias in the perception of scatterplots. *Cognitive Psychology*, *128*, 101406. <https://doi.org/10.1016/j.cogpsych.2021.101406>
- Ciccione, L., Sablé-Meyer, M., & Dehaene, S. (2022). Analyzing the misperception of exponential growth in graphs. *Cognition*, *225*, 105112. <https://doi.org/10.1016/j.cognition.2022.105112>
- Cipriani, M., & Guarino, A. (2008). Herd behavior and contagion in financial markets. *The BE Journal of Theoretical Economics*, *8*(1), Article 24. <https://doi.org/10.2202/1935-1704.1390>
- Clements, M. P. (1995). Rationality and the role of judgement in macroeconomic forecasting. *The Economic Journal*, *105*(429), 410-420. <https://doi.org/10.2307/2235500>
- Clement, M. B. (1999). Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter? *Journal of Accounting and Economics*, *27*(3), 285-303. [https://doi.org/10.1016/S0165-4101\(99\)00013-0](https://doi.org/10.1016/S0165-4101(99)00013-0)
- Clements, M. P. (2010). Explanations of the inconsistencies in survey respondents' forecasts. *European Economic Review*, *54*(4), 536-549. <https://doi.org/10.1016/j.euroecorev.2009.10.003>

- Clements, M. P. (2011). An empirical investigation of the effects of rounding on the SPF probabilities of decline and output growth histograms. *Journal of Money, Credit and Banking*, 43(1), 207-220.  
<https://doi.org/10.1111/j.1538-4616.2010.00371.x>
- Clements, M. P. (2014). Forecast uncertainty—ex ante and ex post: US inflation and output growth. *Journal of Business & Economic Statistics*, 32(2), 206-216.  
<https://doi.org/10.1080/07350015.2013.859618>
- Clements, M. P. (2021). Rounding behaviour of professional macro-forecasters. *International Journal of Forecasting*, 37(4), 1614-1631. <https://doi.org/10.1016/j.ijforecast.2021.03.003>
- Clements, M. P., & Taylor, N. (2003). Evaluating interval forecasts of high-frequency financial data. *Journal of Applied Econometrics*, 18(4), 445-456. <https://doi.org/10.1002/jae.703>
- Cleveland, W. S., McGill, M. E., & McGill, R. (1988). The shape parameter of a two-variable graph. *Journal of the American Statistical Association*, 83(402), 289-300.  
<https://www.tandfonline.com/doi/abs/10.1080/01621459.1988.10478598>
- Cohen, J. & Dearnley, E. J. (1962). Skill and judgment of footballers in attempting to score goals: A study of psychological probability. *British Journal of Psychology*, 53(1), 71-88.  
<https://doi.org/10.1111/j.2044-8295.1962.tb00815.x>
- Cohen, J., Dearnaley, E. J., & Hansel, C. E. M. (1956). Risk and hazard: Influence of training on the performance of bus drivers. *Journal of the Operational Research Society*, 7(3), 67-82.  
<https://doi.org/10.1057/jors.1956.13>
- Coibion, O., D. Georganakos, Y. Gorodnichenko, and M. van Rooij (2019). *How does consumption respond to news about inflation? Field evidence from a randomized control trial* (NBER Working Paper No. 26106). National Bureau of Economic Research.  
<https://www.nber.org/papers/w26106>
- Coibion, O., & Gorodnichenko, Y. (2012). What can survey forecasts tell us about information rigidities? *Journal of Political Economy*, 120(1), 116-159. <https://doi.org/10.1086/665662>
- Coibion, O., & Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, 105(8), 2644-2678.  
<http://dx.doi.org/10.1257/aer.20110306>
- Coibion, O., Gorodnichenko, Y., Kumar, S., & Pedemonte, M. (2020). Inflation expectations as a policy tool? *Journal of International Economics*, 124, 103297.  
<https://doi.org/10.1016/j.jinteco.2020.103297>
- Coibion, O., Gorodnichenko, Y., & Weber, M. (2020). *The cost of the Covid-19 crisis: Lockdowns, macroeconomic expectations, and consumer spending* (NEBR Working Paper No. 27141). National Bureau of Economic Research. <https://www.nber.org/papers/w27141>
- Cole, S. J., & Milani, F. (2021). Heterogeneity in individual expectations, sentiment, and constant-gain learning. *Journal of Economic Behavior & Organization*, 188, 627-650.  
<https://doi.org/10.1016/j.jebo.2021.05.034>

- Conrad, C., Enders, Z., & Glas, A. (2022). The role of information and experience for households' inflation expectations. *European Economic Review*, *143*, 104015. <https://doi.org/10.1016/j.euroecorev.2021.104015>
- Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and applications*. Academic press.
- Corduas, M. (2022). Gender differences in the perception of inflation. *Journal of Economic Psychology*, *90*, 102522. <https://doi.org/10.1016/j.joep.2022.102522>
- Cornand, C., & Hubert, P. (2022). Information frictions across various types of inflation expectations. *European Economic Review*, *46*, 104175. <https://doi.org/10.1016/j.euroecorev.2022.104175>
- Coroneo, L., Iacone, F., & Profumo, F. (2019). *A real-time density forecast evaluation of the ECB Survey of Professional Forecasters* [Discussion Paper]. University of York. <https://ideas.repec.org/p/yor/yorken/19-14.html>
- Costello, F., & Watts, P. (2014). Surprisingly rational: Probability theory plus noise explains biases in judgment. *Psychological Review*, *121*(3), 463–480. <https://doi.org/10.1037/a0037010>
- Cunningham, R., Desroches, B., & Santor, E. (2010). Inflation expectations and the conduct of monetary policy: A review of recent evidence and experience. *Bank of Canada Review*, *Spring*, 13-25. <https://www.bankofcanada.ca/wp-content/uploads/2010/06/cunningham.pdf>
- Curtin, R. (1996). Procedure to estimate price expectations. [Manuscript]. *University of Michigan Survey Research Center*. <https://data.sca.isr.umich.edu/fetchdoc.php?docid=24772>
- Curtin, R. (2006, February). *Inflation expectations: Theoretical models and empirical tests*. The National Bank of Poland Workshop on The Role of Inflation Expectations in Modelling and Monetary Policy Making. [Conference paper]. National Bank of Poland, Warsaw, Poland. <https://www.nbp.pl/konferencje/bbm/curtin.pdf>
- Curtin, R. (2009). Inflation expectations and empirical tests: Theoretical models and empirical tests. In *Inflation Expectations* (pp. 52-79). Routledge.
- Czaczkes, B., & Ganzach, Y. (1996). The natural selection of prediction heuristics: Anchoring and adjustment versus representativeness. *Journal of Behavioral Decision Making*, *9*(2), 125-139. [https://doi.org/10.1002/\(SICI\)1099-0771\(199606\)9:2<125::AID-BDM221>3.0.CO;2-7](https://doi.org/10.1002/(SICI)1099-0771(199606)9:2<125::AID-BDM221>3.0.CO;2-7)
- Czerwonka, M. (2017). Anchoring and overconfidence: The influence of culture and cognitive abilities. *International Journal of Management and Economics*, *53*(3), 48-66. [https://ssl-kolegia.sgh.waw.pl/pl/KGS/publikacje/Documents/srodek\\_IJME\\_53\(3\).pdf](https://ssl-kolegia.sgh.waw.pl/pl/KGS/publikacje/Documents/srodek_IJME_53(3).pdf)
- D'Acunto, F., Hoang, D., Paloviita, M., & Weber, M. (2019). Cognitive abilities and inflation expectations. *American Economic Association Papers and Proceedings*, *109*, 562-66. <https://doi.org/10.1257/pandp.20191050>
- D'Acunto, F., Hoang, D., & Weber, M. (2015). *Inflation expectations and consumption expenditure* (Chicago Booth Global Markets Working Paper Series). National Bureau of Economic Research. [https://conference.nber.org/confer/2015/MEF15/D'Acunto\\_Hoang\\_Weber.pdf](https://conference.nber.org/confer/2015/MEF15/D'Acunto_Hoang_Weber.pdf)

- D'Acunto, F., Hoang, D., & Weber, M. (2018). Unconventional fiscal policy. *American Economic Association Papers and Proceedings*, 108, 519-23.  
<https://www.aeaweb.org/articles?id=10.1257/pandp.20181061>
- D'Acunto, F., Hoang, D., & Weber, M. (2022). Managing households' expectations with unconventional policies. *The Review of Financial Studies*, 35(4), 1597-1642.  
<https://doi.org/10.1093/rfs/hhab083>
- 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.  
<https://doi.org/10.1086/713192>
- D'Acunto, F., Malmendier, U., & Weber, M. (2021). Gender roles produce divergent economic expectations. *Proceedings of the National Academy of Sciences*, 118(21), e2008534118.  
<https://doi.org/10.1073/pnas.2008534118>
- D'Acunto, F., Malmendier, U., & Weber, M. (2022). *What Do the Data Tell Us About Inflation Expectations?* (NBER Working paper No. w29825). National Bureau of Economic Research.  
[https://www.nber.org/system/files/working\\_papers/w29825/w29825.pdf](https://www.nber.org/system/files/working_papers/w29825/w29825.pdf)
- Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, 44(1), 20. <https://doi.org/10.1037/0022-3514.44.1.20>
- Das, A., Lahiri, K., & Zhao, Y. (2016). *Measuring inflationary expectations from cross-sectional surveys: Households vs. experts* [Paper presentation]. The 21st Federal Forecasters Conference, Washington D.C., US.  
<https://cer.columbian.gwu.edu/sites/g/files/zaxdzs4246/files/downloads/FFC2015.pdf>
- Das, A., Lahiri, K., & Zhao, Y. (2019). Inflation expectations in India: Learning from household tendency surveys. *International Journal of Forecasting*, 35(3), 980-993.  
<https://doi.org/10.1016/j.ijforecast.2019.03.007>
- Davydenko, A., & Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts. *International Journal of Forecasting*, 29(3), 510-522. <https://doi.org/10.1016/j.ijforecast.2012.09.002>
- Dawid, H., & Gatti, D. D. (2018). Agent-based macroeconomics. In Hommes, C., & LeBaron, B. (Eds.). *Handbook of Computational Economics: Heterogeneous Agent Modelling* (pp. 63-156). Elsevier.
- Dawson, N. V., Connors, A. F., Speroff, T., Kemka, A., Shaw, P., & Arkes, H. R. (1993). Hemodynamic Assessment in Managing the Critically Ill: Is Physician Confidence Warranted? *Medical Decision Making*, 13(3), 258-266. <https://doi.org/10.1177/0272989X9301300314>
- Deane, D. H., Hammond, K. R., & Summers, D. A. (1972). Acquisition and application of knowledge in complex inference tasks. *Journal of Experimental Psychology*, 92(1), 20-26.  
<https://doi.org/10.1037/h0032162>

- Deaves, R., Lüders, E., & Schröder, M. (2010). The dynamics of overconfidence: Evidence from stock market forecasters. *Journal of Economic Behavior & Organization*, 75(3), 402-412. <https://doi.org/10.1016/j.jebo.2010.05.001>
- De Bondt, W. P. (1993). Betting on trends: Intuitive forecasts of financial risk and return. *International Journal of forecasting*, 9(3), 355-371. [https://doi.org/10.1016/0169-2070\(93\)90030-Q](https://doi.org/10.1016/0169-2070(93)90030-Q)
- De Bondt, W. F., & Thaler, R. (1985). Does the stock market overreact? *The Journal of finance*, 40(3), 793-805. <https://doi.org/10.1111/j.1540-6261.1985.tb05004.x>
- De Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. *International journal of forecasting*, 22(3), 443-473. <https://doi.org/10.1016/j.ijforecast.2006.01.001>
- De Grauwe, P. (2012). *Lectures on behavioral macroeconomics*. Princeton University Press.
- Deese, J., & Kaufman, R. A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of Experimental Psychology*, 54(3), 180-187. <https://doi.org/10.1037/h0040536>
- Del Giovane, P. D., & Sabbatini, R. (2008). Perceived and measured inflation after the launch of the euro: explaining the gap in Italy. In: Del Giovane, P., Sabbatini, R. (eds). *The Euro, Inflation and Consumer's Perceptions* (pp. 13-49). Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-540-78370-1\\_2](https://doi.org/10.1007/978-3-540-78370-1_2)
- Del Missier, F., Ranyard, R., & Bonini, N. (2016). Perceived inflation: The role of product accessibility and attitudes towards inflation. *Journal of Economic Psychology*, 56, 97-106. <https://doi.org/10.1016/j.joep.2016.05.008>
- Del Missier, F., Ranyard, R., & Pietroni, D. (2008). Interazioni virtuose tra psicologia della decisione ed economia comportamentale: il caso dell'inflazione percepita. *Sistemi intelligenti*, 20(1), 39-56. <https://www.rivisteweb.it/doi/10.1422/26716>
- DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic literature*, 47(2), 315-72. <https://www.aeaweb.org/articles?id=10.1257/jel.47.2.315>
- Den Butter, F. A. G. (1992). Scope and limitations of statistical methods for macroeconomic forecasting. *Kwantitatieve Methoden*, 13(40), 5-21. <https://www.vvsor.nl/wp-content/uploads/2020/06/KM1992040001.pdf>
- Denrell, J., & Fang, C. (2010). Predicting the next big thing: Success as a signal of poor judgment. *Management Science*, 56(10), 1653-1667. <https://doi.org/10.1287/mnsc.1100.1220>
- Dickson, G. W., DeSanctis, G., & McBride, D. J. (1986). Understanding the effectiveness of computer graphics for decision support: a cumulative experimental approach. *Communications of the ACM*, 29(1), 40-47. <https://doi.org/10.1145/5465.5469>
- Dickson, P. R., & Sawyer, A. G. (1990). The price knowledge and search of supermarket shoppers. *Journal of marketing*, 54(3), 42-53. <https://doi.org/10.1177/002224299005400304>

- Dietrich, A. M., Knotek II, E. S., Myrseth, K. O. R., Rich, R. W., Schoenle, R. S. & Weber, W. (2022). *Greater Than the Sum of the Parts: Aggregate vs. Aggregated Inflation Expectations* (Working Paper No. 22-20). Federal Reserve Bank of Cleveland.  
<https://doi.org/10.26509/frbc-wp-202220>
- Dietrich, A. M., Kuester, K., Müller, G. J., & Schoenle, R. (2022). News and uncertainty about COVID-19: Survey evidence and short-run economic impact. *Journal of monetary economics*, 129, 35-51. <https://doi.org/10.1016/j.jmoneco.2022.02.004>
- Diebold, F. X., Schorfheide, F., & Shin, M. (2017). Real-time forecast evaluation of DSGE models with stochastic volatility. *Journal of Econometrics*, 201(2), 322-332.  
<https://doi.org/10.1016/j.jeconom.2017.08.011>
- Diebold, F.X., Tay, A. and Wallis, K. (1999). Evaluating Density Forecasts of Inflation: The Survey of Professional Forecasters. In R. Engle and H. White (eds.), *Festschrift in Honor of C.W.J. Granger* (pp. 76-90). Oxford: Oxford University Press.  
<https://www.sas.upenn.edu/~fdiebold/papers/paper20/dtw-f.html>
- Dixon, R., Griffiths, W., & Lim, G. C. (2014). Lay people's models of the economy: A study based on surveys of consumer sentiments. *Journal of Economic Psychology*, 44, 13-20.  
<https://doi.org/10.1016/j.joep.2014.06.001>
- Doherty, M. E., & Balzer, W. K. (1988). Cognitive feedback. In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 163-197). Amsterdam: North-Holland.
- Dohmen, T., Falk, A., Huffman, D., Marklein, F., & Sunde, U. (2009). Biased probability judgment: Evidence of incidence and relationship to economic outcomes from a representative sample. *Journal of Economic Behavior & Organization*, 72(3), 903-915.  
<https://doi.org/10.1016/j.jebo.2009.07.014>
- Dominitz, J., & Manski, C. F. (1997). Using expectations data to study subjective income expectations. *Journal of the American Statistical Association*, 92(439), 855-867.  
<https://doi.org/10.1080/01621459.1997.10474041>
- Donihue, M. R. (1993). Evaluating the role judgment plays in forecast accuracy. *Journal of Forecasting*, 12(2), 81-92. <https://doi.org/10.1002/for.3980120203>
- Döpke, J., Dovern, J., Fritsche, U., & Slacalek, J. (2008). The dynamics of european inflation expectations. *The BE Journal of Macroeconomics*, 8(1), Article 12.  
<https://doi.org/10.2202/1935-1690.1540>
- Döpke, J., & Fritsche, U. (2006). When do forecasters disagree? An assessment of German growth and inflation forecast dispersion. *International Journal of Forecasting*, 22(1), 125-135.  
<https://doi.org/10.1016/j.ijforecast.2005.05.006>
- Dovern, J., & Weisser, J. (2011). Accuracy, unbiasedness and efficiency of professional macroeconomic forecasts: An empirical comparison for the G7. *International Journal of Forecasting*, 27(2), 452-465. <https://doi.org/10.1016/j.ijforecast.2010.05.016>

- Dräger, L. (2015). Inflation perceptions and expectations in Sweden—Are media reports the missing link? *Oxford Bulletin of Economics and Statistics*, 77(5), 681-700.  
<https://doi.org/10.1111/obes.12078>
- Dräger, L., Lamla, M. J., & Pfajfar, D. (2016). Are survey expectations theory-consistent? The role of central bank communication and news. *European Economic Review*, 85, 84-111.  
<https://doi.org/10.1016/j.euroecorev.2016.01.010>
- Dräger, L., Menz, J. O., & Fritsche, U. (2014). Perceived inflation under loss aversion. *Applied Economics*, 46(3), 282-293. <https://doi.org/10.1080/00036846.2013.844328>
- Drakos, K., Konstantinou, P. T., & Thoma, F. A. (2020). Inflation uncertainty and inflation expectations: Micro-level evidence from the Eurozone. *The Journal of Economic Asymmetries*, 21, e00150. <https://doi.org/10.1016/j.jeca.2019.e00150>
- Du, N., Budescu, D. V., Shelly, M. K., & Omer, T. C. (2011). The appeal of vague financial forecasts. *Organizational Behavior and Human Decision Processes*, 114(2), 179-189.  
<https://doi.org/10.1016/j.obhdp.2010.10.005>
- Duffy, D., & Lunn, P. (2008). *The misperception of inflation by Irish consumers* (Working Paper No. 261). The Economic and Social Research Institute. <http://hdl.handle.net/10419/50170>
- Easaw, J., Golinelli, R., & Malgarini, M. (2013). What determines households' inflation expectations? Theory and evidence from a household survey. *European Economic Review*, 61, 1-13.  
<https://doi.org/10.1016/j.euroecorev.2013.02.009>
- Easterwood, J. C., & Nutt, S. R. (1999). Inefficiency in analysts' earnings forecasts: Systematic misreaction or systematic optimism? *The Journal of Finance*, 54(5), 1777-1797.  
<https://doi.org/10.1111/0022-1082.00166>
- Ebrahimi E., Igan D., & Peria S.M. (2020). The Impact of COVID-19 on Inflation: Potential Drivers and Dynamics (Special Notes Series on COVID-19). International monetary fund.  
<https://www.imf.org/-/media/Files/Publications/covid19-special-notes/en-special-series-on-covid-19-the-impact-of-covid-19-on-inflation-potential-drivers-and-dynamics.ashx>
- Edmundson, R. H. (1990). Decomposition; a strategy for judgemental forecasting. *Journal of Forecasting*, 9(4), 305-314. <https://doi.org/10.1002/for.3980090403>
- Edmundson, B., Lawrence, M., & O'Connor, M. (1988). The use of non-time series information in sales forecasting: A case study. *Journal of Forecasting*, 7(3), 201-211. <https://doi.org/10.1002/for.3980070305>
- Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of personality and social psychology*, 71(1), 5-24. <https://psycnet.apa.org/buy/1996-01782-001>
- Eggleton, I. R. (1982). Intuitive time-series extrapolation. *Journal of Accounting Research*, 20(1), 68-102. <https://doi.org/10.2307/2490763>
- Ehrbeck, T., & Waldmann, R. (1996). Why are professional forecasters biased? Agency versus behavioral explanations. *The Quarterly Journal of Economics*, 111(1), 21-40.  
<https://doi.org/10.2307/2946656>

- Ehrmann, M., D. Pfajfar, and E. Santoro (2017). Consumers' attitudes and their inflation expectations. *International Journal of Central Banking* 13(1), 225–259.  
<https://www.ijcb.org/journal/ijcb17q0a6.htm>
- El-Shagi, M., Giesen, S., & Jung, A. (2012, February). *Does Central Bank Staff Beat Private Forecasters?* [Discussion paper]. IWH Discussion Papers 5/2012. Halle Institute for Economic Research (IWH).  
[https://www.econstor.eu/bitstream/10419/79925/1/Vfs\\_2013\\_pid\\_125.pdf](https://www.econstor.eu/bitstream/10419/79925/1/Vfs_2013_pid_125.pdf)
- Engelberg, J., Manski, C. F., & Williams, J. (2009). Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business and Economic Statistics*, 27(1), 30-41. <https://doi.org/10.1198/jbes.2009.0003>
- Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological science*, 17(4), 311-318. <https://doi.org/10.1111/j.1467-9280.2006.01704.x>
- Erdfelder, E., Brandt, M., & Bröder, A. (2007). Recollection biases in hindsight judgments. *Social Cognition*, 25(1), 114–131. <https://doi.org/10.1521/soco.2007.25.1.114>
- Eroglu, C., & Croxton, K. L. (2010). Biases in judgmental adjustments of statistical forecasts: The role of individual differences. *International Journal of Forecasting*, 26(1), 116-133.  
<https://doi.org/10.1016/j.ijforecast.2009.02.005>
- Eubanks, D. L., Read, D., & Grushka-Cockayne, Y. (2015). Biases as constraints on planning performance. In *The Psychology of Planning in Organizations* (pp. 245-257). Routledge.
- Evans, G. W., Honkapohja, S., & Marimon, R. (2001). Convergence in monetary inflation models with heterogeneous learning rules. *Macroeconomic Dynamics*, 5(1), 1-31.  
<https://doi.org/10.1017/S1365100501018016>
- Evans, J. St. B. T. (2008). Dual processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255-278.  
<https://doi.org/10.1146/annurev.psych.59.103006.093629>
- Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, 8(3), 223-241.  
<https://doi.org/10.1177/1745691612460685>
- Fahimnia, B., Pournader, M., Siemsen, E., Bendoly, E., & Wang, C. (2019). Behavioral operations and supply chain management—a review and literature mapping. *Decision Sciences*, 50(6), 1127-1183. <https://doi.org/10.1111/deci.12369>
- Fan, D. K., Lau, K. N., & Leung, P. L. (1996). Combining ordinal forecasts with an application in a financial market. *Journal of Forecasting*, 15(1), 37-48. [https://doi.org/10.1002/\(SICI\)1099-131X\(199601\)15:1<37::AID-FOR603>3.0.CO;2-C](https://doi.org/10.1002/(SICI)1099-131X(199601)15:1<37::AID-FOR603>3.0.CO;2-C)
- Fantino, E., & Navarro, A. (2012). Description–experience gaps: Assessments in other choice paradigms. *Journal of Behavioral Decision Making*, 25(3), 303-314.  
<https://doi.org/10.1002/bdm.737>

- Faust, D. (1986). Research on human judgment and its application to clinical practice. *Professional Psychology: Research and Practice*, 17(5), 420-430. <https://doi.org/10.1037/0735-7028.17.5.420>
- Fechner, G. T. (1860). *Elemente der Psychophysik*. Leipzig (Germany): Breitkopf und Härtel.
- Ferrell, W.R. (1994). Discrete subjective probabilities and decision analysis: Elicitation, calibration and combination. In G. Wright and P. Ayton (Eds.) *Subjective probability*. New York: Wiley (pp 411-451).
- Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological research*, 50(2), 123-129. <https://doi.org/10.1007/BF00309212>
- Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review*, 107, 659–676. <https://doi.org/10.1037/0033-295X.107.4.659>
- Fiedler, K., Kareev, Y., Avrahami, J., Beier, S., Kutzner, F., & Hütter, M. (2016). Anomalies in the detection of change: When changes in sample size are mistaken for changes in proportions. *Memory & Cognition*, 44(1), 143–161. <https://doi.org/10.3758/s13421-015-0537-z>
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage Publications.
- Fildes, R., & Goodwin, P. (2007a). Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces*, 37(6), 570-576. <https://doi.org/10.1287/inte.1070.0309>
- Fildes, R., & Goodwin, P. (2007b). Good and bad judgement in forecasting: Lessons from four companies. *Foresight*, 8(Fall), 5-10. [https://www.researchgate.net/publication/5055580\\_Good\\_and\\_Bad\\_Judgment\\_in\\_Forecasting\\_Lessons\\_from\\_Four\\_Companies](https://www.researchgate.net/publication/5055580_Good_and_Bad_Judgment_in_Forecasting_Lessons_from_Four_Companies)
- Fildes, R., & Goodwin, P. (2021). Stability in the inefficient use of forecasting systems: A case study in a supply chain company. *International Journal of Forecasting*, 37(2), 1031-1046. <https://doi.org/10.1016/j.ijforecast.2020.11.004>
- Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International journal of forecasting*, 25(1), 3-23. <https://doi.org/10.1016/j.ijforecast.2008.11.010>
- Fildes, R., & Hastings, R. (1994). The organization and improvement of market forecasting. *Journal of the Operational Research Society*, 45(1), 1-16. <https://doi.org/10.1057/jors.1994.1>
- Fildes, R., & Petropoulos, F. (2015). Improving forecast quality in practice. *Foresight: The International Journal of Applied Forecasting*, 36 (Winter), 5–12. <https://orca.cardiff.ac.uk/84332>
- Fildes, R., & Stekler, H. (2002). The state of macroeconomic forecasting. *Journal of macroeconomics*, 24(4), 435-468. [https://doi.org/10.1016/S0164-0704\(02\)00055-1](https://doi.org/10.1016/S0164-0704(02)00055-1)

- Filimonov, V., Demos, G., & Sornette, D. (2017). Modified profile likelihood inference and interval forecast of the burst of financial bubbles. *Quantitative finance*, 17(8), 1167-1186. <https://doi.org/10.1080/14697688.2016.1276298>
- Fischer, G. W. (1982). Scoring-rule feedback and the overconfidence syndrome in subjective probability forecasting. *Organizational Behavior and Human Performance*, 29(3), 352-369. [https://doi.org/10.1016/0030-5073\(82\)90250-1](https://doi.org/10.1016/0030-5073(82)90250-1).
- Fischer, I., & Harvey, N. (1999). Combining forecasts: What information do judges need to outperform the simple average? *International journal of forecasting*, 15(3), 227-246. [https://doi.org/10.1016/S0169-2070\(98\)00073-9](https://doi.org/10.1016/S0169-2070(98)00073-9)
- Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1(3), 288–299. <https://doi.org/10.1037/0096-1523.1.3.288>
- Fischhoff, B. (1994). What forecasts (seem to) mean. *International Journal of Forecasting*, 10(3), 387-403. [https://doi.org/10.1016/0169-2070\(94\)90069-8](https://doi.org/10.1016/0169-2070(94)90069-8)
- Fischhoff, B., & MacGregor, D. (1982). Subjective confidence in forecasts. *Journal of Forecasting*, 1(2), 155-172. <https://doi.org/10.1002/for.3980010203>.
- Fisher, P. J., & Yao, R. (2017). Gender differences in financial risk tolerance. *Journal of Economic Psychology*, 61, 191–202. <https://doi.org/10.1016/j.joep.2017.03.006>
- Fitzgerald, L. M., Arvaneh, M., & Dockree, P. M. (2017). Domain-specific and domain-general processes underlying metacognitive judgments. *Consciousness and Cognition*, 49, 264-277. <https://doi.org/10.1016/j.concog.2017.01.011>
- Fluch, M., & Stix, H. (2005). Perceived inflation in Austria—extent, explanations, effects. *Monetary Policy & the Economy*, Q3 (05), 22-47. [https://www.oenb.at/dam/jcr:24ab8fb0-4a9b-4ea9-a88f-974d7ef66e84/mop\\_2005\\_q3\\_analysis2\\_tcm16-34757.pdf](https://www.oenb.at/dam/jcr:24ab8fb0-4a9b-4ea9-a88f-974d7ef66e84/mop_2005_q3_analysis2_tcm16-34757.pdf)
- Foong, S. Y., Lawrence, M., & O'Connor, M. (2003). The interaction of incentives and information disclosure: The case of confidence interval information. *The British Accounting Review*, 35(3), 233-255. [https://doi.org/10.1016/S0890-8389\(03\)00028-3](https://doi.org/10.1016/S0890-8389(03)00028-3)
- Francis, J., & Philbrick, D. (1993). Analysts' decisions as products of a multi-task environment. *Journal of Accounting Research*, 31(2), 216-230. <https://doi.org/10.2307/2491271>
- Franses, P. H., Kranendonk, H. C., & Lanser, D. (2011). One model and various experts: Evaluating Dutch macroeconomic forecasts. *International Journal of Forecasting*, 27(2), 482-495. <https://doi.org/10.1016/j.ijforecast.2010.05.015>
- Frensch, P. A., & Funke, J. (1995). *Complex Problem Solving: The European Perspective*. Hillsdale, NJ: Erlbaum.
- Fried, D., & Givoly, D. (1982). Financial analysts' forecasts of earnings: A better surrogate for market expectations. *Journal of accounting and economics*, 4(2), 85-107. [https://doi.org/10.1016/0165-4101\(82\)90015-5](https://doi.org/10.1016/0165-4101(82)90015-5)

- Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, 40(1), 35-42. <https://doi.org/10.1016/j.socec.2010.10.008>
- Fuster, A., Laibson, D., & Mendel, B. (2010). Natural expectations and macroeconomic fluctuations. *Journal of Economic Perspectives*, 24(4), 67-84. <https://doi.org/10.1257/jep.24.4.67>
- Gabor, A., & Granger, C. W. (1961). On the price consciousness of consumers. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 10(3), 170-188. <https://doi.org/10.2307/2985208>
- Gábel, P., Rariga, J., & Várhegyi, J. (2014). *Inflation expectations in Hungary* (Magyar Nemzeti Bank Occasional Papers No. 113). Magyar Nemzeti Bank, Budapest. <http://hdl.handle.net/10419/141997>
- Galashin, M., Kanz, M., & Perez-Truglia, R. (2020). *Macroeconomic Expectations and Credit Card Spending* (NBER Working paper No. 28281). National Bureau of Economic Research. <https://www.nber.org/papers/w28281>
- Galí, J. (2015). *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*. Princeton University Press. <https://press.princeton.edu/books/hardcover/9780691164786/monetary-policy-inflation-and-the-business-cycle>
- Gaglianone, W. P. (2017). *Empirical findings on inflation expectations in Brazil: A survey* (Banco Central do Brasil Working Papers No. 464). Banco Central do Brasil. <https://www.bcb.gov.br/pec/wps/ingl/wps464.pdf>
- Gennaioli, N., Ma, Y., & Shleifer, A. (2016). Expectations and investment. *National Bureau of Economic Research Macroeconomics Annual*, 30(1), 379-431. <https://doi.org/10.1086/685965>
- Gentzkow, M., & Shapiro, J. M. (2006). Media bias and reputation. *Journal of political Economy*, 114(2), 280-316. <https://doi.org/10.1086/499414>
- Georganas, S., Healy, P. J., & Li, N. (2014). Frequency bias in consumers' perceptions of inflation: An experimental study. *European Economic Review*, 67, 144-158. <https://doi.org/10.1016/j.euroecorev.2014.01.014>
- García, J. A. and Manzanares, A. (2007). *Reporting biases and survey results: evidence from European professional forecasters* (ECB Working Paper No. 836). European Central Bank. <https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp836.pdf>
- Gärbling, T., & Gamble, A. (2008). Perceived inflation and expected future prices in different currencies. *Journal of Economic Psychology*, 29(4), 401-416. <https://doi.org/10.1016/j.joep.2007.04.004>
- Gatsios, R. C., Lima, F. G., Gaio, L. E., & Pimenta Júnior, T. (2021). Re-examining analyst superiority in forecasting results of publicly-traded brazilian companies. *RAM. Revista de Administração Mackenzie*, 22(1), 1-31. <https://doi.org/10.1590/1678-6971/eRAMF210164>

- Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 513-525. <https://doi.org/10.1037/0096-1523.14.3.513>
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684–704. <https://doi.org/10.1037/0033-295X.102.4.684>
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*(4), 506–528. <https://doi.org/10.1037/0033-295X.98.4.506>
- Gigerenzer, G., & Selten, R. (2001). Rethinking rationality. *Bounded rationality: The adaptive toolbox*, (pp. 1-12). MIT Press.
- Gill, M. J., Swann Jr, W. B., & Silvera, D. H. (1998). On the genesis of confidence. *Journal of Personality and Social Psychology*, *75*(5), 1101-1114. <https://doi.org/10.1037/0022-3514.75.5.1101>
- Gillies, D. (2000). *Philosophical theories of probability*. Routledge.
- Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge university press.
- Giordani, P., & Söderlind, P. (2003). Inflation forecast uncertainty. *European Economic Review*, *47*(6), 1037-1059. [https://doi.org/10.1016/S0014-2921\(02\)00236-2](https://doi.org/10.1016/S0014-2921(02)00236-2)
- Glas, A., & Hartmann, M. (2022). Uncertainty measures from partially rounded probabilistic forecast surveys. *Quantitative Economics*, *13*(3), 979-1022. <https://doi.org/10.3982/QE1703>
- Glaser, M., Iliewa, Z., & Weber, M. (2019). Thinking about prices versus thinking about returns in financial markets. *The Journal of Finance*, *74*(6), 2997-3039. <https://doi.org/10.1111/jofi.12835>
- Glaser, M., Langer, T., Reynders, J., & Weber, M. (2007). Framing effects in stock market forecasts: The difference between asking for prices and asking for returns. *Review of Finance*, *11*(2), 325-357. <https://doi.org/10.1093/rof/rfm008>
- Glaser, M., Langer, T., & Weber, M. (2013). True overconfidence in interval estimates: Evidence based on a new measure of miscalibration. *Journal of Behavioral Decision Making*, *26*(5), 405-417. <https://doi.org/10.1002/bdm.1773>
- Glaser, M., Langer, T., & Weber, M. (2007). On the trend recognition and forecasting ability of professional traders. *Decision analysis*, *4*(4), 176-193. <https://doi.org/10.1287/deca.1070.0099>
- Godau, C., Vogelgesang, T., & Gaschler, R. (2016). Perception of bar graphs—a biased impression? *Computers in Human Behavior*, *59*, 67-73. <https://doi.org/10.1016/j.chb.2016.01.036>

- Goidel, R. K., & Langley, R. E. (1995). Media coverage of the economy and aggregate economic evaluations: Uncovering evidence of indirect media effects. *Political Research Quarterly*, 48(2), 313-328. <https://doi.org/10.1177/106591299504800205>
- Goldberg, L. R. (1968). Simple models or simple processes? Some research on clinical judgments. *American Psychologist*, 23(7), 483-496. <https://doi.org/10.1037/h0026206>.
- Goldstein, D. G., & Rothschild, D. (2014). Lay understanding of probability distributions. *Judgment and Decision making*, 9(1), 1-14. <https://journal.sjdm.org/13/131029/jdm131029.pdf>
- Goodwin, P. (2017). *Forewarned: A Sceptic's Guide to Prediction*. Biteback Publishing Ltd.
- Goodwin, P., & Fildes, R. (1999). Judgmental forecasts of time series affected by special events: Does providing a statistical forecast improve accuracy? *Journal of Behavioral Decision Making*, 12(1), 37-53. [https://doi.org/10.1002/\(SICI\)1099-0771\(199903\)12:1<37::AID-BDM319>3.0.CO;2-8](https://doi.org/10.1002/(SICI)1099-0771(199903)12:1<37::AID-BDM319>3.0.CO;2-8)
- Goodwin, P., Fildes, R., Lawrence, M., & Stephens, G. (2011). Restrictiveness and guidance in support systems. *Omega*, 39(3), 242-253. <https://doi.org/10.1016/j.omega.2010.07.001>
- Goodwin, P., Moritz, B., & Siemsen, E. (2018). Forecast decisions. In E., Katok, S. Leider, & K. Donohue (Eds.), *The handbook of behavioral operations* (pp. 433-458). Wiley.
- Goodwin, P., Önkal, D., & Lawrence, M. (2011). Improving the role of judgment in economic forecasting. In M. P. Clements, & D. F. Hendry (Eds.). *The Oxford handbook of economic forecasting* (pp. 163-190). Oxford University Press, USA. <https://doi.org/10.1093/oxfordhb/9780195398649.013.0007>
- Goodwin, P., Önkal, D., & Thomson, M. (2010). Do forecasts expressed as prediction intervals improve production planning decisions? *European Journal of Operational Research*, 205(1), 195-201. <https://doi.org/10.1016/j.ejor.2009.12.020>
- Gorr, W. L. (1986). Use of special event data in government information systems. *Public Administration Review*, 532-539. <https://doi.org/10.2307/975575>
- Goyal, K., & Kumar, S. (2021). Financial literacy: A systematic review and bibliometric analysis. *International Journal of Consumer Studies*, 45(1), 80-105. <https://doi.org/10.1111/ijcs.12605>
- Graham, J. R. (1996). Is a group of economists better than one? Than none? *Journal of Business*, 193-232. <https://www.jstor.org/stable/2353464>
- Granger, C. W. (1989). Invited review combining forecasts—twenty years later. *Journal of forecasting*, 8(3), 167-173. <https://doi.org/10.1002/for.3980080303>
- Granger, C. W. (1996). Can we improve the perceived quality of economic forecasts? *Journal of Applied Econometrics*, 11(5), 455-473. [https://doi.org/10.1002/\(SICI\)1099-1255\(199609\)11:5<455::AID-JAE408>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1099-1255(199609)11:5<455::AID-JAE408>3.0.CO;2-E)
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1, pp. 1969-2012). New York: Wiley.

- Greitemeyer, T., & Schulz-Hardt, S. (2003). Preference-consistent evaluation of information in the hidden profile paradigm: beyond group-level explanations for the dominance of shared information in group decisions. *Journal of personality and social psychology*, *84*(2), 322-339. <https://doi.org/10.1037/0022-3514.84.2.322>
- Greitemeyer, T., Schulz-Hardt, S., Traut-Mattausch, E., & Frey, D. (2005). The influence of price trend expectations on price trend perceptions: Why the Euro seems to make life more expensive? *Journal of Economic Psychology*, *26*(4), 541-548. <https://doi.org/10.1016/j.joep.2004.10.003>
- Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive psychology*, *24*(3), 411-435. [https://doi.org/10.1016/0010-0285\(92\)90013-R](https://doi.org/10.1016/0010-0285(92)90013-R)
- Grounds, M. A., Joslyn, S., & Otsuka, K. (2017). Probabilistic interval forecasts: An individual differences approach to understanding forecast communication. *Advances in Meteorology*, *2017*, 3932565. <https://doi.org/10.1155/2017/3932565>
- Hajdini, I., Knotek, E., Leer, J., Pedemonte, M., Rich, R., & Schoenle, R. (2022). Low Passthrough from Inflation Expectations to Income Growth Expectations: Why People Dislike Inflation. [Discussion paper]. CEPR Discussion Paper No. DP17356. <https://ssrn.com/abstract=4144674>
- Hajdini, I., Knotek II, E. S., Pedemonte, M., Rich, R., Leer, J., & Schoenle, R. (2022). Indirect Consumer Inflation Expectations. *Economic Commentary*, no. 2022-03. <https://doi.org/10.26509/frbc-ec-202203>
- Hamill, T. M., & Wilks, D. S. (1995). A probabilistic forecast contest and the difficulty in assessing short-range forecast uncertainty. *Weather and Forecasting*, *10*(3), 620-631. [https://doi.org/10.1175/1520-0434\(1995\)010<0620:APFCAT>2.0.CO;2](https://doi.org/10.1175/1520-0434(1995)010<0620:APFCAT>2.0.CO;2)
- Hammond, K. R. (1971). Computer graphics as an aid to learning. *Science*, *172*(3986), 903-908. <https://doi.org/10.1126/science.172.3986.903>
- Hammond, K. R., & Boyle, P. J. (1971). Quasi-rationality, quarrels and new conceptions of feedback. *Bulletin of the British Psychological Society*, *24*, 103-113. <https://psycnet.apa.org/record/1972-03933-001>
- Hammond, K. R., & Summers, D. A. (1972). Cognitive control. *Psychological review*, *79*(1), 58-67. <https://doi.org/10.1037/h0031851>
- Hammond, K. R., Summers, D. A., & Deane, D. H. (1973). Negative effects of outcome-feedback in multiple-cue probability learning. *Organizational Behavior and Human Performance*, *9*(1), 30-34. [https://doi.org/10.1016/0030-5073\(73\)90034-2](https://doi.org/10.1016/0030-5073(73)90034-2)
- Handzic, M., Lam, B., Aurum, A., & Oliver, G. R. (2002). A comparative study of two knowledge discovery tools: barchart versus scatterplot. *WIT Transactions on Information and Communication Technologies*, *28*, 167-176. <https://www.witpress.com/elibrary/wit-transactions-on-information-and-communication-technologies/28/1225>

- Hansson, P., Juslin, P. & Winman, A. (2008). The role of short-term memory capacity and task experience for overconfidence in judgment under uncertainty. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(5), 1027-1042. <https://doi.org/10.1037/a0012638>
- Hansson, P., Rönnlund, M., Juslin, P., & Nilsson, L.G. (2008). Adult age differences in the realism of confidence judgments: Overconfidence, format dependence, and cognitive predictors. *Psychology and Aging*, 23(3), 531–544. <https://doi.org/10.1037/a0012782>
- Haran, U., & Moore, D. A. (2014). A better way to forecast. *California Management Review*, 57(1), 5-15. <https://doi.org/10.1525/cmr.2014.57.1.5>
- Haran, U., Moore, D. A., & Morewedge, C. K. (2010). A simple remedy for overprecision in judgment. *Judgment and Decision Making*, 5(7), 467-476. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5386407/>
- Harries, C., & Harvey, N. (2000). Taking advice, using information and knowing what you are doing. *Acta Psychologica*, 104(3), 399-416. [https://doi.org/10.1016/S0001-6918\(00\)00038-X](https://doi.org/10.1016/S0001-6918(00)00038-X).
- Harvey, J. T. (1998). Heuristic judgment theory. *Journal of Economic Issues*, 32(1), 47-64. <https://doi.org/10.1080/00213624.1998.11506010>
- Harvey, N. (1988). Judgmental forecasting of univariate time series. *Journal of Behavioral Decision Making*, 1(2), 95-110. <https://doi.org/10.1002/bdm.3960010204>
- Harvey, N. (1992). Wishful thinking impairs belief-desire reasoning: A case of decoupling failure in adults? *Cognition*, 45(2), 141-162.
- Harvey, N. (1994). Relations between confidence and skilled performance. In G. Wright & P. Ayton (Eds.), *Subjective probability*, (p. 321–352). New York: John Wiley & Sons.
- Harvey, N. (1995). Why are judgments less consistent in less predictable task situations? *Organizational Behavior and Human Decision Processes*, 63(3), 247-263. <https://doi.org/10.1006/obhd.1995.1077>
- Harvey, N. (2007). Use of heuristics: Insights from forecasting research. *Thinking & reasoning*, 13(1), 5-24. <https://doi.org/10.1080/13546780600872502>
- Harvey, N. (2011). Learning judgment and decision making from feedback. In M. K. Dhami, A. Schlottmann, and M. R. Waldmann (Eds.), *Judgment and decision making as a skill: Learning, development, and evolution*, Cambridge: Cambridge University Press. (pp. 406-464).
- Harvey, N., & Bolger, F. (1996). Graphs versus tables: Effects of data presentation format on judgemental forecasting. *International Journal of Forecasting*, 12(1), 119-137. [https://doi.org/10.1016/0169-2070\(95\)00634-6](https://doi.org/10.1016/0169-2070(95)00634-6)
- Harvey, N., Bolger, F., & McClelland, A. (1994). On the nature of expectations. *British Journal of Psychology*, 85(2), 203-229. <https://doi.org/10.1111/j.2044-8295.1994.tb02519.x>
- Harvey, N., Ewart, T., & West, R. (1997). Effects of data noise on statistical judgement. *Thinking & Reasoning*, 3(2), 111-132. <https://doi.org/10.1080/135467897394383>

- Harvey, N., & Fischer, I. (2005). Development of experience-based judgment and decision making: The role of outcome feedback. In T. Betsch & S. Haberstroh (Eds.) *The routines of decision making*. Mahwah, NJ: Lawrence Erlbaum Associates. (pp. 119-137).
- Harvey, N., Garwood, J., & Palencia, M. (1987). Vocal matching of pitch intervals: Learning and transfer effects. *Psychology of Music*, *15*(1), 90-106.  
<https://doi.org/10.1177/0305735687151007>
- Harvey, N., Koehler, D. J., & Ayton, P. (1997). Judgments of decision effectiveness: Actor-observer differences in overconfidence. *Organizational Behavior and Human Decision Processes*, *70*(3), 267-282. <https://doi.org/10.1006/obhd.1997.2710>
- Harvey, N., & Reimers, S. (2013). Trend damping: Under-adjustment, experimental artifact, or adaptation to features of the natural environment? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 589–607. <https://doi.org/10.1037/a0029179>
- Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, *108*(3), 356–388. <https://doi.org/10.1037/0096-3445.108.3.356>
- Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, *39*(12), 1372–1388. <https://doi.org/10.1037/0003-066X.39.12.1372>
- Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological review*, *93*(3), 258-268.  
<https://psycnet.apa.org/buy/1986-29133-001>
- Hastings, J. S., Madrian, B. C., & Skimmyhorn, W. L. (2013). Financial literacy, financial education, and economic outcomes. *Annual Review of Economics*, *5*(1), 347–37.  
<https://www.annualreviews.org/doi/10.1146/annurev-economics-082312-125807>
- Hawkins, E. M. (1974). *A study of the probabilistic nature of financial statements*. [Doctoral dissertation]. Texas Tech University. <https://ttu-ir.tdl.org/bitstream/handle/2346/17693/31295002115987.pdf?sequence=1>
- Heath, C., & Gonzalez, R. (1995). Interaction with others increases decision confidence but not decision quality: Evidence against information collection views of interactive decision making. *Organizational Behavior and Human Decision Processes*, *61*(3), 305-326.  
<https://doi.org/10.1006/obhd.1995.1024>
- Helson, H. (1947). Adaptation-level as frame of reference for prediction of psychophysical data. *The American journal of psychology*, *60*(1), 1-29. <https://doi.org/10.2307/1417326>
- Helson, H. (1964). *Adaptation-level theory: An experimental and systematic approach to behavior*. New York: Harper & Row.
- Henckel, T., Menzies, G. D., Moffatt, P., & Zizzo, D. J. (2019). Three dimensions of central bank credibility and inferential expectations: The Euro zone. *Journal of Macroeconomics*, *60*, 294-308. <https://doi.org/10.1016/j.jmacro.2019.01.012>

- Hertwig, R., & Gigerenzer, G. (1999). The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors. *Journal of behavioral decision making*, *12*(4), 275-305. [https://doi.org/10.1002/\(SICI\)1099-0771\(199912\)12:4<275::AID-BDM323>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1099-0771(199912)12:4<275::AID-BDM323>3.0.CO;2-M)
- Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science*, *20*(2), 231-237. <https://doi.org/10.1111/j.1467-9280.2009.02271.x>
- Herzog, S. M., & Hertwig, R. (2014). Harnessing the wisdom of the inner crowd. *Trends in cognitive sciences*, *18*(10), 504-506. <https://doi.org/10.1016/j.tics.2014.06.009>
- Hey, J. D. (1994). Expectations formation: Rational or adaptive or...? *Journal of Economic Behavior & Organization*, *25*(3), 329-349. [https://doi.org/10.1016/0167-2681\(94\)90104-X](https://doi.org/10.1016/0167-2681(94)90104-X)
- Hilton, D., Regner, I., Cabantous, L., Charalambides, L., & Vautier, S. (2011). Do positive illusions predict overconfidence in judgment? A test using interval production and probability evaluation measures of miscalibration. *Journal of Behavioral Decision Making*, *24*(2), 117-139. <https://doi.org/10.1002/bdm.678>
- Hobijn, B., Mayer, K., Stennis, C., & Topa, G. (2009, September). Household inflation experiences in the US: a comprehensive approach. Federal Reserve Bank of San Francisco. <https://www.frbsf.org/wp-content/uploads/sites/4/wp09-19bk.pdf>
- Hoch, S. J. (1985). Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 719-731. <https://doi.org/10.1037/0278-7393.11.1-4.719>
- Hoch, S. J., & Loewenstein, G. F. (1989). Outcome feedback: Hindsight and information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 605-619. <https://doi.org/10.1037/0278-7393.15.4.605>
- Hofmann, E., Kirchler, E., & Kamleitner, B. (2007). Consumer adaptation strategies: From Austrian shilling to the euro. *Journal of Consumer Policy*, *30*(4), 367-381. <https://doi.org/10.1007/s10603-007-9042-5>
- Hogarth, R. M. (1987). *Judgement and Choice*. Chichester: Wiley.
- Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive psychology*, *24*(1), 1-55. [https://doi.org/10.1016/0010-0285\(92\)90002-J](https://doi.org/10.1016/0010-0285(92)90002-J)
- Hogarth, R. M., & Soyer, E. (2011). Sequentially simulated outcomes: Kind experience versus nontransparent description. *Journal of Experimental Psychology: General*, *140*(3), 434-463. <https://doi.org/10.1037/a0023265>
- Hogarth, R. M., & Soyer, E. (2015). Communicating forecasts: The simplicity of simulated experience. *Journal of Business Research*, *68*(8), 1800-1809. <https://doi.org/10.1016/j.jbusres.2015.03.039>
- Holding, D. H., & Macrae, A. W. (1964). Guidance, restriction and knowledge of results. *Ergonomics*, *7*(3), 289-295. <https://doi.org/10.1080/00140136408930748>.

- Holzworth, R. J., & Doherty, M. E. (1976). Feedback effects in a metric multiple-cue probability learning task. *Bulletin of the Psychonomic Society*, 8(1), 1-3. <https://doi.org/10.3758/BF03337054>.
- Hommes, C. (2011). The heterogeneous expectations hypothesis: Some evidence from the lab. *Journal of Economic dynamics and control*, 35(1), 1-24. <https://doi.org/10.1016/j.jedc.2010.10.003>
- Hommes, C. (2021). Behavioral and experimental macroeconomics and policy analysis: A complex systems approach. *Journal of Economic Literature*, 59(1), 149-219. <https://doi.org/10.1257/jel.20191434>
- Hommes, C., & Makarewicz, T. (2021). Price level versus inflation targeting under heterogeneous expectations: a laboratory experiment. *Journal of Economic Behavior & Organization*, 182, 39-82. <https://doi.org/10.1016/j.jebo.2020.11.011>
- Hong, H., Kubik, J. D., & Solomon, A. (2000). Security analysts' career concerns and herding of earnings forecasts. *The Rand journal of economics*, 121-144. <https://doi.org/10.2307/2601032>
- Hopwood, W. S., & McKeown, J. C. (1990). Evidence on Surrogates for Earnings Expectations Within a Capital Market Context. *Journal of Accounting, Auditing & Finance*, 5(3), 339-363. <https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=bth&AN=7299788&site=ehost-live&scope=site&custid=s8454451>
- Hora, S. C. (2004). Probability judgments for continuous quantities: Linear combinations and calibration. *Management Science*, 50(5), 597-604. <https://doi.org/10.1287/mnsc.1040.0205>
- Hsee, C.K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluation of alternatives. *Organizational Behavior and Human Decision Processes*, 67(3), 247-257. <https://doi.org/10.1006/obhd.1996.0077>
- Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin*, 125(5), 576-590. <https://doi.org/10.1037/0033-2909.125.5.576>
- Huber, O. W. (2011). Frequency of price increases and perceived inflation. An experimental investigation. *Journal of Economic Psychology*, 32(5), 651-661. <https://doi.org/10.1016/j.joep.2011.06.004>
- Huston, S. J. (2010). Measuring financial literacy. *Journal of consumer affairs*, 44(2), 296-316. <https://doi.org/10.1111/j.1745-6606.2010.01170.x>
- Jacobson, R., & Obermiller, C. (1990). The formation of expected future price: A reference price for forward-looking consumers. *Journal of Consumer Research*, 16(4), 420-432. <https://doi.org/10.1086/209227>
- Janis, I. L. (1972). 1972 Victims of groupthink. Boston: Houghton-Mifflin.
- Janiszewski, C., & Lichtenstein, D. R. (1999). A range theory account of price perception. *Journal of consumer Research*, 25(4), 353-368. <https://doi.org/10.1086/209544>

- Johannsen, B. K. (2014). *Inflation experience and inflation expectations: Dispersion and disagreement within demographic groups* (FEDS Working Paper No. 2014-89). Board of Governors of the Federal Reserve System. <http://dx.doi.org/10.2139/ssrn.2529160>
- Johnson, E. J. (1988). Expertise and decision under uncertainty: Performance and process. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 209–228). Lawrence Erlbaum Associates, Inc.
- Johnson, J. E. V., & Bruce, A. C. (2001). Calibration of subjective probability judgments in a naturalistic setting. *Organizational Behavior and Human Decision Processes*, 85(2), 265-290. <https://doi.org/10.1006/obhd.2000.2949>
- Johnson, W. B. (1982). The impact of confidence interval information on probability judgements. *Accounting, Organizations and Society*, 7(4), 349-367. [https://doi.org/10.1016/0361-3682\(82\)90010-1](https://doi.org/10.1016/0361-3682(82)90010-1)
- Johnson, W. B. (1983). "Representativeness" in Judgmental Predictions of Corporate Bankruptcy. *Accounting Review*, 58(1), 78-97. <https://www.jstor.org/stable/246643>
- Jones, G. V. (1979). A generalized polynomial model for perception of exponential series. *Perception & Psychophysics*, 25(3), 232–234. <https://doi.org/10.3758/BF03202992>
- Jonung, L. (1981). Perceived and expected rates of inflation in Sweden. *The American Economic Review*, 71(5), 961-968. <https://www.jstor.org/stable/1803477>
- Jonung, L., & Laidler, D. (1988). Are perceptions of inflation rational? Some evidence for Sweden. *The American Economic Review*, 78(5), 1080-1087. <https://www.jstor.org/stable/1807167>
- Jørgensen, M., & Sjøberg, D. I. (2001). Software process improvement and human judgement heuristics. *Scandinavian Journal of Information Systems*, 13(1), 61-78. <https://aisel.aisnet.org/sjis/vol13/iss1/2/>
- Jørgensen, M., Teigen, K. H., & Moløkken, K. (2004). Better sure than safe? Over-confidence in judgement based software development effort prediction intervals. *Journal of systems and software*, 70(1-2), 79-93. [https://doi.org/10.1016/S0164-1212\(02\)00160-7](https://doi.org/10.1016/S0164-1212(02)00160-7)
- Jungermann, H., Brachinger, H. W., Belting, J., Grinberg, K., & Zacharias, E. (2007). The euro changeover and the factors influencing perceived inflation. *Journal of Consumer Policy*, 30(4), 405-419. <https://doi.org/10.1007/s10603-007-9051-4>
- Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes*, 57(2), 226-246. <https://doi.org/10.1006/obhd.1994.1013>
- Juslin, P., Nilsson, H., & Winman, A. (2009). Probability theory, not the very guide of life. *Psychological Review*, 116(4), 856–874. <https://doi.org/10.1037/a0016979>
- Juslin, P., Wennerholm, P., & Olsson, H. (1999). Format dependence in subjective probability calibration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(4), 1038–1052. <https://doi.org/10.1037/0278-7393.25.4.1038>

- Juslin, P., Winman, A., & Hansson, P. (2007). The naïve intuitive statistician: A naïve sampling model of intuitive confidence intervals. *Psychological Review*, *114*(3), 678-703. <https://doi.org/10.1037/0033-295X.114.3.678>
- Kabus, I. (1976). You can bank on uncertainty. *Harvard Business Review*, *54*(3), 95-105.
- Kahneman, D. (2000). A psychological point of view: Violations of rational rules as a diagnostic of mental processes. *Behavioral and Brain Sciences*, *23*(5), 681-683. <https://doi.org/10.1017/S0140525X00403432>
- Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, *58*(9), 697. <https://doi.org/10.1037/0003-066X.58.9.697>
- Kahneman, D. (2011). *Thinking, fast and slow*. London: Penguin Books Ltd.
- Kahneman, D., & Riepe, M. W. (1998). Aspects of investor psychology. *Journal of portfolio management*, *24*(4), 52-65. [https://obj.portfolioconstructionforum.edu.au/articles\\_perspectives/Portfolio-Construction-Forum\\_Kahneman\\_Aspects-of-investor-psychology.pdf](https://obj.portfolioconstructionforum.edu.au/articles_perspectives/Portfolio-Construction-Forum_Kahneman_Aspects-of-investor-psychology.pdf)
- Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. Little, Brown.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological review*, *80*(4), 237-251. <https://doi.org/10.1037/h0034747>
- Kahneman, D., & Tversky, A. (1982). Intuitive prediction: Biases and corrective procedures. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under Uncertainty: Heuristics and Biases* (pp. 414-421). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511809477.031>
- Kahneman, D., & Tversky, A. (2013a). Prospect theory: An analysis of decision under risk. In L. C. Maclean. & W. T. Ziemba (Eds.). In *Handbook of the fundamentals of financial decision making: Part I* (pp. 99-127). World Scientific. [https://doi.org/10.1142/9789814417358\\_0006](https://doi.org/10.1142/9789814417358_0006)
- Kahneman, D., & Tversky, A. (2013b). Choices, values, and frames. In *Handbook of the fundamentals of financial decision making: Part I* (pp. 269-278). [https://doi.org/10.1142/9789814417358\\_0015](https://doi.org/10.1142/9789814417358_0015)
- Kang, H., Ji, J., Yun, Y., & Han, K. (2021). Estimating Bar Graph Averages: Overcoming Within-the-Bar Bias. *i-Perception*, *12*(1), 2041669520987254. <https://doi.org/10.1177/2041669520987254>
- Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin*, *134*(3), 404-426. <https://doi.org/10.1037/0033-2909.134.3.404>.
- Katona, G. (1975). *Psychological economics*. New York: Elsevier.
- Kaufmann, C., Weber, M., & Haisley, E. (2013). The role of experience sampling and graphical displays on one's investment risk appetite. *Management science*, *59*(2), 323-340. <https://doi.org/10.1287/mnsc.1120.1607>

- Kemp, S. (1987). Estimation of past prices. *Journal of Economic Psychology*, 8(2), 181-189. [https://doi.org/10.1016/0167-4870\(87\)90019-5](https://doi.org/10.1016/0167-4870(87)90019-5)
- Kemp, S. (1991). Remembering and dating past prices. *Journal of Economic Psychology*, 12(3), 431-445. [https://doi.org/10.1016/0167-4870\(91\)90025-0](https://doi.org/10.1016/0167-4870(91)90025-0)
- Kemp, S., & Willetts, K. (1996). Remembering the price of wool. *Journal of economic psychology*, 17(1), 115-125. [https://doi.org/10.1016/0167-4870\(95\)00037-2](https://doi.org/10.1016/0167-4870(95)00037-2)
- Keren, G. (1987). Facing uncertainty in the game of bridge: A calibration study. *Organizational Behavior and Human Decision Processes*, 39(1), 98-114. [https://doi.org/10.1016/0749-5978\(87\)90047-1](https://doi.org/10.1016/0749-5978(87)90047-1)
- Keren, G. (1988). On the ability of monitoring non-veridical perceptions and uncertain knowledge: Some calibration studies. *Acta Psychologica*, 67(2), 95-119. [https://doi.org/10.1016/0001-6918\(88\)90007-8](https://doi.org/10.1016/0001-6918(88)90007-8).
- Keren, G. (1991). Calibration and probability judgements: Conceptual and methodological issues. *Acta Psychologica*, 77(3), 217-273. [https://doi.org/10.1016/0001-6918\(91\)90036-Y](https://doi.org/10.1016/0001-6918(91)90036-Y)
- Keren, G. (1997). On the calibration of probability judgments: Some critical comments and alternative perspectives. *Journal of Behavioral Decision Making*, 10(3), 269-278. [https://doi.org/10.1002/\(SICI\)1099-0771\(199709\)10:3<269::AID-BDM281>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1099-0771(199709)10:3<269::AID-BDM281>3.0.CO;2-L)
- Kerstholt, J. H. (1996). *Dynamic decision making*. Wageningen: Ponsen & Looijen B.V.
- Keys, D. E. (1978). Confidence interval financial statements: An empirical investigation. *Journal of Accounting Research*, 16(2), 389-399. <https://doi.org/10.2307/2490573>
- Khor, L. Y., Sariyev, O., & Loos, T. (2020). Gender differences in risk behavior and the link to household effects and individual wealth. *Journal of Economic Psychology*, 80, 102266. <https://doi.org/10.1016/j.joep.2020.102266>
- King, M. (2022). Monetary policy in a world of radical uncertainty. *Economic Affairs*, 42(1), 2-12. <https://doi.org/10.1111/ecaf.12513>
- Kirchgässner, G., & Müller, U. K. (2006). Are forecasters reluctant to revise their predictions? Some German evidence. *Journal of Forecasting*, 25(6), 401-413. <https://doi.org/10.1002/for.995>
- Klayman, J. (1988). On the how and why (not) of learning from outcomes. In B. Brehmer, & C. J. B. Joyce (Eds.), *Human judgement: The SJT view (Advances in psychology, Vol. 54)*. Amsterdam: Elsevier North-Holland. (pp. 115-162).
- Klayman, J. (1995). Varieties of confirmation bias. *Psychology of learning and motivation*, 32, 385-418. [https://doi.org/10.1016/S0079-7421\(08\)60315-1](https://doi.org/10.1016/S0079-7421(08)60315-1)
- Klayman, J., Soll, J. B., Gonzalez-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational behavior and human decision processes*, 79(3), 216-247. <https://doi.org/10.1006/obhd.1999.2847>

- Klayman, J., Soll, J. B., Justin, P., & Winman, A. (2006). Subjective confidence and the sampling of knowledge. In K. Fiedler & P. Juslin (Eds.), *Information Sampling and Adaptive Cognition* (pp. 153–182). New York: Cambridge University Press.
- Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin*, *107*(3), 296–310. <https://doi.org/10.1037/0033-2909.107.3.296>
- Knox, W. B., Otto, A. R., Stone, P., & Love, B. C. (2012). The nature of belief-directed exploratory choice in human decision-making. *Frontiers in psychology*, *2*, 398. <https://doi.org/10.3389/fpsyg.2011.00398>
- Kóbor, A., Kardos, Z., Takács, Á., Éltető, N., Janacsek, K., Tóth-Fáber, E., ... & Nemeth, D. (2021). Adaptation to recent outcomes attenuates the lasting effect of initial experience on risky decisions. *Scientific reports*, *11*(1), 1-20. <https://www.nature.com/articles/s41598-021-89456-1>
- Koehler, D. J. (1991). Explanation, imagination, and confidence in judgment. *Psychological bulletin*, *110*(3), 499-519. <https://psycnet.apa.org/buy/1992-04108-001>
- Koehler, J. J. (1993). The influence of prior beliefs on scientific judgments of evidence quality. *Organizational behavior and human decision processes*, *56*(1), 28-55. <https://doi.org/10.1006/obhd.1993.1044>
- Koele, P. (1980). The influence of labeled stimuli on nonlinear multiple-cue probability learning. *Organizational Behavior and Human Performance*, *26*(1), 22-31. [https://doi.org/10.1016/0030-5073\(80\)90044-6](https://doi.org/10.1016/0030-5073(80)90044-6)
- Koehler, D. J., Brenner, L., & Griffin, D. (2002). The calibration of expert judgment: Heuristics and biases beyond the laboratory. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 686- 715). New York, NY: Cambridge University Press. <http://bear.warrington.ufl.edu/brenner/mar7588/Papers/koehlerbrennergriffin2002.pdf>
- Kolb, R. A., & Stekler, H. O. (1996). How well do analysts forecast interest rates? *Journal of Forecasting*, *15*(5), 385-394. [https://doi.org/10.1002/\(SICI\)1099-131X\(199609\)15:5<385::AID-FOR627>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1099-131X(199609)15:5<385::AID-FOR627>3.0.CO;2-B)
- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 107–118. <https://doi.org/10.1037/0278-7393.6.2.107>
- Kozyreva, A., & Hertwig, R. (2021). The interpretation of uncertainty in ecological rationality. *Synthese*, *198*(2), 1517-1547. <https://doi.org/10.1007/s11229-019-02140-w>
- Krause, G. A. (2006). Beyond the norm: Cognitive biases and the behavioral underpinnings of US federal agency macroeconomic forecasts. *Rationality and society*, *18*(2), 157-191. <https://doi.org/10.1177/1043463106063322>

- Kremer, M., Moritz, B., & Siemsen, E. (2011). Demand forecasting behavior: System neglect and change detection. *Management Science*, *57*(10), 1827-1843. <https://doi.org/10.1287/mnsc.1110.1382>
- Kremer, M., Siemsen, E., & Thomas, D. J. (2016). The sum and its parts: Judgmental hierarchical forecasting. *Management Science*, *62*(9), 2745-2764. <https://doi.org/10.1287/mnsc.2015.2259>
- Krifka, M. (2002). Be brief and vague! And how bidirectional optimality theory allows for verbosity and precision. In D. Restle, D. Zaefferer, & T. Vennemann (Eds.), *Sounds and systems. Studies in Structure and Change. A Festschrift for Theo Vennemann*. (pp. 439-458). De Gruyter Mouton. <https://doi.org/10.1515/9783110894653>
- Krifka, M. (2009). Approximate Interpretations of Number Words: A Case for Strategic Communication. In H. Erhard, & J. Nerbonne (Eds.), *Theory and Evidence in Semantics* (pp. 109-132). Stanford: CSLI.
- Krizan, Z., & Windschitl, P. D. (2007). The influence of outcome desirability on optimism. *Psychological bulletin*, *133*(1), 95-121. <https://doi.org/10.1037/0033-2909.133.1.95>
- Kruglanski, A. W., Jasko, K., & Friston, K. (2020). All thinking is 'wishful' thinking. *Trends in Cognitive Sciences*, *24*(6), 413-424. <https://doi.org/10.1016/j.tics.2020.03.004>
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480-498. <https://doi.org/10.1037/0033-2909.108.3.480>
- Kurke, L. B., & Aldrich, H. E. (1983). Note—Mintzberg was right! A replication and extension of the nature of managerial work. *Management science*, *29*(8), 975-984. <https://doi.org/10.1287/mnsc.29.8.975>
- Kusev, P., Van Schaik, P., Tsaneva-Atanasova, K., Juliusson, A., & Chater, N. (2018). Adaptive anchoring model: How static and dynamic presentations of time series influence judgments and predictions. *Cognitive science*, *42*(1), 77-102. <https://doi.org/10.1111/cogs.12476>
- Kynn, M. (2008). The 'heuristics and biases' bias in expert elicitation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *171*(1), 239-264. <https://doi.org/10.1111/j.1467-985X.2007.00499.x>
- Lahiri, K., & Sheng, X. (2010a). Measuring forecast uncertainty by disagreement: The missing link. *Journal of Applied Econometrics*, *25*(4), 514-538. <https://doi.org/10.1002/jae.1167>
- Lahiri, K., & Sheng, X. (2010b). Learning and heterogeneity in GDP and inflation forecasts. *International Journal of Forecasting*, *26*(2), 265-292. <https://doi.org/10.1016/j.ijforecast.2009.12.009>
- Lambert, J., Bessièrè, V., & N'Goala, G. (2012). Does expertise influence the impact of overconfidence on judgment, valuation and investment decision? *Journal of Economic Psychology*, *33*(6), 1115-1128. <https://doi.org/10.1016/j.joep.2012.07.007>

- Lamla, M. J., & Lein, S. M. (2014). The role of media for consumers' inflation expectation formation. *Journal of Economic Behavior & Organization*, *106*, 62-77. <https://doi.org/10.1016/j.jebo.2014.05.004>
- Lamla, M. J., & Lein, S. M. (2015). Information rigidities, inflation perceptions, and the media: Lessons from the euro cash changeover. *Economic Inquiry*, *53*(1), 9-22. <https://doi.org/10.1111/ecin.12121>
- Lamla, M. J., & Maag, T. (2012). The role of media for inflation forecast disagreement of households and professional forecasters. *Journal of Money, Credit and Banking*, *44*(7), 1325-1350. <https://doi.org/10.1111/j.1538-4616.2012.00534.x>
- Lamont, O. A. (2002). Macroeconomic forecasts and microeconomic forecasters. *Journal of economic behavior & organization*, *48*(3), 265-280. [https://doi.org/10.1016/S0167-2681\(01\)00219-0](https://doi.org/10.1016/S0167-2681(01)00219-0)
- Larrick, R. P., Burson, K. A., & Soll, J. B. (2007). Social comparison and confidence: When thinking you're better than average predicts overconfidence (and when it does not). *Organizational Behavior and Human Decision Processes*, *102*(1), 76-94. <https://doi.org/10.1016/j.obhdp.2006.10.002>
- Larrick, R. P., & Lawson, M. A. (2021). Judgment and Decision-Making Processes. In *Oxford Research Encyclopedia of Psychology*. <https://doi.org/10.1093/acrefore/9780190236557.013.867>
- Laster, D., Bennett, P., & Geoum, I. S. (1999). Rational bias in macroeconomic forecasts. *The Quarterly Journal of Economics*, *114*(1), 293-318. <https://doi.org/10.1162/003355399555918>
- Lau, N., & Bearden, J. N. (2013). Newsvendor demand chasing revisited. *Management Science*, *59*(5), 1245-1249. <https://doi.org/10.1287/mnsc.1120.1617>
- Lawrence, M. J. (1983). An exploration of some practical issues in the use of quantitative forecasting models. *Journal of Forecasting*, *2*(2), 169-179. <https://doi.org/10.1002/for.3980020207>
- Lawrence, M. J., Edmundson, R. H., & O'Connor, M. J. (1985). An examination of the accuracy of judgmental extrapolation of time series. *International Journal of Forecasting*, *1*(1), 25-35. [https://doi.org/10.1016/S0169-2070\(85\)80068-6](https://doi.org/10.1016/S0169-2070(85)80068-6)
- Lawrence, M., Goodwin, P., O'Connor, M., & Önkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of forecasting*, *22*(3), 493-518. <https://doi.org/10.1016/j.ijforecast.2006.03.007>
- Lawrence, M., & Makridakis, S. (1989). Factors affecting judgmental forecasts and confidence intervals. *Organizational Behavior and Human Decision Processes*, *43*(2), 172-187. [https://doi.org/10.1016/0749-5978\(89\)90049-6](https://doi.org/10.1016/0749-5978(89)90049-6)
- Lawrence, M., & O'Connor, M. (1992). Exploring judgemental forecasting. *International Journal of Forecasting*, *8*(1), 15-26. [https://doi.org/10.1016/0169-2070\(92\)90004-S](https://doi.org/10.1016/0169-2070(92)90004-S)
- Lawrence, M., & O'Connor, M. (1993). Scale, variability, and the calibration of judgmental prediction intervals. *Organizational behavior and human decision processes*, *56*(3), 441-458. <https://doi.org/10.1006/obhd.1993.1063>

- Lawrence, M., & O'connor, M. (1995). The anchor and adjustment heuristic in time-series forecasting. *Journal of Forecasting*, 14(5), 443-451. <https://doi.org/10.1002/for.3980140504>
- Lawrence, M., O'Connor, M., & Edmundson, B. (2000). A field study of sales forecasting accuracy and processes. *European Journal of Operational Research*, 122(1), 151-160. [https://doi.org/10.1016/S0377-2217\(99\)00085-5](https://doi.org/10.1016/S0377-2217(99)00085-5)
- Lawson, M. A., Larrick, R. P., & Soll, J. B. (2020). Comparing fast thinking and slow thinking: The relative benefits of interventions, individual differences, and inferential rules. *Judgment and Decision Making*, 15(5), 660–684. <https://psycnet.apa.org/record/2020-76223-004>
- Lee, Y. S., & Siemsen, E. (2017). Task decomposition and newsvendor decision making. *Management Science*, 63(10), 3226-3245. <https://doi.org/10.1287/mnsc.2016.2521>
- Legerstee, R., & Franses, P. H. (2014). Do experts' SKU forecasts improve after feedback? *Journal of Forecasting*, 33(1), 69-79. <https://doi.org/10.1002/for.2274>
- Legoux, R., Leger, P. M., Robert, J., & Boyer, M. (2014). Confirmation biases in the financial analysis of IT investments. *Journal of the Association for Information Systems*, 15(1), 33-52. <https://aisel.aisnet.org/jais/vol15/iss1/1>
- Lein, S. M., & Maag, T. (2011). The formation of inflation perceptions: Some empirical facts for European countries. *Scottish Journal of Political Economy*, 58(2), 155-188. <https://doi.org/10.1111/j.1467-9485.2011.00541.x>
- Leiser, D., & Aroch, R. (2009). Lay understanding of macroeconomic causation: The good-begets-good heuristic. *Applied Psychology*, 58(3), 370-384. <https://doi.org/10.1111/j.1464-0597.2009.00396.x>
- Leiser, D., & Drori, S. (2005). Naïve understanding of inflation. *The Journal of Socio-Economics*, 34(2), 179-198. <https://doi.org/10.1016/j.socec.2004.09.006>
- Leiser, D., & Krill, Z. (2018). How laypeople understand the economy. In R. Ranyard (Ed.), *Economic Psychology*. Chichester: Wiley (pp 139-154).
- Leiser, D., & Shemesh, Y. (2018). *How we misunderstand economics and why it matters: The psychology of bias, distortion and conspiracy*. Routledge. <https://doi.org/10.4324/9781315675343>
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125(2), 255-275. <https://doi.org/10.1037/0033-2909.125.2.255>.
- Leung, C. (2009). The demographics of household inflation perceptions and expectations. *Reserve Bank of New Zealand Bulletin*, 72(2), 34-42. <https://www.rbnz.govt.nz/-/media/project/sites/rbnz/files/publications/bulletins/2009/2009jun72-2.pdf#page=34>
- Levenko, N. (2020). Rounding bias in forecast uncertainty. *Research in Economics*, 74(4), 277-291. <https://doi.org/10.1016/j.rie.2020.08.001>

- Levy, E., Zacks, J., Tversky, B., & Schiano, D. (1996). Gratuitous graphics? Putting preferences in perspective. *Proceedings of the SIGCHI conference on human factors in computing systems*, 42-49. <https://dl.acm.org/doi/fullHtml/10.1145/238386.238400>
- Levy, M., & Tasoff, J. (2016). Exponential-growth bias and lifecycle consumption. *Journal of the European Economic Association*, 14(3), 545-583. <https://doi.org/10.1111/jeea.12149>
- Liao, K., Zhang, Y., Lei, H., Peng, G., & Kong, W. (2022). A comparative analysis of the effects of objective and self-assessed financial literacy on stock investment return. *Frontiers in Psychology*, 13, Article 842277. <https://doi.org/10.3389/fpsyg.2022.842277>
- Lieberman, V. (2004). Local and Global Judgments of Confidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(3), 729–732. <https://doi.org/10.1037/0278-7393.30.3.729>
- Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational behavior and human performance*, 20(2), 159-183. [https://doi.org/10.1016/0030-5073\(77\)90001-0](https://doi.org/10.1016/0030-5073(77)90001-0)
- Lichtenstein, S., & Fischhoff, B. (1980). Training for calibration. *Organizational behavior and human performance*, 26(2), 149-171. [https://doi.org/10.1016/0030-5073\(80\)90052-5](https://doi.org/10.1016/0030-5073(80)90052-5)
- Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 305-334). Cambridge University Press.
- Lim, J. S., & O'Connor, M. (1996a). Judgmental forecasting with time series and causal information. *International Journal of Forecasting*, 12(1), 139-153. [https://doi.org/10.1016/0169-2070\(95\)00635-4](https://doi.org/10.1016/0169-2070(95)00635-4)
- Lim, J. S., & O'Connor, M. (1996b). Judgmental forecasting with interactive forecasting support systems. *Decision Support Systems*, 16(4), 339-357. [https://doi.org/10.1016/0167-9236\(95\)00009-7](https://doi.org/10.1016/0167-9236(95)00009-7)
- Lobo, G. J., & Nair, R. D. (1990). Combining judgmental and statistical forecasts: An application to earnings forecasts. *Decision Sciences*, 21(2), 446-460. <https://doi.org/10.1111/j.1540-5915.1990.tb01696.x>
- Løhre, E., & Teigen, K. H. (2016). There is a 60% probability, but I am 70% certain: Communicative consequences of external and internal expressions of uncertainty. *Thinking & Reasoning*, 22(4), 369-396. <https://doi.org/10.1080/13546783.2015.1069758>
- Lombardelli, C., & Saleheen, J. (2003). Public expectations of UK inflation. *SSRN 706944*. <https://ssrn.com/abstract=706944>
- Lopes, L. L., & Oden, G. C. (1987). Distinguishing between random and nonrandom events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(3), 392-400. <https://doi.org/10.1037/0278-7393.13.3.392>

- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of personality and social psychology*, 37(11), 2098. <https://doi.org/10.1037/0022-3514.37.11.2098>
- Lucas Jr, R. E. (1972). Expectations and the Neutrality of Money. *Journal of economic theory*, 4(2), 103-124.  
<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.592.6178&rep=rep1&type=pdf>
- Lusardi, A. (2019). Financial literacy and the need for financial education: evidence and implications. *Swiss Journal of Economics and Statistics*, 155(1), 1-8.  
<https://sjes.springeropen.com/articles/10.1186/s41937-019-0027-5>
- Lusardi, A., & Mitchell, O. S. (2007). Financial literacy and retirement planning: New evidence from the Rand American Life Panel. *University of Michigan Retirement Research Center Research Paper*, WP 2007-157. <http://dx.doi.org/10.2139/ssrn.1095869>
- Lusardi, A., & Mitchell, O. S. (2011a). *Financial literacy and planning: Implications for retirement wellbeing* (No. w17078). National Bureau of Economic Research.  
<https://www.nber.org/papers/w17078>
- Lusardi, A., & Mitchell, O. S. (2011b). Financial literacy around the world: an overview. *Journal of pension economics & finance*, 10(4), 497-508. <https://doi.org/10.1017/S1474747211000448>
- Lusardi, A., Mitchell, O. S., & Curto, V. (2010). Financial literacy among the young. *Journal of consumer affairs*, 44(2), 358-380. <https://doi.org/10.1111/j.1745-6606.2010.01173.x>
- Lusk, C. M., & Hammond, K. R. (1991). Judgment in a dynamic task: Microburst forecasting. *Journal of Behavioral Decision Making*, 4(1), 55-73. <https://doi.org/10.1002/bdm.3960040105>
- MacDonald, R., & Marsh, I. W. (1994). Combining exchange rate forecasts: What is the optimal consensus measure? *Journal of Forecasting*, 13(3), 313-332. <https://doi.org/10.1002/for.3980130306>
- MacGregor, D. G. (2001). Decomposition for judgmental forecasting and estimation. In J. S. Armstrong (Ed.), *Principles of forecasting: a handbook for researchers and practitioners* (pp. 107-123). Springer, Boston, MA.
- Maćkowiak, B., Matějka, F., & Wiederholt, M. (2021). *Rational inattention: A review* (ECB Working Paper Series No. 2570). European Central Bank.  
<https://www.ecb.europa.eu/pub/pdf/scpwps/ecb.wp2570~a3979fbfa5.en.pdf>
- Maćkowiak, B., & Wiederholt, M. (2009). Optimal sticky prices under rational inattention. *American Economic Review*, 99(3), 769-803. <https://doi.org/10.1257/aer.99.3.769>
- Macrae, A. W. & Holding, D. H. (1965). Guided practice in direct and reversed serial tracking. *Ergonomics*, 8(4), 487-92. <https://doi.org/10.1080/00140136508930830>.
- Maddock, R., & Carter, M. (1982). A child's guide to rational expectations. *Journal of Economic Literature*, 20(1), 39-51. <https://www.jstor.org/stable/2724658>

- Madeira, C. & Zafar, B. (2015). Heterogeneous inflation expectations and learning. *Journal of Money, Credit and Banking*, 47(5), 867-896. <https://doi.org/10.1111/jmcb.12230>
- Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K., & Simmons, L. F. (1993). The M2-competition: A real-time judgmentally based forecasting study. *International Journal of forecasting*, 9(1), 5-22. [https://doi.org/10.1016/0169-2070\(93\)90044-N](https://doi.org/10.1016/0169-2070(93)90044-N)
- Makridakis, S., & Hibon, M. (1979). Accuracy of forecasting: An empirical investigation. *Journal of the Royal Statistical Society: Series A (General)*, 142(2), 97-125. <https://doi.org/10.2307/2345077>
- Makridakis, S., Spiliotis, E., Assimakopoulos, V., Chen, Z., Gaba, A., Tsetlin, I., & Winkler, R. (2021). The M5 Uncertainty Competition: Results, Findings and Conclusions. *International journal of forecasting*. <https://doi.org/10.1016/j.ijforecast.2021.10.009>
- Makridakis, S., Wheelwright, S., & Hyndman, R. J. (1998). *Forecasting: methods and applications*. (3<sup>rd</sup> ed.). John Wiley & Sons. <http://hdl.handle.net/11728/7113>
- Malmendier, U., & Nagel, S. (2016). Learning from inflation experiences. *The Quarterly Journal of Economics*, 131(1), 53-87. <https://doi.org/10.1093/qje/qjv037>
- Mandel, D. R., Collins, R. N., Risko, E. F., & Fugelsang, J. A. (2020). Effect of confidence interval construction on judgment accuracy. *Judgment and Decision Making*, 15(5), 783-797. <https://sjdm.org/journal/20/200516/jdm200516.pdf>
- Manski, C. F. (2004). Measuring expectations. *Econometrica*, 72(5), 1329-1376. <https://doi.org/10.1111/j.1468-0262.2004.00537.x>
- Mankiw, N. G., & Reis, R. (2002). Sticky information versus sticky prices: a proposal to replace the New Keynesian Phillips curve. *The Quarterly Journal of Economics*, 117(4), 1295-1328. <https://doi.org/10.1162/003355302320935034>
- Mankiw, N. G., Reis, R., & Wolfers, J. (2003). Disagreement about inflation expectations. *NBER macroeconomics annual*, 18, 209-248. <https://doi.org/10.1086/ma.18.3585256>
- Massey, C., & Wu, G. (2005). Detecting regime shifts: The causes of under- and overreaction. *Management Science*, 51(6), 932-947. <https://doi.org/10.1287/mnsc.1050.0386>
- Mathews, B. P., & Diamantopoulos, A. (1989). Judgemental revision of sales forecasts: A longitudinal extension. *Journal of Forecasting*, 8(2), 129-140. <https://doi.org/10.1002/for.3980080206>
- Mayraz, G. (2011, October). *Wishful Thinking*. SSRN 1955644. <http://dx.doi.org/10.2139/ssrn.1955644>
- Mazumdar, T., Raj, S. P., & Sinha, I. (2005). Reference price research: Review and propositions. *Journal of marketing*, 69(4), 84-102. <https://doi.org/10.1509/jmkg.2005.69.4.84>

- McClelland, A. G., & Bolger, F. (1994). The calibration of subjective probability: Theories and models 1980–94. In: Wright G and Ayton P (eds). *Subjective probability*. Wiley: Chichester. (pp. 453-482).
- McFarland, S. G. (1981). Effects of question order on survey responses. *Public Opinion Quarterly*, 45, 208-215. <https://doi.org/10.1086/268651>
- McGranahan, L., & Paulson, A. (2006). *Constructing the Chicago fed income based economic index – Consumer price index: Inflation experiences by demographic group: 1983–2005* (Federal Reserve Bank of Chicago Working Paper No. WP2005-20). Federal Reserve Bank of Chicago. [https://fraser.stlouisfed.org/files/docs/historical/frbchi/workingpapers/frbchi\\_workingpaper\\_2005-20.pdf](https://fraser.stlouisfed.org/files/docs/historical/frbchi/workingpapers/frbchi_workingpaper_2005-20.pdf)
- McGregor, D. (1938). The major determinants of the prediction of social events. *The Journal of Abnormal and Social Psychology*, 33(2), 179-204. <https://doi.org/10.1037/h0062931>
- McKenzie, C. R., Liersch, M. J., & Yaniv, I. (2008). Overconfidence in interval estimates: What does expertise buy you? *Organizational Behavior and Human Decision Processes*, 107(2), 179-191. <https://doi.org/10.1016/j.obhdp.2008.02.007>
- McNees, S. K. (1990). The role of judgment in macroeconomic forecasting accuracy. *International Journal of Forecasting*, 6(3), 287-299. [https://doi.org/10.1016/0169-2070\(90\)90056-H](https://doi.org/10.1016/0169-2070(90)90056-H)
- Medin, D. L., Altom, M. W., & Murphy, T. D. (1984). Given versus induced category representations: use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(3), 333. <https://doi.org/10.1037/0278-7393.10.3.333>
- Meehl, P. E. (1957). When shall we use our heads instead of the formula?. *Journal of Counseling Psychology*, 4(4), 268. <https://doi.org/10.1037/h0047554>
- Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, 52(3), 194–216. <https://doi.org/10.1037/h0048070>
- Meikle, N. L., Tenney, E. R., & Moore, D. A. (2016). Overconfidence at work: Does overconfidence survive the checks and balances of organizational life? *Research in Organizational Behavior*, 36, 121-134. <https://doi.org/10.1016/j.riob.2016.11.005>
- Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., Scott, S.E., Moore, D., Atanasov, P., Swift, S. A., Murray, T., Stone, E. & Tetlock, P. (2014). Psychological strategies for winning a geopolitical forecasting tournament. *Psychological Science*, 25(5), 1106-1115. <https://doi.org/10.1177/0956797614524255>.
- Mello, J. (2009). The impact of sales forecast game playing on supply chains. *Foresight: The International Journal of Applied Forecasting*, (13), 13-22. <https://ideas.repec.org/a/for/ijafaa/y2009113p13-22.html>
- Menkhoff, L., Schmeling, M., & Schmidt, U. (2010). Are all professional investors sophisticated? *German Economic Review*, 11(4), 418-440. <https://doi.org/10.1111/j.1468-0475.2009.00497.x>

- Milani, F. (2007). Expectations, learning and macroeconomic persistence. *Journal of monetary Economics*, 54(7), 2065-2082. <https://doi.org/10.1016/j.jmoneco.2006.11.007>
- Miller, P. M. (1971). Do labels mislead? A multiple cue study, within the framework of Brunswik's probabilistic functionalism. *Organizational Behavior and Human Performance*, 6(4), 480-500. [https://doi.org/10.1016/0030-5073\(71\)90029-8](https://doi.org/10.1016/0030-5073(71)90029-8)
- Mintzberg, H. (1973). *Nature of managerial work*. New York: Harper & Row.
- Moiseev, N. A., Demenko, O. G., & Savina, N. P. (2019). Computation of reliable interval forecast for dynamic averaging of economic time series regression models. *International Journal of Civil Engineering and Technology*, 10(2), 1594-1602. <http://iaeme.com/Home/issue/IJCIET?Volume=10&Issue=2>
- Monetary Policy Committee. (2021). *Monetary Policy Report: February 2021*. Bank of England. <https://www.bankofengland.co.uk/-/media/boe/files/monetary-policy-report/2021/february/monetary-policy-report-february-2021.pdf>
- Monetary Policy Committee. (2022a). *Monetary policy report: May 2022*. Bank of England. <https://www.bankofengland.co.uk/-/media/boe/files/monetary-policy-report/2022/may/monetary-policy-report-may-2022.pdf>
- Monetary Policy Committee. (2022b). *Monetary policy report: August 2022*. Bank of England. <https://www.bankofengland.co.uk/-/media/boe/files/monetary-policy-report/2022/august/monetary-policy-report-august-2022.pdf>
- Monroe, K. B. (1973). Buyers' subjective perceptions of price. *Journal of marketing research*, 10(1), 70-80. <https://journals.sagepub.com/doi/pdf/10.1177/002224377301000110>
- Monroe, K.B. (1990). *Pricing: Making Profitable Decisions*. (2nd Eds.). McGraw Hill, New York.
- Monroe, K. B., & Lee, A. Y. (1999). Remembering versus knowing: Issues in buyers' processing of price information. *Journal of the Academy of Marketing Science*, 27(2), 207-225. <https://doi.org/10.1177/0092070399272006>
- Montgomery, L. E., & Lee, M. (2022). The Wisdom of the Crowd and Framing Effects in Spatial Knowledge. In *Proceedings of the Annual Meeting of the Cognitive Science Society*. 44. <https://escholarship.org/uc/item/0h95m7m4>
- Moore, D. (2003). *Survey of financial literacy in Washington State: Knowledge, Behavior, Attitudes, and Experiences* (SESRC Technical Report 03-39). Social and Economic Sciences Research Center, Pullman, Washington. <https://dx.doi.org/10.13140/2.1.4729.4722>
- Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2), 502-517. <https://doi.org/10.1037/0033-295X.115.2.502>
- Moore, D. A., & Schatz, D. (2017). The three faces of overconfidence. *Social and Personality Psychology Compass*, 11(8), e12331. <https://doi.org/10.1111/spc3.12331>

- Moore, D. A., Tenney, E. R., & Haran, U. (2015). Overprecision in judgment. In G. Keren, & G. Wu, (Eds.). *The Wiley-Blackwell handbook of judgment and decision making* (pp 182-209). Wiley-Blackwell. <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118468333.ch6>
- Morewedge, C. K., Gilbert, D. T., & Wilson, T. D. (2005). The least likely of times: How remembering the past biases forecasts of the future. *Psychological Science, 16*(8), 626-630. <https://doi.org/10.1111/j.1467-9280.2005.01585.x>
- Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights from the Behavioral and Brain Sciences, 2*(1), 129-140. <https://doi.org/10.1177/2372732215600886>
- Moritz, B., Siemsen, E., & Kremer, M. (2014). Judgmental forecasting: Cognitive reflection and decision speed. *Production and Operations Management, 23*(7), 1146-1160. <https://doi.org/10.1111/poms.12105>
- Moscovici, S. (1981). On social representations. In J. Forgas (Ed.), *Social cognition: Perspectives on everyday understanding* (pp. 181–210). London: Academic Press.
- Moscovici, S. (1984). The phenomenon of social representations. In R.M. Farr & S. Moscovici (Eds.), *Social representations* (pp. 3–69). Cambridge: Cambridge University Press.
- Mosteller, F., Siegel, A. F., Trapido, E., & Youtz, C. (1981). Eye fitting straight lines. *The American Statistician, 35*(3), 150-152. <https://www.tandfonline.com/doi/abs/10.1080/00031305.1981.10479335>
- Murphy, A. H., & Winkler, R. L. (1977). Reliability of subjective probability forecasts of precipitation and temperature. *Journal of the Royal Statistical Society Series C: Applied Statistics, 26*(1), 41-47. <https://doi.org/10.2307/2346866>
- Murphy, A. H., & Winkler, R. L. (1984). Probability forecasting in meteorology. *Journal of the American Statistical Association, 79*(387), 489-500. <https://doi.org/10.1080/01621459.1984.10478075>
- Murphy, G. L., & Wright, J. C. (1984). Changes in conceptual structure with expertise: Differences between real-world experts and novices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(1), 144–155. <https://doi.org/10.1037/0278-7393.10.1.144>
- Muth, J. (1961). Rational expectations and the theory of price movements. *Econometrica, 29*(3), 315–335. <https://doi.org/10.2307/1909635>
- Navajas, J., Hindocha, C., Foda, H., Keramati, M., Latham, P. E., & Bahrami, B. (2017). The idiosyncratic nature of confidence. *Nature human behaviour, 1*(11), 810-818. <https://www.nature.com/articles/s41562-017-0215-1>
- Nerlove, M. (1983). Expectations, plans, and realizations in theory and practice. *Econometrica: Journal of the Econometric Society, 51*(5), 1251-1279. <https://doi.org/10.2307/1912274>

- Newman, G. E., & Scholl, B. J. (2012). Bar graphs depicting averages are perceptually misinterpreted: The within-the-bar bias. *Psychonomic bulletin & review*, *19*(4), 601-607. <https://doi.org/10.3758/s13423-012-0247-5>
- Niedrich, R. W., Sharma, S., & Wedell, D. H. (2001). Reference price and price perceptions: A comparison of alternative models. *Journal of Consumer Research*, *28*(3), 339-354. <https://doi.org/10.1086/323726>
- Nikolopoulos, K., Lawrence, M., Goodwin, P., & Fildes, R. A. (2005). *On the accuracy of judgmental interventions on forecasting support systems* (Working paper No. 2005/022). Lancaster University. <https://eprints.lancs.ac.uk/id/eprint/48768/1/Document.pdf>
- Niu, X., & Harvey, N. (2022a). Context effects in inflation surveys: The influence of additional information and prior questions. *International Journal of Forecasting*, *38*(3), 988-1004. <https://doi.org/10.1016/j.ijforecast.2021.07.009>
- Niu, X., & Harvey, N. (2022b). Outcome feedback reduces over-forecasting of inflation and overconfidence in forecasts. *Judgment and Decision Making*, *17*(1), 124-163. <http://journal.sjdm.org/vol17.1.html>
- Niu, X., & Harvey, N. (2022c). Point, interval, and density forecasts: Differences in bias, judgment noise, and overall accuracy. *Futures & Foresight Science*, 1-18. <https://doi.org/10.1002/ffo2.124>
- Nyamekye, G. E. & Poku, E. A. (2017). What is the effect of inflation on consumer spending behavior in Ghana? *Munich Personal RePEc Archive*, Paper No. 81081. [https://mpra.ub.uni-muenchen.de/81081/1/MPRA\\_paper\\_81081.pdf](https://mpra.ub.uni-muenchen.de/81081/1/MPRA_paper_81081.pdf)
- Oberlechner, T., & Osler, C. (2012). Survival of overconfidence in currency markets. *Journal of Financial and Quantitative Analysis*, *47*(1), 91-113. <https://doi.org/10.1017/S0022109012000038>
- O'Connor, M., & Lawrence, M. (1989). An examination of the accuracy of judgemental confidence intervals in time series forecasting. *Journal of Forecasting*, *8*(2), 141-155. <https://doi.org/10.1002/for.3980080207>
- O'Connor, M., & Lawrence, M. (1992). Time series characteristics and the widths of judgemental confidence intervals. *International Journal of Forecasting*, *7*(4), 413-420. [https://doi.org/10.1016/0169-2070\(92\)90026-6](https://doi.org/10.1016/0169-2070(92)90026-6)
- O'Connor, M., & Lawrence, M. (1998). Judgmental forecasting and the use of available information', in Wright, G. and Goodwin, P. (eds), *Forecasting with Judgment* (pp. 65-90), Chichester: John Wiley.
- O'Connor, M., Remus, W., & Griggs, K. (1993). Judgmental forecasting in times of change. *International Journal of Forecasting*, *9*(2), 163-172. [https://doi.org/10.1016/0169-2070\(93\)90002-5](https://doi.org/10.1016/0169-2070(93)90002-5)
- O'Connor, M., Remus, W., & Griggs, K. (1997). Going up—going down: How good are people at forecasting trends and changes in trends? *Journal of Forecasting*, *16*(3), 165-176. [https://doi.org/10.1002/\(SICI\)1099-131X\(199705\)16:3<165::AID-FOR653>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1099-131X(199705)16:3<165::AID-FOR653>3.0.CO;2-Y)

- O'Connor, M., Remus, W., & Griggs, K. (2001). The asymmetry of judgemental confidence intervals in time series forecasting. *International Journal of Forecasting*, 17(4), 623-633.  
[https://doi.org/10.1016/S0169-2070\(01\)00103-0](https://doi.org/10.1016/S0169-2070(01)00103-0)
- Office for National Statistics (2022a). *Consumer price inflation, UK: May 2022*.  
<https://www.ons.gov.uk/economy/inflationandpriceindices/bulletins/consumerpriceinflation/may2022/pdf>
- Office for National Statistics. (2022b). *Consumer price inflation, UK: September 2022*.  
<https://www.ons.gov.uk/economy/inflationandpriceindices/bulletins/consumerpriceinflation/september2022/pdf>
- Ogburn, W. F. (1934). Studies in prediction and the distortion of reality. *Social Forces*, 13, 224-229.  
<https://doi.org/10.2307/2570338>
- O'Hagan, A., Buck, C.E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E. & Rakow, T. (2006). *Uncertain judgments: Eliciting experts' probabilities*. Chichester: Wiley.
- Okan, Y., Garcia-Retamero, R., Cokely, E. T., & Maldonado, A. (2018). Biasing and debiasing health decisions with bar graphs: Costs and benefits of graph literacy. *Quarterly Journal of Experimental Psychology*, 71(12), 2506-2519. <https://doi.org/10.1177/17470218177445>
- Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, 8(4), 434-447.  
<http://dx.doi.org/10.1037/1082-989X.8.4.434>
- Oliver, B. L. (1972). A study of confidence interval financial statements. *Journal of Accounting Research*, 10(1), 154-166. <https://doi.org/10.2307/2490224>
- Önkal, D., & Bolger, F. (2004). Provider–user differences in perceived usefulness of forecasting formats. *Omega*, 32(1), 31-39. <https://doi.org/10.1016/j.omega.2003.09.007>
- Önkal, D., & Muradoğlu, G. (1994). Evaluating probabilistic forecasts of stock prices in a developing stock market. *European Journal of Operational Research*, 74(2), 350-358.  
[https://doi.org/10.1016/0377-2217\(94\)90102-3](https://doi.org/10.1016/0377-2217(94)90102-3)
- Önkal, D., & Muradoğlu, G. (1995). Effects of feedback on probabilistic forecasts of stock prices. *International Journal of Forecasting*, 11(2), 307-319. [https://doi.org/10.1016/0169-2070\(94\)00572-T](https://doi.org/10.1016/0169-2070(94)00572-T)
- Önkal, D., & Muradoglu, G. (1996). Effects of task format on probabilistic forecasting of stock prices. *International Journal of Forecasting*, 12(1), 9-24. [https://doi.org/10.1016/0169-2070\(95\)00633-8](https://doi.org/10.1016/0169-2070(95)00633-8)
- Önkal, D., Sayım, K. Z., & Gönül, M. S. (2013). Scenarios as channels of forecast advice. *Technological Forecasting and Social Change*, 80(4), 772-788.  
<https://doi.org/10.1016/j.techfore.2012.08.015>

- Önkal, D., Yates, J. F., Simga-Mugan, C., & Öztin, Ş. (2003). Professional vs. amateur judgment accuracy: The case of foreign exchange rates. *Organizational behavior and human decision processes*, *91*(2), 169-185. [https://doi.org/10.1016/S0749-5978\(03\)00058-X](https://doi.org/10.1016/S0749-5978(03)00058-X)
- Önkal-Atay, D., Thomson, M. E., & Pollock, A. C. (2002). Judgmental forecasting. In M. P. Clements and D. F. Hendry (Ed.). *A companion to economic forecasting* (pp.133-151). Oxford: Black-well.
- Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology*, *29*(3), 261–265. <https://doi.org/10.1037/h0022125>
- Osman, M. (2010). Controlling uncertainty: A review of human behaviour in complex dynamic environments. *Psychological Bulletin*, *136*(1), 65-86. <https://doi.org/10.1037/a0017815>.
- Ottaviani, M., & Sørensen, P. N. (2006). The strategy of professional forecasting. *Journal of Financial Economics*, *81*(2), 441-466. <https://doi.org/10.1016/j.jfineco.2005.08.002>
- Pachur, T., & Biele, G. (2007). Forecasting from ignorance: The use and usefulness of recognition in lay predictions of sports events. *Acta Psychologica*, *125*(1), 99-116. <https://doi.org/10.1016/j.actpsy.2006.07.002>.
- Paese, P. W., & Snizek, J. A. (1991). Influences on the appropriateness of confidence in judgment: Practice, effort, information, and decision-making. *Organizational Behavior and Human Decision Processes*, *48*(1), 100-130. [https://doi.org/10.1016/0749-5978\(91\)90008-H](https://doi.org/10.1016/0749-5978(91)90008-H)
- Palardy, J., & Ovaska, T. (2015). Decomposing household, professional and market forecasts on inflation: a dynamic factor model analysis. *Applied Economics*, *47*(20), 2092-2101. <https://doi.org/10.1080/00036846.2014.1002889>
- Papastamos, D., Matysiak, G., & Stevenson, S. (2018). A comparative analysis of the accuracy and uncertainty in real estate and macroeconomic forecasts. *Journal of Real Estate Research*, *40*(3), 309-346. <https://doi.org/10.1080/10835547.2018.12091503>
- Parducci, A. (1965). Category judgment: a range-frequency model. *Psychological review*, *72*(6), 407. <https://doi.org/10.1037/h0022602>
- Parducci, A. (1995). *Happiness, pleasure, and judgment: The contextual theory and its applications*. Lawrence Erlbaum Associates, Inc.
- Parker, A. M., & Fischhoff, B. (2005). Decision-making competence: External validation through an individual-differences approach. *Journal of Behavioral Decision Making*, *18*(1), 1-27. <https://doi.org/10.1002/bdm.481>
- Patton, A. J., & Timmermann, A. (2010). Why do forecasters disagree? Lessons from the term structure of cross-sectional dispersion. *Journal of Monetary Economics*, *57*(7), 803-820. <https://doi.org/10.1016/j.jmoneco.2010.07.001>
- Payne, J. W., Bettman, J. R. & Johnson, E. J. (1993). *The Adaptive Decision-Maker*. Cambridge: Cambridge University Press.

- Peterson, D. K., & Pitz, G. F. (1986). Effects of amount of information on predictions of uncertain quantities. *Acta psychologica*, *61*(3), 229-241. [https://doi.org/10.1016/0001-6918\(86\)90083-1](https://doi.org/10.1016/0001-6918(86)90083-1)
- Peterson, D. K., & Pitz, G. F. (1988). Confidence, uncertainty, and the use of information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 85–92. <https://doi.org/10.1037/0278-7393.14.1.85>
- Petropoulos, F., Goodwin, P., & Fildes, R. (2017). Using a rolling training approach to improve judgmental extrapolations elicited from forecasters with technical knowledge. *International Journal of Forecasting*, *33*(1), 314-324. <https://doi.org/10.1016/j.ijforecast.2015.12.006>
- Pfajfar, D., & Santoro, E. (2010). Heterogeneity, learning and information stickiness in inflation expectations. *Journal of Economic Behavior & Organization*, *75*(3), 426-444. <https://doi.org/10.1016/j.jebo.2010.05.012>
- Pfajfar, D., & Žakelj, B. (2014). Experimental evidence on inflation expectation formation. *Journal of Economic Dynamics and Control*, *44*, 147-168. <https://doi.org/10.1016/j.jedc.2014.04.012>
- Pfajfar, D., & Žakelj, B. (2016). Uncertainty in forecasting inflation and monetary policy design: Evidence from the laboratory. *International Journal of Forecasting*, *32*(3), 849-864. <https://doi.org/10.1016/j.ijforecast.2016.01.005>
- Phelps, E. S. (1969). The new microeconomics of inflation and employment theory. *The American Economic Review*, *59*(2), 147-160. [https://econpapers.repec.org/article/aeaaecrev/v\\_3a59\\_3ay\\_3a1969\\_3ai\\_3a2\\_3ap\\_3a147-60.htm](https://econpapers.repec.org/article/aeaaecrev/v_3a59_3ay_3a1969_3ai_3a2_3ap_3a147-60.htm)
- Phillips, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861-1957. *Economica*, *25*. No.100, 283-299. <https://doi.org/10.2307/2550759>
- Phillot, M., & Rosenblatt-Wisch, R. (2018). *Inflation expectations: The effect of question ordering on forecast inconsistencies* (Swiss National Bank No. 11/2018). Swiss National Bank. [https://www.snb.ch/en/mmr/papers/id/working\\_paper\\_2018\\_11](https://www.snb.ch/en/mmr/papers/id/working_paper_2018_11)
- Picart, J. A. (1990). *Social judgment theory: investigating expert and novice judgment in a complex human inference task*. [Unpublished Ph.D. dissertation]. The University of Oklahoma. [https://www.snb.ch/n/mmr/reference/working\\_paper\\_2018\\_11/source/working\\_paper\\_2018\\_11.n.pdf](https://www.snb.ch/n/mmr/reference/working_paper_2018_11/source/working_paper_2018_11.n.pdf)
- Pierdzioch, C., Reid, M. B., & Gupta, R. (2016). Inflation forecasts and forecaster herding: Evidence from South African survey data. *Journal of Behavioral and Experimental Economics*, *62*, 42-50. <https://doi.org/10.1016/j.socec.2016.03.004>
- Pitz, G. F. (1974). Subjective probability distributions for imperfectly known quantities. In L. W. Gregg (Ed.), *Knowledge and cognition*. Lawrence Erlbaum. <https://psycnet.apa.org/record/1975-20374-004>
- Pohl, R. F., & Erdfelder, E. (2016). Hindsight bias. In R. F., Pohl (Ed.), *Cognitive illusions* (pp. 434-455). Psychology Press.

- Pohl, R. F., & Hell, W. (1996). No reduction in hindsight bias after complete information and repeated testing. *Organizational Behavior and Human Decision Processes*, 67(1), 49-58. <https://doi.org/10.1006/obhd.1996.0064>
- Potrich, A. C. G., Vieira, K. M., & Kirch, G. (2015). Determinantes da alfabetização financeira: Análise da influência de variáveis socioeconômicas e demográficas. *Revista Contabilidade & Finanças*, 26, 362-377. <https://doi.org/10.1590/1808-057x201501040>
- Pownall, G., Wasley, C., & Waymire, G. (1993). The stock price effects of alternative types of management earnings forecasts. *Accounting Review*, 64 (4), 896-912. <https://www.jstor.org/stable/248512>
- Price, P. C., & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *Journal of Behavioral Decision Making*, 17(1), 39-57. <https://doi.org/10.1002/bdm.460>
- Ramos, M. H., Van Andel, S. J., & Pappenberger, F. (2013). Do probabilistic forecasts lead to better decisions?. *Hydrology and Earth System Sciences*, 17(6), 2219-2232. <https://doi.org/10.5194/hess-17-2219-2013>
- Ranjan, R., & Gneiting, T. (2010). Combining probability forecasts. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(1), 71-91. <https://doi.org/10.1111/j.1467-9868.2009.00726.x>
- Ranyard, R., Missier, F. D., Bonini, N., Duxbury, D., & Summers, B. (2008). Perceptions and expectations of price changes and inflation: A review and conceptual framework. *Journal of Economic Psychology*, 29(4), 378-400. <https://doi.org/10.1016/j.joep.2008.07.002>
- Ranyard, R., Missier, F. D., Bonini, N., & Pietroni, D. (2017). The Citizen's Judgements of Prices and Inflation. In R. Ranyard (Ed.), *Economic Psychology*, (pp. 155-170). Wiley. <https://doi.org/10.1002/9781118926352.ch10>
- Rast, S. (2022). *Essays on the dynamics of inflation expectations*. [Unpublished Doctoral dissertation]. European University Institute.
- Reiche, L., & Meyler, A. (2022). *Making sense of consumer inflation expectations: the role of uncertainty* (ECB Working Paper No. 2022/2642). European Central Bank. <http://dx.doi.org/10.2139/ssrn.4036097>
- Reid, M., & Siklos, P. (2021). Inflation expectations surveys: a review of some survey design choices and their implications. *Studies in Economics and Econometrics*, 45(4), 283-303. <https://doi.org/10.1080/03796205.2022.2060299>
- Reimers, S., & Harvey, N. (2011). Sensitivity to autocorrelation in judgmental time series forecasting. *International Journal of Forecasting*, 27, 1196-1214. <https://doi.org/10.1016/j.ijforecast.2010.08.004>
- Reimers, S., & Harvey, N. (2022). Bars, lines, and points: The effect of graph format on judgmental forecasting. *International Journal of Forecasting*, In Press. <https://doi.org/10.1016/j.ijforecast.2022.11.003>

- Reips, U-D. (2002). Context effects in web-surveys. In B. Batinic, U-D Reips and N. Bosnjac (Eds.), *Online social sciences*. Seattle: Hogrefe and Huber.
- Reis, R. (2006). Inattentive producers. *The Review of Economic Studies*, 73(3), 793-821. <https://doi.org/10.1111/j.1467-937X.2006.00396.x>
- Remund, D. L. (2010). Financial literacy explicated: The case for a clearer definition in an increasingly complex economy. *Journal of consumer affairs*, 44(2), 276-295. <https://doi.org/10.1111/j.1745-6606.2010.01169.x>
- Remus, W., O'Connor, M., & Griggs, K. (1995). Does reliable information improve the accuracy of judgmental forecasts? *International Journal of Forecasting*, 11(2), 285-293. [https://doi.org/10.1016/0169-2070\(94\)00578-Z](https://doi.org/10.1016/0169-2070(94)00578-Z)
- Remus, W., O'Conner, M., & Griggs, K. (1996). Does feedback improve the accuracy of recurrent judgmental forecasts? *Organizational Behavior and Human Decision Processes*, 66(1), 22-30. <https://doi.org/10.1006/obhd.1996.0035>.
- Remus, W., O'Connor, M., & Griggs, K. (1998). The impact of information of unknown correctness on the judgmental forecasting process. *International Journal of Forecasting*, 14(3), 313-322. [https://doi.org/10.1016/S0169-2070\(98\)00002-8](https://doi.org/10.1016/S0169-2070(98)00002-8)
- Ren, Y., & Croson, R. (2013). Overconfidence in newsvendor orders: An experimental study. *Management Science*, 59(11), 2502-2517. <https://doi.org/10.1287/mnsc.2013.1715>
- Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on psychological science*, 7(5), 411-426. <https://doi.org/10.1177/1745691612454303>
- Romer, C. D., & Romer, D. H. (2000). Federal Reserve information and the behavior of interest rates. *American economic review*, 90(3), 429-457. <https://www.aeaweb.org/articles?id=10.1257/aer.90.3.429>
- Ronis, D. L., & Yates, J. F. (1987). Components of probability judgment accuracy: Individual consistency and effects of subject matter and assessment method. *Organizational Behavior and Human Decision Processes*, 40(2), 193-218. [https://doi.org/10.1016/0749-5978\(87\)90012-4](https://doi.org/10.1016/0749-5978(87)90012-4).
- Roos, M. W., & Schmidt, U. (2012). The importance of time-series extrapolation for macroeconomic expectations. *German Economic Review*, 13(2), 196-210. <https://doi.org/10.1111/j.1468-0475.2011.00551.x>
- Rötheli, T. F. (2020). *The Behavioral Economics of Inflation Expectations: Macroeconomics Meets Psychology*. Cambridge University Press.
- Roulston, M. S., & Kaplan, T. R. (2009). A laboratory-based study of understanding of uncertainty in 5-day site-specific temperature forecasts. *Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling*, 16(2), 237-244. <https://doi.org/10.1002/met.113>

- Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: issues and analysis. *International journal of forecasting*, 15(4), 353-375. [https://doi.org/10.1016/S0169-2070\(99\)00018-7](https://doi.org/10.1016/S0169-2070(99)00018-7)
- Russo, J. E., & Schoemaker, P. J. (1992). Managing overconfidence. *Sloan management review*, 33(2), 7-17. <https://sloanreview.mit.edu/article/managing-overconfidence/>
- Saakshi, S., & Sahu, S. (2019). An analysis of heterogeneity in inflation expectations across cities in India. *Journal of Economic Studies*, 46(5), 1116-1136. <https://www.emerald.com/insight/content/doi/10.1108/JES-08-2018-0297/full/html>
- Sanders, N. R. (1992). Accuracy of judgmental forecasts: A comparison. *Omega*, 20(3), 353-364. [https://doi.org/10.1016/0305-0483\(92\)90040-E](https://doi.org/10.1016/0305-0483(92)90040-E)
- Sanders, N. R., & Ritzman, L. P. (1992). The need for contextual and technical knowledge in judgmental forecasting. *Journal of Behavioral Decision Making*, 5(1), 39-52. <https://doi.org/10.1002/bdm.3960050106>
- Sanders, N. R., & Ritzman, L. P. (1995). Bringing judgment into combination forecasts. *Journal of Operations Management*, 13(4), 311-321. [https://doi.org/10.1016/0272-6963\(95\)00039-9](https://doi.org/10.1016/0272-6963(95)00039-9)
- Sanderson, P. M. (1989). Verbalizable knowledge and skilled task performance: Association, dissociation, and mental models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(4), 729-747. <https://doi.org/10.1037/0278-7393.15.4.729>
- Savadori, L., Nicotra, E., Rumiati, R., & Tamborini, R. (2001). Mental representation of economic crisis in Italian and Swiss samples. *Swiss Journal of Psychology/Schweizerische Zeitschrift für Psychologie/Revue Suisse de Psychologie*, 60(1), 11-14. <https://doi.org/10.1024%2F1421-0185.60.1.11>
- Schipper, K. (1991). Analysts' forecasts. *Accounting horizons*, 5(4), 105-121. <https://www.proquest.com/scholarly-journals/analysts-forecasts/docview/208907126/se-2>
- Schmalensee, R. (1976). An experimental study of expectation formation. *Econometrica: Journal of the Econometric Society*, 44(1), 17-41. <https://doi.org/10.2307/1911378>
- Schmitt, N., Coyle, B. W., & King, L. (1976). Feedback and task predictability as determinants of performance in multiple cue probability learning tasks. *Organizational Behavior and Human Performance*, 16(2), 388-402. [https://doi.org/10.1016/0030-5073\(76\)90023-4](https://doi.org/10.1016/0030-5073(76)90023-4)
- Schmitt, N., Coyle, B. W., & Saari, B. B. (1977). Types of task information feedback in multiple-cue probability learning. *Organizational Behavior and Human Performance*, 18(2), 316-328. [https://doi.org/10.1016/0030-5073\(77\)90033-2](https://doi.org/10.1016/0030-5073(77)90033-2)
- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3(4), 207-218. <https://doi.org/10.1111/j.1467-9280.1992.tb00029.x>
- Schmidt, R. A., Young, D. E., Swinnen, S., & Shapiro, D. C. (1989). Summary knowledge of results for skill acquisition: Support for the guidance hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(2), 352. <https://doi.org/10.1037/0278-7393.15.2.352>

- Schuman, H., Kalton, G. & Ludwig, J. (1983). Context and contiguity in survey questionnaires. *Public Opinion Quarterly*, 47, 112-115. <https://doi.org/10.1086/268771>
- Schwarz, N., & Sudman, S. (Eds.) (1992). *Context effects in social and psychological research*. New York: Springer-Verlag.
- Schwenk, C. R., & Cosier, R. A. (1980). Effects of the expert, devil's advocate, and dialectical inquiry methods on prediction performance. *Organizational Behavior and Human Performance*, 26(3), 409-424. [https://doi.org/10.1016/0030-5073\(80\)90076-8](https://doi.org/10.1016/0030-5073(80)90076-8)
- Seaver, D. A., Von Winterfeldt, D., & Edwards, W. (1978). Eliciting subjective probability distributions on continuous variables. *Organizational Behavior and Human Performance*, 21(3), 379-391. [https://doi.org/10.1016/0030-5073\(78\)90061-2](https://doi.org/10.1016/0030-5073(78)90061-2)
- Sedlmeier, P. (1997). BasicBayes: A tutor system for simple Bayesian inference. *Behavior Research Methods, Instruments, & Computers*, 29(3), 328-336. <https://doi.org/10.3758/BF03200584>
- Sedlmeier, P. (2000). How to improve statistical thinking: Choose the task representation wisely and learn by doing. *Instructional Science*, 28(3), 227-262. <https://doi.org/10.1023/A:1003802232617>.
- Serwe, S., & Frings, C. (2006). Who will win Wimbledon? The recognition heuristic in predicting sports events. *Journal of Behavioral Decision Making*, 19(4), 321-332. <https://doi.org/10.1002/bdm.530>.
- Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational behavior and human decision processes*, 53(2), 252-266. [https://doi.org/10.1016/0749-5978\(92\)90064-E](https://doi.org/10.1016/0749-5978(92)90064-E)
- Sharot, T. (2011). The optimism bias. *Current biology*, 21(23), R941-R945. <https://doi.org/10.1016/j.cub.2011.10.030>
- Sharot, T., Korn, C. W., & Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. *Nature neuroscience*, 14(11), 1475-1479. <https://doi.org/10.1038/nn.2949>
- Sharp, G. L., Cutler, B. L., & Penrod, S. D. (1988). Performance feedback improves the resolution of confidence judgments. *Organizational Behavior and Human Decision Processes*, 42(3), 271-283. [https://doi.org/10.1016/0749-5978\(88\)90001-5](https://doi.org/10.1016/0749-5978(88)90001-5).
- Sherden, W. A. (1998). *The fortune sellers: The big business of buying and selling predictions*. Chichester: Wiley.
- Sherif, M., & Hovland, C. I. (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change*. Yale University Press.
- Shiller, R. J. (1997). Why do people dislike inflation? In C. D. Romer & D. H. Romer (Eds.), *Reducing inflation: Motivation and strategy* (pp. 13–70). Chicago: University of Chicago Press.

- Sieck, W. R., & Arkes, H. R. (2005). The recalcitrance of overconfidence and its contribution to decision aid neglect. *Journal of Behavioral Decision Making*, 18(1), 29-53. <https://doi.org/10.1002/bdm.486>
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological review*, 63(2), 129. <https://doi.org/10.1037/h0042769>
- Simon, H. A. (1957). *Models of man, social and rational: Mathematical essays on rational human behavior in a social setting*. London: Wiley.
- Simonsohn, U., Karlsson, N., Loewenstein, G., & Ariely, D. (2008). The tree of experience in the forest of information: Overweighing experienced relative to observed information. *Games and Economic Behavior*, 62(1), 263-286. <https://doi.org/10.1016/j.geb.2007.03.010>
- Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3), 665-690. [https://doi.org/10.1016/S0304-3932\(03\)00029-1](https://doi.org/10.1016/S0304-3932(03)00029-1)
- Slobodyan, S., & Wouters, R. (2012a). Learning in a medium-scale DSGE model with expectations based on small forecasting models. *American Economic Journal: Macroeconomics*, 4(2), 65-101. <http://dx.doi.org/10.1257/mac.4.2.65>
- Slobodyan, S., & Wouters, R. (2012b). Learning in an estimated medium-scale DSGE model. *Journal of Economic Dynamics and control*, 36(1), 26-46. <https://doi.org/10.1016/j.jedc.2011.01.016>
- Slooman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological bulletin*, 119(1), 3. <https://doi.org/10.1037/0033-2909.119.1.3>
- Slooman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational behavior and human decision processes*, 91(2), 296-309. [https://doi.org/10.1016/S0749-5978\(03\)00021-9](https://doi.org/10.1016/S0749-5978(03)00021-9)
- Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational behavior and human performance*, 6(6), 649-744. [https://doi.org/10.1016/0030-5073\(71\)90033-X](https://doi.org/10.1016/0030-5073(71)90033-X)
- Smyth, J.D., Dillman, D.A. & Christian, L. M. (2009). Context effects in internet surveys: New issues and evidence. In A. D. Joynson, K. Y. A. McKenna, T. Postmes & U-D Reips (Eds.), *Oxford Handbook of Internet Psychology*. Oxford: Oxford University Press.
- Sniashko, S. (2019). Uncertainty in decision-making: A review of the international business literature. *Cogent Business & Management*, 6(1), 1650692. <https://doi.org/10.1080/23311975.2019.1650692>
- Sniezek, J. A. (1986). The role of variable labels in cue probability learning tasks. *Organizational Behavior and Human Decision Processes*, 38(2), 141-161. [https://doi.org/10.1016/0749-5978\(86\)90013-0](https://doi.org/10.1016/0749-5978(86)90013-0)
- Sniezek, J. A. (1989). An examination of group process in judgmental forecasting. *International Journal of Forecasting*, 5(2), 171-178. [https://doi.org/10.1016/0169-2070\(89\)90084-8](https://doi.org/10.1016/0169-2070(89)90084-8)

- Sniezek, J. A. (1990). A comparison of techniques for judgmental forecasting by groups with common information. *Group & Organization Studies*, 15(1), 5-19. <https://doi.org/10.1177/105960119001500102>
- Sniezek, J. A., & Buckley, T. (1991). Confidence depends on level of aggregation. *Journal of Behavioral Decision Making*, 4(4), 263-272. <https://doi.org/10.1002/bdm.3960040404>
- Sniezek, J. A., & Henry, R. A. (1989). Accuracy and confidence in group judgment. *Organizational behavior and human decision processes*, 43(1), 1-28. [https://doi.org/10.1016/0749-5978\(89\)90055-1](https://doi.org/10.1016/0749-5978(89)90055-1)
- Soll, J. B., & Klayman, J. (2004). Overconfidence in Interval Estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2), 299–314. <https://doi.org/10.1037/0278-7393.30.2.299>
- Sorić, P., Lolić, I., & Matošec, M. (2020). Some properties of inflation expectations in the euro area. *Metroeconomica*, 71(1), 176-203. <https://doi.org/10.1111/meca.12273>
- Soroka, S. N. (2006). Good news and bad news: Asymmetric responses to economic information. *The journal of Politics*, 68(2), 372-385. <https://doi.org/10.1111/j.1468-2508.2006.00413.x>
- Souleles, N. S. (2004). Expectations, heterogeneous forecast errors, and consumption: Micro evidence from the Michigan consumer sentiment surveys. *Journal of Money, Credit and Banking*, 39-72. <https://www.jstor.org/stable/3839047>
- Speirs-Bridge, A., Fidler, F., McBride, M., Flander, L., Cumming, G., & Burgman, M. (2010). Reducing overconfidence in the interval judgments of experts. *Risk Analysis: An International Journal*, 30(3), 512-523. <https://doi.org/10.1111/j.1539-6924.2009.01337.x>
- Spellman, B. A. (1996). Acting as intuitive scientists: Contingency judgments are made while controlling for alternative potential causes. *Psychological Science*, 7(6), 337-342. <https://doi.org/10.1111/j.1467-9280.1996.tb00385.x>
- Spence, M. T. (1996). Problem–problem solver characteristics affecting the calibration of judgments. *Organizational behavior and human decision processes*, 67(3), 271-279. <https://doi.org/10.1006/obhd.1996.0079>
- Spiwoks, M., Bedke, N., & Hein, O. (2008). Forecasting the past: the case of US interest rate forecasts. *Financial Markets and Portfolio Management*, 22(4), 357-379. <https://doi.org/10.1007/s11408-008-0087-5>
- Staël von Holstein, C.A.S. (1972). Probabilistic forecasting: an experiment related to the stock market. *Organizational Behavior and Human Performance*, 8 (1), 139-58. [https://doi.org/10.1016/0030-5073\(72\)90041-4](https://doi.org/10.1016/0030-5073(72)90041-4)
- Stanisławska, E., & Paloviita, M. (2021). Medium-vs. short-term consumer inflation expectations: evidence from a new euro area survey [Discussion Paper]. *Bank of Finland Research Discussion Paper No. 10/2021*. SSRN. <https://ssrn.com/abstract=3877041>

- Stanisławska, E., Paloviita, M., & Łyziak, T. (2019, July). Assessing reliability of aggregated inflation views in the European Commission Consumer Survey [Discussion Paper]. *Bank of Finland Research Discussion Paper No. 10/2019*. <http://dx.doi.org/10.2139/ssrn.3415820>
- Stengård, E., Juslin, P., Hahn, U., & van den Berg, R. (2022). On the generality and cognitive basis of base-rate neglect. *Cognition*, 226, 105160. <https://doi.org/10.1016/j.cognition.2022.105160>
- Stephensen, M. B., Martiny-Huenger, T., & Schulze, C. (2021). Confidence in complex risk judgments: the roles of uncertainty, experience, and affect. *PsyArXiv*. <https://psyarxiv.com/gk7xc/>
- Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. *Cognitive psychology*, 53(1), 1-26. <https://doi.org/10.1016/j.cogpsych.2005.10.003>
- Stock, J. H., & Watson, M. W. (2009). Phillips curve inflation forecasts. In: Fuhrer J, Kodrzycki Y, Little J, Olivei G (Ed.). *Understanding Inflation and the Implications for Monetary Policy* (pp. 99-202). Cambridge: MIT Press.
- Stone, E. R., & Opel, R. B. (2000). Training to improve calibration and discrimination: The effects of performance and environmental feedback. *Organizational behavior and human decision processes*, 83(2), 282-309. <https://doi.org/10.1006/obhd.2000.2910>
- Suantak, L., Bolger, F., & Ferrell, W. R. (1996). The hard–easy effect in subjective probability calibration. *Organizational Behavior and Human Decision Processes*, 67(2), 201-221. <https://doi.org/10.1006/obhd.1996.0074>
- Subbotin, V. (1996). Outcome feedback effects on under- and overconfident judgments (general knowledge tasks). *Organizational Behavior and Human Decision Processes*, 66(3), 268-276. <https://doi.org/10.1006/obhd.1996.0055>
- Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. New York: Random House.
- Svenson, O., & Nilsson, G. (1986). Mental economics: Subjective representations of factors related to expected inflation. *Journal of Economic Psychology*, 7(3), 327-349. [https://doi.org/10.1016/0167-4870\(86\)90024-3](https://doi.org/10.1016/0167-4870(86)90024-3)
- Tannenbaum, D., Fox, C. R. & Ülküman, G. (2017). Judgment extremity and accuracy under epistemic vs. aleatory uncertainty. *Management Science*, 63(2), 497-518. <https://doi.org/10.1287/mnsc.2015.2344>
- Tanner Jr, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological review*, 61(6), 401. <https://doi.org/10.1037/h0058700>
- Tape, T. G., Kripal, J., & Wigton, R. S. (1992). Comparing methods of learning clinical prediction from case simulations. *Medical Decision Making*, 12(3), 213-221. <https://doi.org/10.1177/0272989X9201200307>
- Taylor, J. W., & Bunn, D. W. (1999). Investigating improvements in the accuracy of prediction intervals for combinations of forecasts: A simulation study. *International Journal of Forecasting*, 15(3), 325-339. [https://doi.org/10.1016/S0169-2070\(99\)00002-3](https://doi.org/10.1016/S0169-2070(99)00002-3)

- Taylor, S. E. (1989). *Positive illusions: Creative self-deception and the healthy mind*. Basic Books/Hachette Book Group.
- Teigen, K. H., Halberg, A. M., & Fostervold, K. I. (2007). Single-limit interval estimates as reference points. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 21(3), 383-406. <https://doi.org/10.1002/acp.1283>
- Teigen, K. H., & Jørgensen, M. (2005). When 90% confidence intervals are 50% certain: On the credibility of credible intervals. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 19(4), 455-475. <https://doi.org/10.1002/acp.1085>
- Theocharis, Z., & Harvey, N. (2016). Order effects in judgmental forecasting. *International Journal of Forecasting*, 32(1), 44-60. <https://doi.org/10.1016/j.ijforecast.2015.01.007>
- Theocharis, Z., & Harvey, N. (2019). When does more mean worse? Accuracy of judgmental forecasting is nonlinearly related to length of data series. *Omega*, 87, 10-19. <https://doi.org/10.1016/j.omega.2018.11.009>
- Theocharis, Z., Smith, L. A., & Harvey, N. (2019). The influence of graphical format on judgmental forecasting accuracy: Lines versus points. *Futures & Foresight Science*, 1(1), e7. <https://doi.org/10.1002/ffo2.7>
- Thomson, M. E., Önkal-Atay, D., & Pollock, A. C. (1997). Currency forecasting: an investigation of extrapolative judgement. *International Journal of Forecasting*, 13(4), 509-526. [https://doi.org/10.1016/S0169-2070\(97\)00036-8](https://doi.org/10.1016/S0169-2070(97)00036-8)
- Thomson, M. E., Önkal-Atay, D., Pollock, A. C., & Macaulay, A. (2003). The influence of trend strength on directional probabilistic currency predictions. *International Journal of Forecasting*, 19(2), 241-256. [https://doi.org/10.1016/S0169-2070\(01\)00132-7](https://doi.org/10.1016/S0169-2070(01)00132-7)
- Thomson, M. E., Pollock, A. C., Henriksen, K. B., & Macaulay, A. (2004). The influence of the forecast horizon on judgemental probability forecasts of exchange rate movements. *The European Journal of Finance*, 10(4), 290-307. <https://doi.org/10.1080/13518470110047620>
- Thomson, M. E., Pollock, A. C., Gönül, M. S., & Önkal, D. (2013). Effects of trend strength and direction on performance and consistency in judgmental exchange rate forecasting. *International Journal of Forecasting*, 29(2), 337-353. <https://doi.org/10.1016/j.ijforecast.2012.03.004>
- Thurstone, L. L. (1926). The scoring of individual performance. *Journal of Educational Psychology*, 17(7), 446-457. <https://doi.org/10.1037/h0075125>.
- Timmers, H., & Wagenaar, W. A. (1977). Inverse statistics and misperception of exponential growth. *Perception & Psychophysics*, 21(6), 558-562. <https://doi.org/10.3758/BF03198737>
- Todd, F. J., & Hammond, K. R. (1965). Differential feedback in two multiple-cue probability learning tasks. *Behavioral Science*, 10(4), 429-435. <https://doi.org/10.1002/bs.3830100406>.

- Tong, J., & Feiler, D. (2017). A behavioral model of forecasting: Naive statistics on mental samples. *Management Science*, *63*(11), 3609-3627. <https://doi.org/10.1287/mnsc.2016.2537>
- Tourangeau, R., Rips, L. J. & Rasinski, K.A. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.
- Traut-Mattausch, E., Greitemeyer, T., Frey, D., & Schulz-Hardt, S. (2007). Illusory price increases after the euro changeover in Germany: An expectancy-consistent bias. *Journal of Consumer Policy*, *30*(4), 421-434. <https://doi.org/10.1007/s10603-007-9049-y>
- Traut-Mattausch, E., Schulz-Hardt, S., Greitemeyer, T., & Frey, D. (2004). Expectancy confirmation in spite of disconfirming evidence: The case of price increases due to the introduction of the Euro. *European Journal of Social Psychology*, *34*(6), 739-760. <https://doi.org/10.1002/ejsp.228>
- Tsai, C. I., Klayman, J., & Hastie, R. (2008). Effects of amount of information on judgment accuracy and confidence. *Organizational Behavior and Human Decision Processes*, *107*(2), 97-105. <https://doi.org/10.1016/j.obhdp.2008.01.005>
- Tuffour, J. K., Amoako, A. A., & Amartey, E. O. (2020). Assessing the effect of financial literacy among managers on the performance of small-scale enterprises. *Global Business Review*, *0972150919899753*. <https://doi.org/10.1177/0972150919899753>
- Tumen C., & Boduroglu, A. (2022). Judgmental Time Series Forecasting: A systematic analysis of graph format and trend type. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*, 3265-5271. <https://escholarship.org/uc/item/0233x259>
- Turner, D. S. (1990). The role of judgement in macroeconomic forecasting. *Journal of Forecasting*, *9*(4), 315-345. <https://doi.org/10.1002/for.3980090404>
- Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological bulletin*, *76*(2), 105. <https://doi.org/10.1037/h0031322>
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, *5*(2), 207-232. [https://doi.org/10.1016/0010-0285\(73\)90033-9](https://doi.org/10.1016/0010-0285(73)90033-9)
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124-1131. <https://www.science.org/doi/epdf/10.1126/science.185.4157.1124>
- Tversky, A., & Kahneman, D. (1982). Judgment of and by representativeness. In D. Kahneman, P. Slovic and A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 84–100). New York: Cambridge University Press.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293–315. <https://doi.org/10.1037/0033-295X.90.4.293>

- Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The quarterly journal of economics*, *106*(4), 1039-1061.  
<https://doi.org/10.2307/2937956>
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, *5*(4), 297-323.  
<https://doi.org/10.1007/BF00122574>
- Ülkümen, G., Fox, C. R., & Malle, B. F. (2016). Two dimensions of subjective uncertainty: Clues from natural language. *Journal of Experimental Psychology: General*, *145*(10), 1280-1297.  
<https://doi.org/10.1037/xge0000202>
- Van der Klaauw, W., Bruine de Bruin, W., Topa, G., Potter, S., & Bryan, M. F. (2008). Rethinking the measurement of household inflation expectations: preliminary findings. *Staff Reports 359*, Federal Reserve Bank of New York. <http://dx.doi.org/10.2139/ssrn.1320196>
- Van Dolder, D. & Van den Assem, M. J. (2018). The wisdom of the inner crowd in three large natural experiments. *Nature Human Behavior*, *2*(1), 21-26.  
<https://www.nature.com/articles/s41562-017-0247-6>
- Vanhuele, M., & Drèze, X. (2002). Measuring the price knowledge shoppers bring to the store. *Journal of marketing*, *66*(4), 72-85. <https://doi.org/10.1509/jmkg.66.4.72.18516>
- Van Raaij, W. (1989). Economic news, expectations and macro-economic behaviour. *Journal of Economic Psychology*, *10*(4), 473-493. [https://doi.org/10.1016/0167-4870\(89\)90039-1](https://doi.org/10.1016/0167-4870(89)90039-1)
- Volkman, J. (1951). *Scales of judgement and their implications for social psychology*. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads* (pp. 273–296). New York: Harper.
- Voorhees, C. M., Beck, J. M., Randhawa, P., DeTienne, K. B., & Bone, S. A. (2021). Assessing the Effects of Service Variability on Consumer Confidence and Behavior. *Journal of Service Research*, *24*(3), 405-420. <https://doi.org/10.1177/1094670520952145>
- Vul, E., & Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science*, *19*(7), 645– 647. <https://doi.org/10.1111/j.1467-9280.2008.02136.x>.
- Wagenaar, W. A., & Sagaria, S. D. (1975). Misperception of exponential growth. *Perception & Psychophysics*, *18*(6), 416-422. <https://doi.org/10.3758/BF03204114>
- Wagenaar, W. A., & Timmers, H. (1978). Extrapolation of exponential time series is not enhanced by having more data points. *Perception & Psychophysics*, *24*(2), 182-184.  
<https://doi.org/10.3758/BF03199548>
- Wagenaar, W. A., & Timmers, H. (1979). The pond-and-duckweed problem; Three experiments on the misperception of exponential growth. *Acta Psychologica*, *43*(3), 239-251.  
[https://doi.org/10.1016/0001-6918\(79\)90028-3](https://doi.org/10.1016/0001-6918(79)90028-3)
- Wallis, K. F., & Whitley, J. D. (1991). Sources of error in forecasts and expectations: UK economic models, 1984–8. *Journal of Forecasting*, *10*(3), 231-253.  
<https://doi.org/10.1002/for.3980100302>

- Wärneryd, K. E. (1986). Introduction The psychology of inflation. *Journal of economic psychology*, 7(3), 259-268. [https://doi.org/10.1016/0167-4870\(86\)90020-6](https://doi.org/10.1016/0167-4870(86)90020-6)
- Webby, R. G. (1994), *Graphical Support for the Integration of Event Information into Time-series Forecasting: An Empirical Investigation*[Unpublished Ph.D. dissertation]. University of New South Wales. <https://unsworks.unsw.edu.au/handle/1959.4/62728>
- Webby, R., & O'Connor, M. (1996). Judgemental and statistical time series forecasting: a review of the literature. *International Journal of forecasting*, 12(1), 91-118. [https://doi.org/10.1016/0169-2070\(95\)00644-3](https://doi.org/10.1016/0169-2070(95)00644-3)
- Webby, R., O'Connor, M., & Lawrence, M. (2001). Judgmental time-series forecasting using domain knowledge. In J. S. Armstrong (Ed.), *Principles of forecasting: a handbook for researchers and practitioners* (pp. 389-403). Springer, Boston, MA.
- Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual review of psychology*, 60, 53-85. <https://doi.org/10.1146/annurev.psych.60.110707.163633>
- Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, 39(5), 806–820. <https://doi.org/10.1037/0022-3514.39.5.806>
- Weiss-Cohen, L., Konstantinidis, E., Speekenbrink, M., & Harvey, N. (2016). Incorporating conflicting descriptions into decisions from experience. *Organizational Behavior and Human Decision Processes*, 135, 55-69. <https://doi.org/10.1016/j.obhdp.2016.05.005>
- Whitcomb, K. M., Önköl, D., Curley, S. P., & George Benson, P. (1995). Probability judgment accuracy for general knowledge. Cross-national differences and assessment methods. *Journal of Behavioral Decision Making*, 8(1), 51-67. <https://doi.org/10.1002/bdm.3960080105>
- Whitecotton, S. M. (1996). The effects of experience and a decision aid on the slope, scatter, and bias of earnings forecasts. *Organizational Behavior and Human Decision Processes*, 66(1), 111-121. <https://doi.org/10.1006/obhd.1996.0042>
- Wilcox, R. R. (2017). *Introduction to robust estimation and hypothesis testing* (4th ed.). Academic press.
- Winkler, R.L. and Murphy, A.H. (1979) The Use of Probabilities in Forecasts of Maximum and Minimum Temperatures. *The Meteorological Magazine*, 108(1288), 317-329. [https://digital.nmla.metoffice.gov.uk/download/file/IO\\_19898460-3d53-4bf9-9040-33aff06a7fab](https://digital.nmla.metoffice.gov.uk/download/file/IO_19898460-3d53-4bf9-9040-33aff06a7fab)
- Winman, A., Hansson, P., & Juslin, P. (2004). Subjective Probability Intervals: How to Reduce Overconfidence by Interval Evaluation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6), 1167–1175. <https://doi.org/10.1037/0278-7393.30.6.1167>
- Winman, A., & Juslin, P. (1993). Calibration of sensory and cognitive judgments: Two different accounts. *Scandinavian Journal of Psychology*, 34(2), 135-148. <https://doi.org/10.1111/j.1467-9450.1993.tb01109.x>

- Winstein, C. J., & Schmidt, R. A. (1990). Reduced frequency of knowledge of results enhances motor skill learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(4), 677-691. <https://doi.org/10.1037/0278-7393.16.4.677>.
- Wolfe, C., & Flores, B. (1990). Judgmental adjustment of earnings forecasts. *Journal of Forecasting*, 9(4), 389-405. <https://doi.org/10.1002/for.3980090407>
- Woodford, M. (2003). *Interest and prices: Foundations of a theory of monetary policy*. Princeton: Princeton University Press.
- Wright, G. (1982). Changes in the realism and distribution of probability assessments as a function of question type. *Acta Psychologica*, 52(1-2), 165-174. [https://doi.org/10.1016/0001-6918\(82\)90033-6](https://doi.org/10.1016/0001-6918(82)90033-6).
- Wright, G., & Ayton, P. (1989). Judgemental probability forecasts for personal and impersonal events. *International Journal of Forecasting*, 5(1), 117-125. [https://doi.org/10.1016/0169-2070\(89\)90069-1](https://doi.org/10.1016/0169-2070(89)90069-1)
- Wright, G., & Ayton, P. (1992). Judgmental probability forecasting in the immediate and medium term. *Organizational Behavior and Human Decision Processes*, 51(3), 344-363. [https://doi.org/10.1016/0749-5978\(92\)90017-2](https://doi.org/10.1016/0749-5978(92)90017-2)
- Wright, G., & Goodwin, P. (2009). Decision making and planning under low levels of predictability: Enhancing the scenario method. *International Journal of Forecasting*, 25(4), 813-825. <https://doi.org/10.1016/j.ijforecast.2009.05.019>
- Wulf, G., & Schmidt, R. A. (1989). The learning of generalized motor programs: Reducing the relative frequency of knowledge of results enhances memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(4), 748-757. <https://doi.org/10.1037/0278-7393.15.4.748>.
- Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational behavior and human decision processes*, 93(1), 1-13. <https://doi.org/10.1016/j.obhdp.2003.08.002>
- Yaniv, I., & Choshen-Hillel, S. (2012). Exploiting the wisdom of others to make better decisions: Suspending judgment reduces egocentrism and increases accuracy. *Journal of Behavioral Decision Making*, 25(5), 427-434. <https://doi.org/10.1002/bdm.740>
- Yaniv, I., & Foster, D. P. (1995). Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *Journal of Experimental Psychology: General*, 124(4), 424-432. <https://doi.org/10.1037/0096-3445.124.4.424>
- Yaniv, I., & Foster, D. P. (1997). Precision and accuracy of judgmental estimation. *Journal of behavioral decision making*, 10(1), 21-32. [https://doi.org/10.1002/\(SICI\)1099-0771\(199703\)10:1<21::AID-BDM243>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1099-0771(199703)10:1<21::AID-BDM243>3.0.CO;2-G)
- Yaniv, I., & Schul, Y. (1997). Elimination and inclusion procedures in judgment. *Journal of Behavioral Decision Making*, 10(3), 211-220. [https://doi.org/10.1002/\(SICI\)1099-0771\(199709\)10:3<211::AID-BDM250>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1099-0771(199709)10:3<211::AID-BDM250>3.0.CO;2-J)

- Yates, J. F. (1982). External correspondence: Decompositions of the mean probability score. *Organizational Behavior and Human Performance*, 30(1), 132-156. [https://doi.org/10.1016/0030-5073\(82\)90237-9](https://doi.org/10.1016/0030-5073(82)90237-9).
- Yates, J. F. (1990). *Judgment and decision making*. Englewood Cliffs, NJ: Prentice Hall.
- Yates, J. F. (1994). Subjective probability accuracy analysis. In G. Wright and P. Ayton (Eds.), *Subjective probability*. (pp. 381-410). John Wiley & Sons.
- Yates, J. F., & Curley, S. P. (1985). Conditional distribution analyses of probabilistic forecasts. *Journal of Forecasting*, 4(1), 61-73. <https://doi.org/10.1002/for.3980040110>.
- Yates, J. F., McDaniel, L. S., & Brown, E. S. (1991). Probabilistic forecasts of stock prices and earnings: The hazards of nascent expertise. *Organizational behavior and human decision processes*, 49(1), 60-79. [https://doi.org/10.1016/0749-5978\(91\)90042-R](https://doi.org/10.1016/0749-5978(91)90042-R)
- Yates, J. F., Price, P. C., Lee, J. W., & Ramirez, J. (1996). Good probabilistic forecasters: The 'consumer's' perspective. *International Journal of Forecasting*, 12(1), 41-56. [https://doi.org/10.1016/0169-2070\(95\)00636-2](https://doi.org/10.1016/0169-2070(95)00636-2)
- Yechiam, E., & Rakow, T. (2012). The effect of foregone outcomes on choices from experience: An individual-level modeling analysis. *Experimental Psychology*, 59(2), 55. <https://doi.org/10.1027/1618-3169/a000126>
- Zacks, J., & Tversky, B. (1999). Bars and lines: A study of graphic communication. *Memory & cognition*, 27(6), 1073-1079. <https://doi.org/10.3758/BF03201236>
- Zakay, D. (1983). The relationship between the probability assessor and the outcomes of an event as a determiner of subjective probability. *Acta psychologica*, 53(3), 271-280. [https://doi.org/10.1016/0001-6918\(83\)90008-2](https://doi.org/10.1016/0001-6918(83)90008-2)
- Zakay, D. (1992). The influence of computerized feedback on overconfidence in knowledge. *Behaviour & Information Technology*, 11(6), 329-333. <https://doi.org/10.1080/01449299208924354>.
- Zarnani, A., Karimi, S., & Musilek, P. (2019). Quantile regression and clustering models of prediction intervals for weather forecasts: A comparative study. *Forecasting*, 1(1), 169-188. <https://doi.org/10.3390/forecast1010012>
- Zarnowitz, V., & Braun, P. (1993). Twenty-two years of the NBER-ASA quarterly economic outlook surveys: aspects and comparisons of forecasting performance. In *Business cycles, indicators, and forecasting* (pp. 11-94). University of Chicago Press. <https://www.nber.org/system/files/chapters/c7189/c7189.pdf>
- Zhu, D., Wang, Q., & Goddard, J. (2022). A new hedging hypothesis regarding prediction interval formation in stock price forecasting. *Journal of Forecasting*, 41(4), 697-717. <https://doi.org/10.1002/for.2830>
- Ziv, I., & Leiser, D. (2013). The need for central resources in answering questions in different domains: Folk psychology, biology and economics. *Journal of Cognitive Psychology*, 25(7), 816-832. <https://doi.org/10.1080/20445911.2013.826663>

Zsombok, C. E., & Klein, G. (2014). *Naturalistic decision making*. Psychology Press.

## Appendices

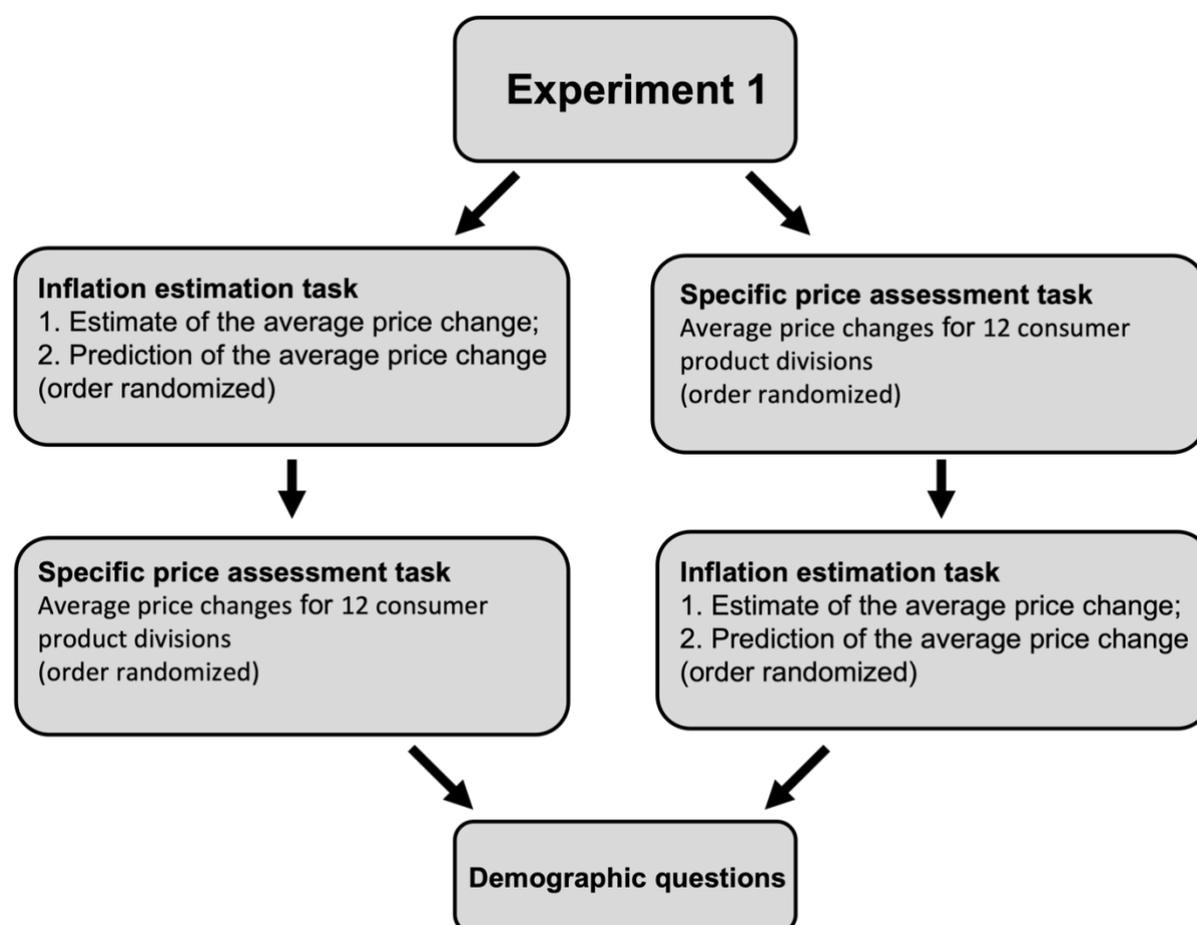
### Appendix 1

#### Method

One reviewer of research reported in Chapter 3 asked us to provide graphical illustrations of the experimental designs to increase clarity and help readers' comprehension. In this section, we do that for each of the three experiments.

**Figure A3.1**

*Experiment 1: Graphical illustration of the experiment design*



#### Specific price assessment task

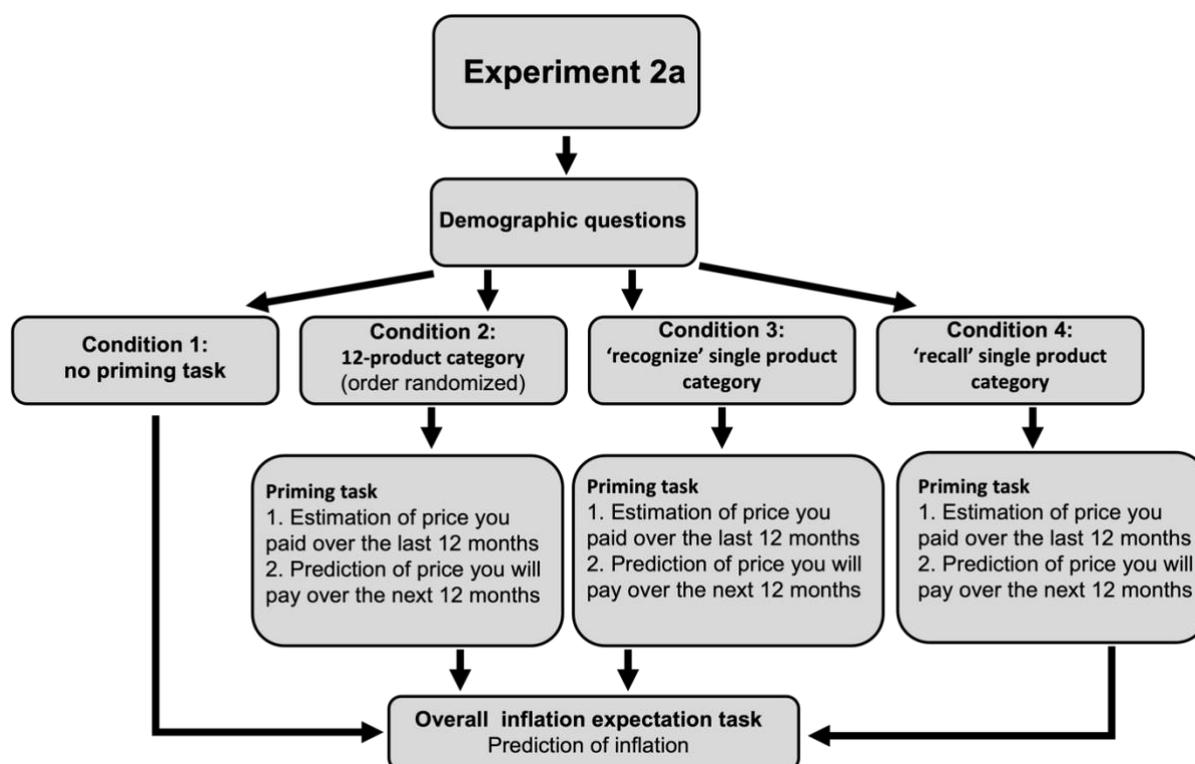
1. How many times do you buy this item per month?
2. What is the average cost of a single payment this year (June 2018-June 2019)?
3. What was the average cost of a single payment last year (June 2017-June 2018)?
4. What will be the average cost of a single payment next year (June 2019-June 2020)?
5. Self-reported reliance on personal experience

### Inflation estimation task

1. What is your prediction of the average price change (as a %) from the year June 2018-June 2019 to the year June 2019-June 2020?
2. What is your estimate of the average price change (as a %) from the year June 2017-June 2018 to the year June 2018-June 2019?

**Figure A3.2**

*Experiment 2a: Graphical illustration of the experiment design*



### Priming task

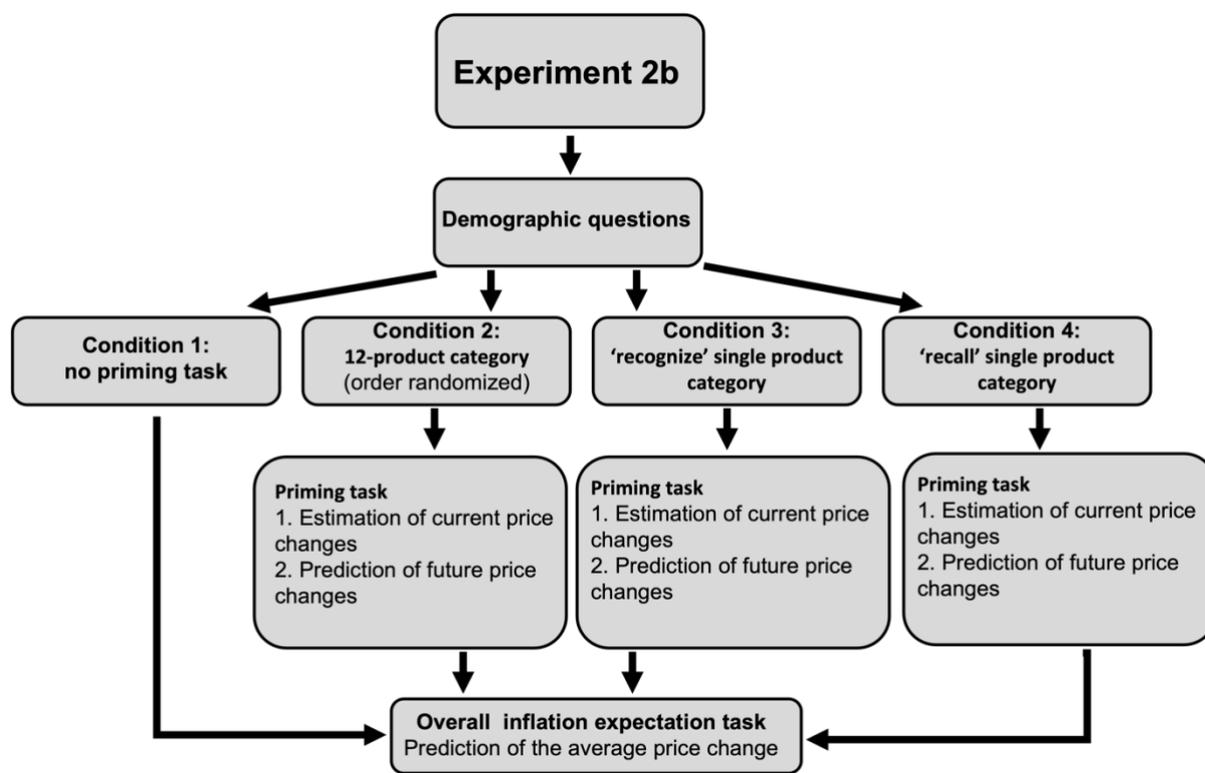
1. Over the past 12 months, did the prices you pay for [a category] go up, go down, or stay the same? If choosing change, by about what percent (%) did the prices of [the category] change?
2. During the next 12 months, do you think that the prices you pay for [a category] will go up, go down, or stay where they are now? If choosing change, by about what percent (%) do you expect prices of [the category] to change?

### Overall inflation expectation task

During the next 12 months, do you think that inflation will go up, go down, or stay where it is now? If choosing change, by about what percent (%) do you expect inflation to change?

**Figure A3.3**

*Experiment 2b: Graphical illustration of the experiment design*



The priming tasks and the overall inflation expectation task used similar question wordings as Experiment 1.

#### **Priming task**

1. How many times do you buy this item per month?
2. What is the average cost of a single payment this year (September 2021- September 2022)?
3. What was the average cost of a single payment last year (September 2020- September 2021)?
4. What will be the average cost of a single payment next year (September 2022- September 2023)?
5. Self-reported reliance on personal experience

#### **Overall inflation expectation task**

What is your prediction of the average price change (as a %) from the year September 2021- September 2022 to the year September 2022- September 2023?

## Appendix 2

### Additional results and discussion

#### Results

Here we present data tables and the results of additional analyses of the three experiments reported in Chapter 3. These analyses are not critical for testing our hypotheses but throw light on more peripheral matters that are nevertheless of some interest. For example, we show that people sampled from different countries in Experiment 1 did not differ in their judgments of inflation and we provide analyses of demographic data.

#### Experiment 1

**Table A3.1.**

*Experiment 1: Means and standard deviations (in parentheses) of inflation judgements, their constant errors, and their absolute errors*

	Direct Estimates	Indirect Estimates
<b>Inflation estimation before specific price assessment (N = 48)</b>		
Judgment		
Current inflation 2019	6.31(5.44)	18.29(29.33)
Expected inflation 2020	5.52(3.58)	18.12(22.09)
Constant Error		
Current inflation 2019	4.23(5.45)	16.21(29.42)
Expected inflation 2020	3.40(3.83)	16.00(21.86)
Absolute Error		
Current inflation 2019	4.58(5.15)	20.22(26.77)
Expected inflation 2020	4.09(3.06)	17.94(20.27)
<b>Specific price assessment before inflation estimation (N = 44)</b>		
Judgment		
Current inflation 2019	7.74(8.96)	15.54(25.45)
Expected inflation 2020	9.12(11.08)	11.66(14.24)
Constant Error		
Current inflation 2019	5.43(8.84)	13.23(25.42)
Expected inflation 2020	6.97(11.13)	9.51(14.51)
Absolute Error		
Current inflation 2019	5.62(8.72)	16.44(23.42)
Expected inflation 2020	7.20(10.97)	11.95(12.52)

### *Experiment completion times and earnings*

Mean completion time was 12 minutes (725 seconds) with a range between seven and 25 minutes. Five participants were excluded because they took longer than 30 minutes to complete the experiment. (The participant details given in the paper are *after* these five were excluded.) Average hourly earnings were £6.25 or 15RMB.

### *Analysis of CE and AE scores of participants from different countries*

The ANOVAs were carried out on 31 participants from China and 61 participants from Western countries. We conducted three-way mixed ANOVAs on judgments, CEs and VEs with within participant-variables of Task Type and Judgment Type and a between-participant variable of Country Type (East versus West). There was a main effect of Task Type on Judgment ( $F(1, 90) = 19.10$ ;  $p < 0.001$ ,  $ges = 0.0565$ ), CE ( $F(1, 90) = 19.10$ ;  $p < 0.001$ ,  $ges = 0.0564$ ), and AE ( $F(1, 90) = 39.00$ ;  $p < 0.001$ ,  $ges = 0.1077$ ). None of these analyses showed significant main or interactive effects of Country Type. In addition, independent t-tests conducted on the frequency weighted inflation judgments and on direct inflation judgments, their CEs and their AEs also failed to show any significant effects of Country Type ( $p > 0.05$ ).

### *Self-reported reliance on personal experience when estimating prices in different categories*

In this experiment, participants carrying out the specific price estimation task were asked “what percentage of your knowledge of prices in this category comes your knowledge of prices in this category comes from your personal experience rather than hearing about the prices from other people or the media”. A two-way mixed model ANOVA on self-reported reliance using Task Order as a between-participants factor and Product Category as a within-participants factor required use of Greenhouse–Geisser corrections to degrees of freedom and p values because Mauchly’s test showed a deviation from sphericity. It showed only an effect of Product Category ( $F(7.69, 692.41) = 5.62$ ,  $p < 0.001$ ,  $ges = 0.0341$ ). People gave high percentage ratings for reliance on personal experience for food and non-alcoholic beverages, for clothing and footwear, for communication, for restaurants and hotels, and for recreation and culture; they gave low ratings for housing, water, electricity, gas, and other fuels, and for furniture, household equipment and maintenance. Pairwise comparisons showed that ratings for items in the former set of categories were all significantly higher than ratings in the latter set of categories. However, there was no evidence that ratings of reliance on personal experience in different categories correlated with judged frequency of purchase in those categories.

We should not assume that people have insight into the cognitive process that they use to produce these ratings of reliance on personal experience in each category. The processing needed to generate those ratings may be intuitive; people may not be consciously aware of how it occurs. In other words, within a dual system model of cognition (e.g., Evans, 2008, 2010; Kahneman, 2011), we would say that the processing required to produce judgments of

reliance on personal experiences in each category is carried out by System 1. However, when participants are later asked how much they relied on personal experience when producing those price estimates, those who had produced particularly high expectations may then have consciously sought to account for them by using System 2 to infer how important personal experience was producing rice estimates. One function of System 2 processing is to provide a post-hoc rationale for the results of System 1 processing (Evans, 2008). So we can see that categories listed above as being given high ratings for the role of personal experience in price estimation were categories where personal choice of items is recognised as important (e.g., clothing, food, restaurant meals) whereas categories given low ratings are those where it is not recognised as important (e.g., housing, water, fuel, household equipment and maintenance).

### *Analysis of effects of demographic variables*

We conducted multiple regressions using all five demographical factors (age, gender, culture, education, and school exposure to an economics-related major) as predictors. For direct current inflation estimates, no models predicting CE or AE were significant. For direct estimates of future inflation, CE scores were predicted ( $F(5, 86) = 2.84, p = 0.02, \text{Adjusted } R^2 = 0.09$ ) only by a model that included age as the sole predictor ( $\beta = 0.31, t(91) = 2.99, p = 0.004$ ). Older people overestimated future inflation more. For these estimates, the overall model predicting AE scores was marginally significant ( $F(5, 86) = 2.21, p = 0.06, \text{Adjusted } R^2 = 0.06$ ), with age again as the sole predictor ( $\beta = 0.30, t(91) = 2.90, p = 0.005$ ). Older people made less accurate estimates of future inflation.

No overall model significantly predicted CE scores or AE scores of indirect estimates of current or future inflation.

## **Experiment 2a**

### *Experiment completion times and earnings*

Mean completion time was 11 minutes (668 seconds) with a range between five and 26 minutes. (This range is partly explained by the fact that participants were assigned to conditions that took different times to complete.) Five participants were excluded because they took longer than 30 minutes to complete the experiment. (The participant details given in the paper are *after* these five were excluded.) Average hourly earnings were £5.45.

### *Comparison of CE scores of indirect estimates of current and future inflation*

A two-way mixed ANOVA on CE scores of the indirect estimates obtained from the priming tasks with Estimate Type (Current versus Future Inflation) as a within-participant factor and Task Type (the three priming conditions) as a between-participant factor showed that people expected future inflation to be lower than current inflation ( $F(1, 146) = 5.80, p = 0.02, \text{ges} =$

0.0078). Also, inflation was judged to be higher in the two priming conditions in which people assessed inflation for the single product category with the largest price rise than in the condition in which people assessed inflation across all 12 product categories  $F(2, 146) = 24.65$ ,  $p < 0.001$ ,  $ges = 0.2133$ ). These effects are shown in Figure SM1.

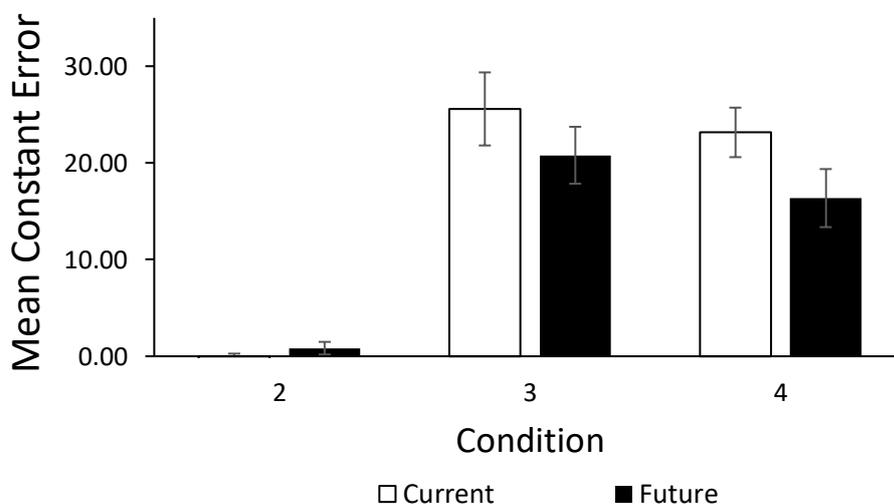
**Table A3.2.**

*Experiment 2a: Means and standard deviations (in parentheses) of estimates of future inflation, their constant errors, and their absolute errors*

	Direct Estimates	Indirect Estimates
<b>Judgment</b>		
Condition 1	6.56(6.07)	
Condition 2	6.82(4.59)	7.49(4.35)
Condition 3	7.36(6.50)	27.44(21.43)
Condition 4	8.71(7.08)	23.01(21.67)
<b>Constant Error</b>		
Condition 1	-0.09(6.07)	
Condition 2	0.17(4.59)	0.84(4.35)
Condition 3	0.71(6.50)	20.79(21.43)
Condition 4	2.06(7.08)	16.36(21.67)
<b>Absolute Error</b>		
Condition 1	4.48(4.05)	
Condition 2	3.60(2.80)	3.33(2.87)
Condition 3	4.56(4.64)	21.84(20.34)
Condition 4	4.91(5.47)	17.16(21.03)

**Figure A3.4.**

*Experiment 2a: CE scores of indirect estimates of inflation*



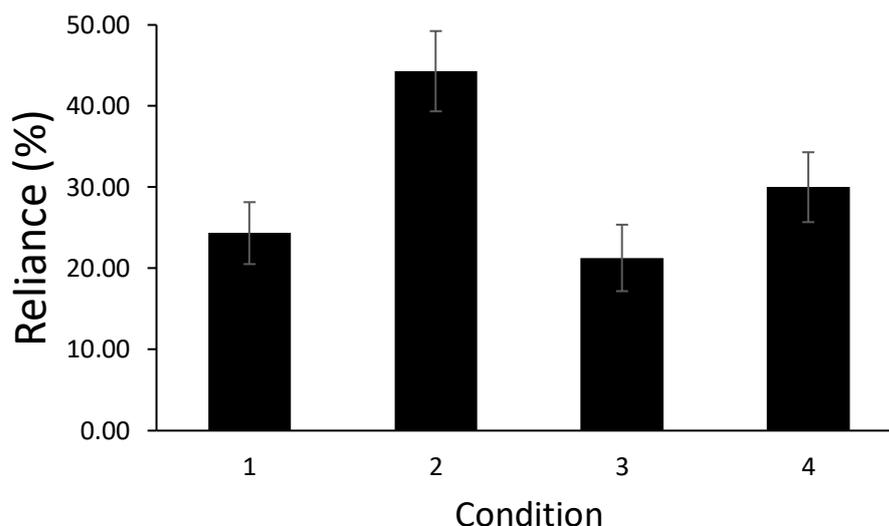
Greater overestimation of current than of future inflation occurred only in Conditions 3 and 4. In these conditions, participants were primed by estimating price changes for just the product category that they judged had shown the largest price change over the previous 12 months. It appears that thinking of the largest price change in the previous 12 months has a greater priming effect on judging current price changes than on judging future ones. In other words, temporal distance influences the size of the priming effect.

#### *Self-reported reliance on experience of specific price changes*

Figure SM2 shows a significant effect of Task Type on reported degree of reliance on specific price changes when directly estimating level of future inflation ( $F(3, 201) = 5.28, p = 0.002, \eta^2 = 0.0730$ ). Pairwise comparisons using the BH adjustment method revealed a significant difference between Condition 2 and other three conditions: Condition 1 ( $p = 0.004$ ), Condition 3 ( $p = 0.002$ ), and Condition 4 ( $p = 0.046$ ).

#### **Figure A3.5**

*Experiment 2a: Judged % reliance on past experiences with specific price changes when making direct estimates of the level of future inflation*



These data indicate that people in Condition 2 who were first required to recall prices in all categories reported that they later relied more on recall of specific price changes when directly estimating level of future inflation than those who were not initially required to recall any prices or those who were required to recall prices for just a single product category. However, we should not assume that people have insight into the cognitive process that they use to produce direct estimates of inflation. The processing needed to generate inflation expectations may be intuitive; people may not be consciously aware of how it occurs. In other words, within a dual system model of cognition (e.g., Evans, 2008, 2010; Kahneman, 2011), we would say that the processing required to estimate inflation is carried out by System 1. However, when participants are later asked whether they had thought of prices for any specific things when generating their inflation expectations, those who had produced

particularly high expectations may then have consciously sought to account for them by using System 2 to think of goods or services that had shown particularly high price rises over the previous year. One function of System 2 processing is to provide a post-hoc rationale for the results of System 1 processing (Evans, 2008).

### *Analysis of effects of demographic variables*

In this experiment, our battery of demographic variables included data from two questionnaires. The first of these was an 18-question financial literacy questionnaire. Five of these questions that measured basic financial literacy and eight that measured advanced financial literacy were drawn from Lusardi and Mitchell (2007). The remaining five questions were taken from those listed by Liao et al. (2022). There were no significant correlations across the whole set of participants between scores on this questionnaire and either direct or indirect estimates of future inflation.

The second questionnaire that participants completed was Rammstedt and John's (2007) brief Big Five Personality Inventory (BFI-10). This uses five-point Likert scales to measure extraversion, agreeableness, conscientiousness, neuroticism and openness to experience. Conscientiousness was correlated with of indirect estimates of current inflation ( $r = 0.24$ ,  $t(147) = 2.94$ ,  $p = 0.004$ ), and with the CE scores ( $r = 0.24$ ,  $t(147) = 2.94$ ,  $p = 0.004$ ) and AE scores ( $r = 0.25$ ,  $t(147) = 3.19$ ,  $p = 0.002$ ) associated with those estimates. Conscientiousness was also correlated with of indirect estimates of future inflation ( $r = 0.24$ ,  $t(147) = 3.02$ ,  $p = 0.003$ ), and with the CE scores ( $r = 0.24$ ,  $t(147) = 3.02$ ,  $p = 0.003$ ) and AE scores ( $r = 0.24$ ,  $t(147) = 3.01$ ,  $p = 0.003$ ) associated with those estimates.

We conducted multiple regressions using all eight demographical factors (age, gender, income, education, marital status, ethnicity, school exposure of financial education and work exposure of financial education) as predictors. The model that explained a significant amount of the variance in CE scores of direct estimates of future inflation ( $F(8, 195) = 2.60$ ,  $p = 0.01$ , Adjusted  $R^2 = 0.06$ ) included income ( $\beta = -2.24$ ,  $t(203) = -2.49$ ,  $p = 0.01$ ), marital status ( $\beta = 1.88$ ,  $t(203) = 1.99$ ,  $p = 0.048$ ) and school exposure of financial education ( $\beta = 3.15$ ,  $t(203) = 2.01$ ,  $p = 0.046$ ) as predictors. People with lower income, who lived with others, and who had more school exposure of financial education overestimated future inflation more. Poorer people are likely to spend relatively high proportions of their income on products in categories that have relatively high inflation rates (e.g., food). No model significantly predicted AE scores of direct estimates of future inflation.

Again, no overall model significantly predicted CE scores or AE scores of indirect estimates of current or future inflation.

## Experiment 2b

**Table A3.3**

*Experiment 2b: Means and standard deviations (in parentheses) of estimates of future inflation, their constant errors, and their absolute errors*

	Direct Estimates	Indirect Estimates
<b>Judgment</b>		
Condition 1	10.96(9.78)	
Condition 2	11.80(10.33)	20.67(13.06)
Condition 3	24.89(21.01)	28.98(27.95)
Condition 4	19.39(18.64)	19.50(22.57)
<b>Constant Error</b>		
Condition 1	1.43(9.78)	
Condition 2	2.27(10.33)	11.14(13.06)
Condition 3	15.36(21.01)	19.45(27.95)
Condition 4	9.86(18.64)	9.97(22.57)
<b>Absolute Error</b>		
Condition 1	5.77(7.99)	
Condition 2	6.64(8.19)	12.42(11.83)
Condition 3	17.43(19.29)	23.92(24.16)
Condition 4	13.04(16.53)	16.55(18.20)

### *Experiment completion times and earnings*

Mean completion time was seven minutes (440 seconds) with a range between two and 24 minutes. (This range is partly explained by the fact that participants were assigned to conditions that took different times to complete.) Three participants were excluded because they took longer than 30 minutes to complete the experiment. (The participant details given in the paper are *after* these three were excluded.) Average hourly earnings were £8.57.

#### 3.1. Robust ANOVAs

Data were analysed in R using robust tests on 20% trimmed means (to reduce skew) and a bootstrap procedure ( $nboot = 2000$ ) to obtain empirically derived critical values ( $p < 0.05$ ) against which test statistics were compared.

Robust one-way ANOVAs (Wilcox, 2017) using function *t1waybt* were conducted to determine whether there was an effect of different types of priming on the main task of producing direct estimates of future inflation. For inflation forecasts and CEs, the effect was significant ( $F_t = 5.43$ ,  $p = 0.005$ ). Pairwise comparisons using *linconb* showed significant differences between condition 1 and condition 3 ( $p = 0.007$ ), between condition 1 and condition 4 ( $p = 0.03$ ), between condition 2 and condition 3 ( $p = 0.009$ ) and marginally

between condition 2 and condition 4 ( $p = 0.05$ ). For AEs, the effect of priming was also significant ( $F_t = 4.56$ ,  $p = 0.02$ ). Pairwise comparisons showed significant differences between condition 1 and condition 3 ( $p = 0.02$ ), between condition 1 and condition 4 ( $p = 0.03$ ), marginally between condition 2 and condition 3 ( $p = 0.05$ ) and marginally between condition 2 and condition 4 ( $p = 0.08$ ).

Robust two-way mixed ANOVA (Wilcox, 2017) using function *bwtrimbt* were carried out on each dependent variable using Task Type (i.e., the three different types of priming task) as a between-participants factor and Estimate Type (Direct versus Indirect) as a within-participant factor to examine the relation between the two types of estimate in the three priming conditions.

The analyses of forecasts and CEs showed main effects of Task Type ( $Q = 4.81$ ,  $p = 0.01$ ) and Estimate Type ( $Q = 10.92$ ,  $p = 0.002$ ), but no interaction between these variables. Pairwise comparisons using *linconb* showed significant difference between condition 2 and condition 3 ( $p < 0.001$ ) and between condition 3 and condition 4 ( $p = 0.02$ ). The analyses of AEs showed main effects of Task Type ( $Q = 6.52$ ,  $p = 0.003$ ) and Estimate Type ( $Q = 10.14$ ,  $p = 0.002$ ), but no interaction between these variables. Pairwise comparisons showed significant difference between condition 2 and condition 3 ( $p < 0.001$ ), between condition 2 and condition 4 ( $p = 0.008$ ) and between condition 3 and condition 4 ( $p = 0.04$ ).

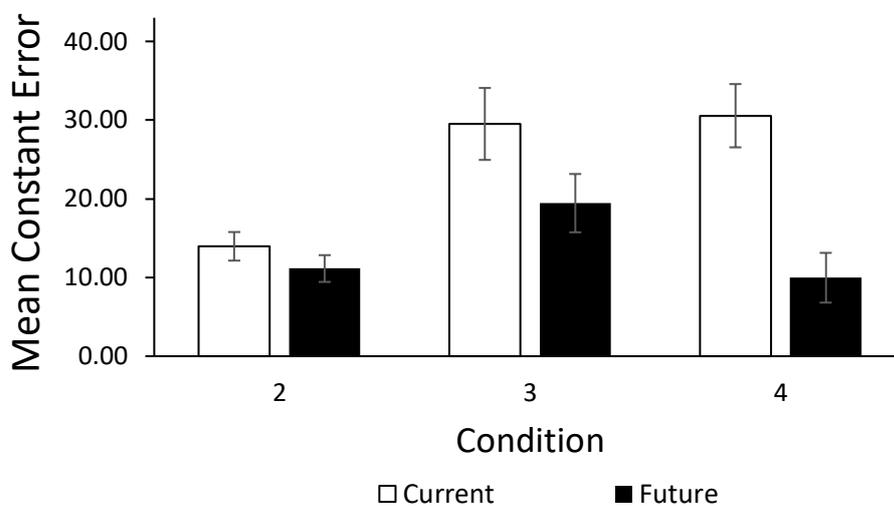
#### *Comparison of CE scores of indirect estimates of current and future inflation*

A robust two-way mixed ANOVAs was carried out on the CE scores of the indirect estimates obtained from the priming tasks with Estimate Type (Current versus Future Inflation) as a within-participant factor and Task Type (the three priming conditions) as a between-participant factor were conducted. Figure SM3 shows a main effect of Estimate Type ( $Q = 8.13$ ,  $p < 0.001$ ), a main effect of Task Type ( $Q = 29.39$ ,  $p < 0.001$ ) and interaction between these variables ( $Q = 6.69$ ,  $p = 0.002$ ). The simple effect of Estimation type showed significance for Condition 3 ( $p = 0.004$ ) and condition 4 ( $p < 0.001$ ). The effect of Task type on CEs was only significant for the indirect estimates of current inflation by using function *t1waybt* ( $F_t = 10.53$ ,  $p < 0.001$ ). Pairwise comparisons revealed significant difference between Condition 2 and Condition 3 ( $p < 0.001$ ) and Condition 2 and Condition 4 ( $p = 0.004$ ).

As in Experiment 2a, greater overestimation of current than of future inflation occurred only in Conditions 3 and 4. In these conditions, participants were primed by estimating price changes for just the product category that they judged had shown the largest price change over the previous 12 months. It appears that thinking of the largest price change in the previous 12 months has a greater priming effect on judging current price changes than on judging future ones. In other words, temporal distance influences the size of the priming effect.

**Figure A3.6**

*Experiment 2b: CE scores of indirect estimates of inflation*



*Self-reported reliance on experience of specific price changes*

**Figure A3.7**

*Experiment 2b: Judged % reliance on past experiences with specific price changes when making direct estimates of the level of future inflation*

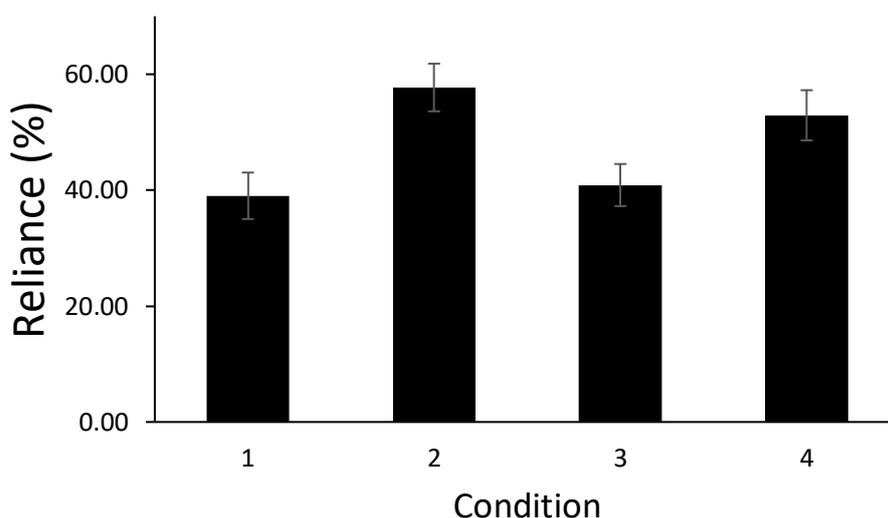


Figure SM4 shows a significant effect of Task Type on reported degree of reliance on specific price changes when directly estimating level of future inflation ( $F(3, 221) = 5.25, p = 0.002, ges = 0.0666$ ). Pairwise comparisons using the BH adjustment method revealed a significant difference between Condition 2 and Condition 1 ( $p = 0.006$ ), between Condition 2 and Condition 3 ( $p = 0.009$ ), and between Condition 1 and Condition 4 ( $p = 0.03$ ). Thus, results

here are similar to those obtained in Experiment 2a except that reported reliance on recall of specific prices was somewhat higher in Condition 4.

#### *Analysis of effects of demographic variables*

We conducted multiple regressions using all six demographical factors (age, gender, income, educational level, marital status, ethnicity) as predictors. CE scores of direct estimates of future inflation were significantly predicted by a model ( $F(6, 214) = 2.38, p = 0.03, \text{Adjusted } R^2 = 0.04$ ) that included education as the sole predictor ( $\beta = 5.27, t(220) = 2.15, p = 0.03$ ). More educated people overestimated future inflation by a greater amount. No model significantly predicted AE scores of direct estimates of future inflation.

No overall model significantly predicted either the CE or AE scores of indirect estimates of current inflation or the CE scores of indirect estimates of future inflation. Though an overall model including all predictors significantly predicted AE scores of indirect estimates of future inflation ( $F(6, 156) = 2.34, p = 0.03, \text{Adjusted } R^2 = 0.05$ ), none of the individual predictors was significant.

#### **Discussion**

The price-free approach to estimating inflation relies on use of media reports and other secondary sources. In other words, it relies on descriptions received from other people. Most decision making research up to now has been based in paradigms that require decisions from descriptions (e.g., Kahneman and Tversky, 1979). In contrast, the price-recall approach to estimating inflation relies on use of personal experience accumulated over time. Recently, there has been increased interest in how people make decisions from experience (e.g., Ert and Erev, 2007; Yechiam and Rakow, 2012). In particular, a number of studies have examined whether there are differences between the characteristics of decisions from experience and the characteristics of decisions from description (e.g., Camilleri and Newell, 2011; Fantino and Navarro, 2012). It is possible that any such differences are reflected in differences in inflation estimates produced by the price-free and price-recall approaches. In addition, under certain conditions, people may make decisions based on both descriptions and experience (e.g., Weiss-Cohen et al., 2016; 2018). Hence, we recognise that there may be times, particularly during shifts in the inflation environment, when estimates of inflation are based on both personal experience and on media reports and other secondary sources.

## References

- Camilleri, A. R., & Newell, B. R. (2011). Description- and experience-based choice: Does equivalent information equal equivalent choice? *Acta Psychologica, 136*, 276–284. <http://dx.doi.org/10.1016/j.actpsy.2010.11.007>.
- Ert, E., & Erev, I. (2007). Replicated alternatives and the role of confusion, chasing, and regret in decisions from experience. *Journal of Behavioral Decision Making, 20*, 305–322. <http://dx.doi.org/10.1002/bdm.556>
- Evans, J. St. B. T. (2008). Dual processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255-278. <https://doi.org/10.1146/annurev.psych.59.103006.093629>
- Evans, J. St. B. T. (2010). Intuition and reasoning: A dual-process perspective. *Psychological Inquiry, 21*(4), 313-326. <https://doi.org/10.1080/1047840X.2010.521057>
- Fantino, E., & Navarro, A. (2012). Description-experience gaps: Assessments in other choice paradigms. *Journal of Behavioral Decision Making, 25*, 303–314. <http://dx.doi.org/10.1002/bdm.737>
- Kahneman, D. (2011). *Thinking, fast and slow*. London: Penguin Books Ltd.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–292. <http://dx.doi.org/10.2307/1914185>
- Liao, K., Zhang, Y., Lei, H., Peng, G., & Kong, W. (2022). A comparative analysis of the effects of objective and self-assessed financial literacy on stock investment return. *Frontiers in Psychology, 13*, Article 842277. <https://doi.org/10.3389/fpsyg.2022.842277>
- Lusardi, A., & Mitchell, O. S. (2007). Financial literacy and retirement planning: New evidence from the Rand American Life Panel. *University of Michigan Retirement Research Center Research Paper, WP 2007-157*. <http://dx.doi.org/10.2139/ssrn.1095869>
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41*(1), 203-212. <https://doi.org/10.1016/j.jrp.2006.02.001>
- Weiss-Cohen, L., Konstantinidis, E., Speekenbrink, M., & Harvey, N. (2016). Incorporating conflicting descriptions into decisions from experience. *Organizational Behavior and Human Decision Processes, 135*, 55–69. <http://dx.doi.org/10.1016/j.obhdp.2016.05.005>
- Weiss-Cohen, L., Konstantinidis, E., Speekenbrink, M., & Harvey, N. (2018). Task complexity moderates the influence of descriptions in decisions from experience. *Cognition, 170*, 209–227. <https://doi.org/10.1016/j.cognition.2017.10.005>
- Wilcox, R. R. (2017). *Introduction to robust estimation and hypothesis testing (4th ed.)*. Academic press.

Yechiam, E., & Rakow, T. (2012). The effect of foregone outcomes on choices from experience. *Experimental Psychology*, *59*, 55–67. <https://doi.org/10.1027/1618-3169/a000126>

### Appendix 3

**Table A4.1**

*Chapter 4: Demographical statistics for participants whose data were analysed in three experiments (percentages or standard deviations in parentheses)*

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

*Note.* one participant in Experiment 5 did not report her age.

## Appendix 4

Chapter 5: This appendix includes tables of summarized data showing inflation judgments and confidence judgments on each trial in each session in Experiment 6 (Table A5.1) and Experiment 7 (Tables A5.2 and A5.3). Full data for each participant in both experiments are available at <https://osf.io/k8vx9>

**Table A5.1**

*Experiment 6: Means and standard deviations (in parentheses) of inflation judgments and confidence judgments*

Inflation judgments		Trial										mean
		1	2	3	4	5	6	7	8	9	10	
Feedback	Session 1	2.34(1.52)	1.96(1.11)	1.95(1.48)	1.67(1.29)	2.15(1.45)	1.86(1.05)	1.50(1.04)	1.94(1.47)	2.36(1.73)	2.29(1.71)	2.00(0.42)
	Session 2	2.44(1.50)	1.77(1.19)	2.02(1.24)	1.77(1.52)	2.13(1.47)	1.65(1.15)	1.82(1.21)	1.75(1.16)	2.22(1.32)	2.28(1.43)	1.98(0.34)
	mean	2.39(0.98)	1.87(0.79)	1.99(0.99)	1.72(0.92)	2.14(0.92)	1.76(0.76)	1.66(0.80)	1.85(0.88)	2.29(1.01)	2.28(1.03)	1.99(0.22)
No-feedback	Session 1	2.64(1.71)	2.11(1.27)	2.14(1.51)	2.37(1.51)	1.81(1.04)	1.84(1.14)	2.13(1.21)	2.09(1.40)	2.26(1.51)	2.29(1.20)	2.17(0.40)
	Session 2	2.34(1.46)	1.96(1.48)	1.98(1.25)	2.14(1.22)	2.15(1.52)	2.32(1.47)	2.28(1.53)	2.08(1.23)	2.04(1.52)	2.10(1.06)	2.14(0.42)
	mean	2.49(1.15)	2.04(0.97)	2.06(1.04)	2.25(0.96)	1.98(0.94)	2.08(0.95)	2.20(1.10)	2.09(0.88)	2.15(0.97)	2.20(0.62)	2.15(0.28)
Confidence judgments												
Feedback	Session 1	53.04(20.31)	51.72(19.51)	49.39(21.28)	47.09(22.04)	46.37(20.30)	49.76(23.17)	49.76(24.86)	45.09(24.87)	44.50(24.73)	49.89(24.40)	48.66(18.91)
	Session 2	47.78(22.37)	46.74(23.01)	47.33(22.65)	47.35(21.40)	46.20(23.53)	48.43(21.73)	46.85(24.46)	46.43(22.70)	48.65(23.33)	49.17(23.98)	47.49(20.20)
	mean	50.41(17.05)	49.23(19.64)	48.36(20.61)	47.22(19.88)	46.28(20.44)	49.10(20.75)	48.30(23.43)	45.76(22.59)	46.58(22.45)	49.53(22.52)	48.08(19.19)
No-feedback	Session 1	55.80(22.67)	58.76(22.24)	55.15(25.51)	56.12(21.98)	54.71(19.52)	59.39(24.95)	59.17(25.34)	52.71(23.32)	57.27(21.64)	52.88(24.90)	56.20(18.91)
	Session 2	60.17(21.39)	58.61(21.83)	59.41(23.12)	62.37(22.46)	58.27(23.35)	61.41(21.79)	56.05(24.98)	63.66(21.25)	60.49(22.94)	59.41(24.88)	59.99(19.15)
	mean	57.99(20.58)	58.68(20.43)	57.28(21.80)	59.24(19.55)	56.49(18.49)	60.40(20.34)	57.61(22.14)	58.18(19.88)	58.88(20.29)	56.15(22.54)	58.09(18.50)

**Table A5.2***Experiment 7: Means and standard deviations (in parentheses) of inflation judgments*

Simple OFB		Trial										mean
		1	2	3	4	5	6	7	8	9	10	
No-feedback	Session 1	2.19(1.43)	2.07(1.40)	1.80(0.99)	1.95(1.15)	2.48(1.59)	1.94(1.07)	1.86(1.14)	2.23(1.70)	2.26(1.23)	2.97(1.89)	2.17(0.38)
	Session 2	2.38(1.72)	1.99(1.36)	2.08(1.21)	2.19(1.45)	2.27(1.60)	2.16(1.43)	2.02(1.36)	2.26(1.40)	2.44(1.75)	2.14(1.40)	2.19(0.39)
	mean	2.28(1.25)	2.03(1.02)	1.94(0.85)	2.07(0.97)	2.38(1.26)	2.05(0.84)	1.94(0.94)	2.24(1.12)	2.35(1.05)	2.55(1.19)	2.18(0.26)
2-feedback	Session 1	2.33(1.61)	1.91(1.04)	2.32(1.47)	1.94(1.26)	2.12(1.45)	1.94(1.16)	1.93(1.13)	2.20(1.47)	2.16(1.04)	2.02(1.36)	2.09(0.36)
	Session 2	1.93(1.26)	2.12(1.42)	2.16(1.33)	2.02(1.18)	2.02(1.26)	2.16(1.62)	1.87(1.08)	2.31(1.54)	1.86(1.08)	2.02(1.11)	2.05(0.41)
	mean	2.13(0.87)	2.01(0.88)	2.24(0.83)	1.98(0.86)	2.07(0.94)	2.05(0.99)	1.90(0.72)	2.25(0.99)	2.01(0.74)	2.02(0.87)	2.07(0.20)
5-feedback	Session 1	2.13(1.31)	1.99(1.07)	2.06(1.25)	1.94(1.34)	1.98(1.32)	1.92(1.25)	1.85(1.30)	2.09(1.38)	2.08(1.29)	1.84(1.36)	1.99(0.39)
	Session 2	1.83(1.23)	1.89(1.74)	1.82(1.23)	1.99(1.20)	2.04(1.37)	2.53(1.35)	1.91(1.39)	2.12(1.58)	1.68(1.17)	2.21(1.49)	2.00(0.41)
	mean	1.98(0.93)	1.94(0.88)	1.94(0.89)	1.97(0.80)	2.01(0.95)	2.22(0.77)	1.88(0.95)	2.11(1.07)	1.88(0.72)	2.02(1.08)	2.00(0.20)
10-feedback	Session 1	2.13(1.48)	1.96(1.40)	1.86(1.41)	2.09(1.44)	2.38(1.64)	2.09(1.39)	1.89(1.61)	2.30(1.69)	1.90(1.19)	2.09(1.38)	2.07(0.44)
	Session 2	2.12(1.62)	2.02(1.31)	1.85(1.06)	1.98(1.39)	1.64(0.76)	2.02(1.33)	2.58(1.50)	2.07(1.30)	2.03(1.28)	2.06(1.47)	2.04(0.40)
	mean	2.13(0.89)	1.99(0.85)	1.86(0.87)	2.04(1.02)	2.01(0.94)	2.06(0.84)	2.23(1.13)	2.19(1.00)	1.97(0.79)	2.07(1.02)	2.05(0.25)
<b>Summarized OFB</b>												
No-feedback	Session 1	2.09(1.34)	2.33(1.51)	2.58(1.48)	1.96(1.28)	1.97(1.41)	2.16(1.11)	1.99(1.15)	2.11(1.16)	1.66(1.11)	2.11(1.60)	2.10(0.35)
	Session 2	2.40(1.54)	2.19(1.40)	2.15(1.43)	1.76(1.37)	2.27(1.42)	2.02(1.40)	2.34(1.38)	2.21(1.53)	2.14(1.44)	2.13(1.33)	2.16(0.40)
	mean	2.25(0.92)	2.26(0.97)	2.36(0.80)	1.86(0.87)	2.12(0.96)	2.09(0.79)	2.17(0.92)	2.16(0.92)	1.90(1.00)	2.12(1.03)	2.13(0.19)
2-feedback	Session 1	1.92(1.27)	1.93(1.07)	2.11(1.46)	1.98(1.29)	2.31(1.49)	1.88(1.12)	1.94(1.09)	1.95(1.27)	2.22(1.44)	2.23(1.49)	2.05(0.35)
	Session 2	1.87(1.02)	1.97(1.20)	2.05(1.25)	1.97(1.11)	1.97(1.20)	2.02(1.30)	2.07(1.38)	1.93(1.37)	2.14(1.35)	2.26(1.55)	2.03(0.35)
	mean	1.89(0.87)	1.95(0.69)	2.08(0.95)	1.98(0.81)	2.14(0.95)	1.95(0.85)	2.00(0.80)	1.94(1.06)	2.18(0.91)	2.25(1.15)	2.04(0.18)
5-feedback	Session 1	2.18(1.31)	2.28(1.40)	2.10(1.30)	1.88(1.38)	2.01(1.52)	1.92(1.33)	2.03(1.11)	2.11(1.41)	2.06(1.34)	1.80(1.03)	2.04(0.35)
	Session 2	1.91(1.19)	1.97(1.34)	2.18(1.47)	2.17(1.40)	1.87(1.23)	2.03(1.39)	2.13(1.44)	2.02(1.50)	2.28(1.53)	2.09(1.32)	2.07(0.36)
	mean	2.05(0.77)	2.12(0.95)	2.14(0.88)	2.02(0.96)	1.94(0.90)	1.98(1.02)	2.08(0.87)	2.07(1.06)	2.17(1.03)	1.95(0.84)	2.05(0.21)
10-feedback	Session 1	2.25(1.46)	2.41(1.54)	2.00(1.22)	2.14(1.62)	1.98(1.09)	1.80(1.31)	2.00(1.22)	2.15(1.45)	1.94(1.42)	2.19(1.25)	2.09(0.34)
	Session 2	2.06(1.10)	2.00(1.38)	1.86(1.31)	2.12(1.22)	2.35(1.51)	2.15(1.62)	1.97(1.31)	2.14(1.55)	2.10(1.43)	1.88(1.14)	2.06(0.33)
	mean	2.16(1.01)	2.20(1.04)	1.93(0.91)	2.13(1.10)	2.16(0.93)	1.98(0.94)	1.98(0.88)	2.15(1.15)	2.02(0.98)	2.03(0.94)	2.07(0.20)

**Table A5.3***Experiment 7: Means and standard deviations (in parentheses) of confidence judgments*

		Trial										
<b>Simple OFB</b>		1	2	3	4	5	6	7	8	9	10	mean
No-feedback	Session 1	51.54(21.52)	55.04(19.17)	51.85(21.23)	53.60(17.76)	53.83(20.71)	52.21(20.86)	57.17(21.85)	56.94(20.94)	54.92(23.28)	54.38(21.81)	54.15(17.20)
	Session 2	54.38(22.41)	56.88(21.00)	56.81(19.65)	54.17(21.12)	54.81(20.32)	53.17(20.52)	53.54(21.81)	55.81(20.67)	50.96(22.57)	56.25(21.30)	54.68(18.13)
	mean	52.96(19.51)	55.96(17.66)	54.33(18.62)	53.89(16.90)	54.32(19.08)	52.69(18.41)	55.35(20.15)	56.38(19.12)	52.94(20.57)	55.31(20.37)	54.41(17.29)
2-feedback	Session 1	50.00(21.11)	50.91(21.70)	50.33(21.44)	51.20(20.33)	50.70(20.61)	49.63(21.57)	49.04(21.94)	48.74(22.27)	52.26(22.00)	50.70(22.99)	50.35(19.08)
	Session 2	52.17(20.26)	50.78(22.31)	51.89(21.45)	47.65(20.38)	51.80(19.76)	48.24(23.68)	50.30(20.65)	49.52(22.15)	53.37(21.70)	51.83(20.18)	50.76(18.99)
	mean	51.09(19.03)	50.85(20.21)	51.11(19.38)	49.42(18.58)	51.25(19.24)	48.93(21.25)	49.67(20.19)	49.13(21.12)	52.82(20.98)	51.26(20.88)	50.55(18.69)
5-feedback	Session 1	53.60(17.52)	50.55(16.70)	51.00(17.84)	52.74(15.39)	47.83(18.08)	50.45(17.98)	50.96(16.82)	51.96(20.20)	51.79(20.02)	53.57(19.74)	51.44(14.01)
	Session 2	52.09(17.82)	50.15(17.31)	48.81(16.32)	49.64(19.49)	47.43(17.61)	50.17(18.76)	50.43(19.52)	50.81(19.11)	49.00(19.85)	51.51(19.07)	50.00(15.39)
	mean	52.84(15.65)	50.35(14.10)	49.90(14.74)	51.19(15.51)	47.63(16.04)	50.31(16.73)	50.69(15.70)	51.38(17.02)	50.39(18.56)	52.54(18.22)	50.72(14.20)
10-feedback	Session 1	52.21(20.42)	47.70(20.51)	46.81(18.96)	47.98(21.11)	47.47(20.51)	46.74(20.50)	49.49(21.02)	46.72(22.80)	45.47(22.07)	46.58(22.39)	47.72(16.35)
	Session 2	49.47(21.68)	50.14(20.87)	46.72(23.42)	45.21(20.33)	49.14(22.03)	48.98(22.99)	48.53(20.39)	47.98(21.73)	47.72(21.02)	45.84(22.40)	47.97(18.71)
	mean	50.84(18.20)	48.92(17.97)	46.77(18.76)	46.59(16.85)	48.30(17.74)	47.86(18.82)	49.01(19.40)	47.35(20.06)	46.59(19.78)	46.21(20.53)	47.84(16.79)
<b>Summarized OFB</b>												
No-feedback	Session 1	60.43(17.47)	55.43(22.20)	55.55(21.31)	52.28(21.34)	52.60(22.75)	54.64(18.34)	56.38(20.89)	52.06(20.27)	54.51(20.91)	52.96(22.07)	54.68(17.57)
	Session 2	58.11(23.40)	57.55(25.00)	58.55(25.46)	55.91(23.87)	56.85(24.35)	56.96(24.09)	55.47(24.99)	57.70(26.19)	57.11(27.34)	52.47(27.19)	56.67(23.61)
	mean	59.27(18.94)	56.49(22.08)	57.05(21.07)	54.10(19.96)	54.72(22.09)	55.80(18.69)	55.93(21.39)	54.88(21.26)	55.81(22.66)	52.71(22.51)	55.68(19.85)
2-feedback	Session 1	57.56(21.51)	55.15(21.44)	53.23(24.01)	56.65(22.08)	54.37(21.84)	54.63(21.85)	54.37(21.41)	55.33(22.75)	55.12(22.41)	54.12(22.26)	55.05(18.53)
	Session 2	58.40(22.15)	55.92(21.28)	56.35(20.53)	57.06(21.66)	57.31(21.25)	55.25(21.54)	54.60(22.15)	57.46(20.86)	57.65(22.17)	54.81(23.24)	56.48(18.95)
	mean	57.98(20.20)	55.54(19.45)	54.79(19.74)	56.86(20.73)	55.84(19.60)	54.94(18.71)	54.48(19.66)	56.39(20.97)	56.38(21.28)	54.46(20.73)	55.77(18.44)
5-feedback	Session 1	53.70(17.82)	50.44(17.11)	51.11(18.02)	47.26(17.38)	50.52(19.99)	51.07(18.03)	51.28(18.97)	51.11(19.66)	53.19(19.04)	47.57(18.24)	50.73(15.07)
	Session 2	52.09(17.55)	50.11(19.48)	49.91(20.35)	52.06(22.07)	48.98(20.85)	51.87(21.38)	47.04(20.09)	48.56(21.80)	49.56(19.94)	48.50(21.62)	49.87(17.58)
	mean	52.90(15.67)	50.28(16.46)	50.51(17.20)	49.66(17.92)	49.75(18.11)	51.47(17.99)	49.16(16.98)	49.83(17.88)	51.37(17.67)	48.04(18.37)	50.30(15.90)
10-feedback	Session 1	55.94(19.10)	52.00(17.06)	50.04(18.43)	47.12(21.01)	52.61(21.08)	51.06(21.94)	49.94(23.54)	53.78(24.74)	50.71(21.53)	50.14(21.82)	51.33(16.51)
	Session 2	51.96(21.47)	51.78(22.18)	50.49(22.67)	51.29(22.32)	53.02(22.92)	55.41(23.78)	54.37(23.71)	52.37(24.22)	54.16(22.10)	51.65(23.82)	52.65(20.76)
	mean	53.95(16.96)	51.89(17.02)	50.26(18.33)	49.21(19.95)	52.81(20.56)	53.24(21.40)	52.16(21.38)	53.08(21.89)	52.43(20.62)	50.89(21.34)	51.99(18.22)

## Appendix 5

Chapter 5: Comparisons of error scores in each experiment with two simple algorithmic models.

**Table A5.4**

*Means and standard deviations (in parentheses) of different types of inflation forecast over 20 countries*

Inflation forecasts	Judgment level	Absolute error	Constant error	Variable error
Linear regression model	1.61(1.64)	0.90(1.15)	-0.17(1.47)	0.91(1.13)
Mean model	2.31(1.86)	0.75(0.67)	0.53(0.86)	0.65(0.54)
Judgments in Experiment 6	2.07(1.38)	0.88(0.86)	0.29(1.19)	0.73(0.68)
Judgments in Experiment 7	2.07(1.36)	0.85(0.83)	0.30(1.15)	0.71(0.65)

*Note.* This table summarizes the accuracy levels of judgmental forecasts in Experiment 6 and Experiment 7 and of forecasts from two simple algorithmic models a) Linear regression model: each forecast was produced by regressing the 10 historical data points for a country and using the model to predict the inflation rate for 2019, b) Mean model: each forecast was produced by extracting the mean of the 10 years' historical inflation data points for a country and using it as the forecast for 2019. (The mean of actual inflation rates in 2019 for the 20 countries was 1.77% with the SD of 1.72%.)

## Appendix 6

### Chapter 6. Experiment 8: Uncertainty estimation in interval and density forecasting

It is well-established that interval forecasts tend to be much too narrow, thus implying that forecasters are highly overconfident. For example, Hansson, Juslin, et al. (2008) point out: “If people, for example, produce intuitive 90% confidence intervals for unknown quantities, the percentage of intervals that include the true value is often closer to 40% or 50% than to the normatively expected 90% (see Block & Harper, 1991; Lichtenstein, Fischhoff, & Phillips, 1982; Russo & Schoemaker, 1992; Soll & Klayman, 2004)”.

In the interval condition of my experiment in which people were asked to set a 90% interval for each of 10 inflation series, the outcome should be within the set interval in nine of those 10 series. In fact, however, outcomes were within the set interval in an average of only 5.265 series out of 10 (52.65%). A one-sample t-test showed that this was significantly below 90% ( $t(41) = 9.67$ ;  $p < 0.001$ ). This result therefore replicates the overconfidence found in the previous work cited above.

Hansson, Juslin, et al.'s (2008) experiments showed that if, instead of setting an interval for a given probability, people were asked to provide a probability for a given interval, overconfidence all but vanished. In the density condition, I required people to provide a probability equivalent (i.e., a number of pounds sterling out of a maximum of £100) for each of a number of intervals (i.e., bins). As this corresponds closely to the task examined by Hansson, Juslin, et al. (2008), I expected little, if any, overconfidence to be present.

For each series, I first calculated the optimal forecast; in the seven un-trended series, this corresponded to the series mean; in the three series with shallow trends, it corresponded to an extrapolation of the trend to the period to be forecasted. After excluding one series (Kiribati) because points were not normally distributed around the mean or trend line, I calculated the 90% prediction interval for each series. The judged probability of the outcome being within this interval should be 90%. In other words, the model fitted to each participant's data on each trial should show that £90 out of the available £100 was allocated to that interval. In fact, I found that, on average, a total sum of £82.92 out of the £100 was allocated to the 90% interval. A one-sample t-test showed that this mean value was different from £90 ( $t(30) = 5.90$ ;  $p < 0.001$ ). Thus some overconfidence was still present in the density forecasting group.

A two-sample t-test showed that the mean probability equivalent assigned to the 90% interval in the density forecasting condition (i.e., 82.92%) was significantly different from the mean probability of outcomes appearing within the judged 90% interval in the interval condition (i.e., 52.65%):  $t(48.69) = 7.48$ ;  $p < 0.001$ . Thus, although interval and density forecasts do not differ in terms of the accuracy with which they provide an estimate of the expected value of the point to be forecast, density forecasts provide a better estimate of the uncertainty associated with the prediction.

## Appendix 7

**Table A6.1**

*Chapter 6: A summary table of the hypotheses and results in the three experiments*

Hypotheses	Experiment	Results
H <sub>1</sub> : Point forecast = Mid-point of the interval forecast = Mean of the density forecast	Experiment 8, Experiment 9	Supported
H <sub>2</sub> : ME > 0	Experiment 8, Experiment 9	Supported
H <sub>3</sub> : ME <sub>Point</sub> = ME <sub>Interval</sub> (= ME <sub>Density</sub> )	Experiment 8, Experiment 9	Supported
H <sub>4</sub> : VE <sub>Point</sub> > VE <sub>Interval</sub> (> VE <sub>Density</sub> )	Experiment 8, Experiment 9	Partially supported: VE <sub>Point</sub> > VE <sub>Interval</sub> = VE <sub>Density</sub>
H <sub>5</sub> : RMSE <sub>Point</sub> > RMSE <sub>Interval</sub> > RMSE <sub>Density</sub>	Experiment 8	Partially supported: RMSE <sub>Point</sub> > RMSE <sub>Interval</sub> = RMSE <sub>Density</sub>
H <sub>6</sub> :  VE <sub>Interval second</sub> – VE <sub>Interval first</sub>   >  VE <sub>Point second</sub> – VE <sub>Point first</sub>	Experiment 9	Not supported
H <sub>7</sub> : Confidence <sub>Point Forecasts</sub> < Confidence <sub>Interval Forecasts</sub>	Experiment 9	Supported
H <sub>8</sub> : VE of average of two point forecasts < Average VE of two point forecasts	Experiment 10	Supported
H <sub>9</sub> : RMSE of average of two point forecasts < Average RMSE of two point forecasts	Experiment 10	Supported
H <sub>10</sub> : VE of average of two point forecasts = VE of the mid-point of interval forecast bounds	Experiment 10	Supported
H <sub>11</sub> : RMSE of average of two point forecasts = RMSE of the mid-point of interval forecast bounds	Experiment 10	Not supported
H <sub>12</sub> : Average VE of two point forecasts > VE of the mid-point of interval forecast bounds	Experiment 10	Supported
H <sub>13</sub> : Average RMSE of two point forecasts > RMSE of the mid-point of interval forecast bounds	Experiment 10	Supported
H <sub>14</sub> : The effect identified in H <sub>8</sub> and H <sub>9</sub> will be greater with noisier inflation series	Experiment 10	Supported
H <sub>15</sub> : The effect identified in H <sub>12</sub> and H <sub>13</sub> will be greater with noisier inflation series	Experiment 10	Supported only for H <sub>13</sub>