Getting through:
Communicating complex central bank messages

Michael McMahon,[*] Matthew Naylor[†]

August 2023

PRELIMINARY:
DO NOT CITE WITHOUT AUTHOR'S PERMISSION.

Abstract

Policymakers communicate complex messages to multiple audiences; we investigate how complexity impacts messages 'getting through' effectively. We formalise the argument for simplicity in a rational inattention model; complexity reduces attention to communication leading to poorly anchored expectations. We show that recent Bank of England efforts to simplify its communications have reduced traditional measures of 'semantic' complexity, but 'conceptual' complexity, captured by a novel measure that we construct, declines less uniformly. Experimental evaluation indicates that conceptual, not semantic, complexity is what matters for getting through. This is true even for individuals with economics degrees suggesting conceptual complexity matters for all audiences.

Keywords: Central bank communications, linguistic complexity, rational inattention, inflation expectations
JEL classification: C83, E58, E61, E70

[*]University of Oxford, CEPR, and Irish Fiscal Council. Email: michael.mcmahon@economics.ox.ac.uk
[†]Corresponding author. University of Oxford. Email: matthew.naylor@economics.ox.ac.uk.
We thank Ambrogio Cesa-Bianchi, Jonathan Haskell, Simon Lloyd, Alastair Macaulay, Catherine Mann, Rebecca Maule, and Silvana Tenreyro for helpful comments at various stages of the paper. We are also grateful to seminar participants at Bank of England, EEA-ESEM 2023, EconDAT (KCL) 2023, Money Macro and Finance (Portsmouth) 2023, and Money-Macro-Finance (Warsaw) 2023. We also thank Abel Hofflin and Hannah Slade Walker for their excellent research assistance. The views expressed in this paper are ours, and not necessarily those of the Irish Fiscal Council. All errors are our own.

"I think our challenge is to speak in plain English as opposed to in a high-tech scientific language which only about half a dozen people understand and even less are interested in" Adrian Orr, RBNZ Governor (2018)

1 Introduction

Central banks and other policy institutions often have to communicate inherently complex messages to a range of different audiences. But can these complex messages be communicated in a way that is accessible? Do (and can) they 'get through'? These have become pertinent questions for central banks in recent years, as communication has increasingly become a key tool in the central bank policy shed. Broadly, there are two primary and very distinct intended audiences: financial markets; and the general public. Evidence indicates that communications with the former have been largely successful in shaping medium- and long-term market expectations for inflation and interest rates as intended (**??**). However, in relation to the latter, evidence indicates that the general public's understanding even of the central bank's role in the economy, let alone its monetary policy strategy, remains very limited. Yet, households and firms (henceforth, the general public) form expectations in similar ways (**???**) and are key actors in the macroeconomy. Their consumption and saving, and wage- and price-setting behaviours directly impact the inflationary environment, and thus, central banks' ability to successfully meet their primary price stability function.

So, why aren't central bank messages getting through to the public? One possible explanation is that, in low inflation economies, the benefits to the average member of the general public of paying attention to information about the macroeconomic outlook and monetary policy strategy are, frankly, very low. Whether the inflation rate is expected to be 1.5% or 2% in a years' time is unlikely to make much difference to day-to-day economic decisions. However, evidence suggests that inflation expectations across the general public have not become significantly better anchored in the high inflationary environment since 2022.[1] Another possible explanation is that the cost of paying attention to and trying to understand central bank communications is high. Publications have typically been relatively complex, with high economic literacy required to fully process messages communicated. Taken together, in a world in which attention is a scarce resource, it may be *rational* for the public to be inattentive (**?**).

In this paper, we seek to draw insights into this question through the lens of this cost channel. We do so with three primary contributions. First, we develop a simple rational inattention model, based on **?**, that embeds complexity of communications in an individual's optimal attention allocation decision. We show that linguistic complexity reduces the degree to which people are willing to pay attention to messages communicated by the central bank, and thus, the degree to which they form accurate beliefs about the economy and well-anchored expectations.

Second, we seek to broaden our understanding of what linguistic complexity actually is, by constructing novel quantitative measures of it. To date, both the research literature and policy institutions have focused primarily on rather restrictive measures that capture only narrow 'semantic' dimensions of complexity (e.g. average word and sentence length), such as the Flesch-Kincaid score. These measures provide no insight into how complex the *content* of the text may be. Yet, an extensive theoretical literature on information processing stresses the importance of new information conveyed by *specific* words (Attneave 1959, Frank 2012, and Goodkind and Bicknell 2018). We construct the McMahon-Naylor Conceptual Complexity (MNCC) index that seeks to measure the 'conceptual' complexity of text. Utilising a dictionary of economic and financial jargon terms, the MNCC index captures the quantity and breadth of jargon used in a text, as well as the range of different technical topics covered. Focusing on quarterly Bank of England publications, we show

[1] The August 2022 Inflation Attitudes Survey found that fewer than 20% of the UK public expect inflation to be at the 2% target rate in 5-years time.

that efforts to simplify language have been successful if one focuses only on *semantic* dimensions of complexity (e.g. the FK score). However, the *conceptual* complexity of its reports has not followed the same trend-decline, instead having *increased* over the same period for certain publications and demonstrated far greater volatility.

Finally, we test the relative importance of these dimensions of complexity in an experimental study with 1,800 representative members of the public. We randomly assign respondents to hypothetical central bank reports that vary in complexity across 'semantic' and 'conceptual' dimensions. We find that '*conceptual*' complexity, captured by the MNCC index, matters more than 'semantic'. It reduces: (i) respondents' *perceived* understanding of the report they read, (ii) their *actual* understanding of the information conveyed, and (iii) their sentiments towards the central bank (such as trust), with some evidence of a potential 'goldilocks' level of complexity. Moreover, each of these results hold focusing on a sub-sample of highly educated respondents who studied economics at university, with potentially important implications for the effectiveness of communications with a *range* of actors in the economy, not just the general public.

Our findings have important and clear policy implications. If central banks and other policy institutions wish to communicate complex messages effectively and 'get through' to their broad audiences, they should pay close consideration to the complexity of the language they use. Specifically, *conceptual* dimensions of complexity are particularly important, as captured by the novel MNCC index we construct, not only for effective communications with the general public but potentially also for *all* economic agents.

The rest of this paper is structured as follows. In Section 1.1 we review the related literatures and detail how our paper contributes to each. In Section 2, we provide a theoretical argument for simplicity, by developing a simple rational inattention model of central bank communications. In Section 3 we construct our novel measures of *conceptual* complexity, including the McMahon-Naylor Conceptual Complexity (MNCC) index, and apply these to BoE publications. Section 4 then details the experimental study we carry out and the empirical strategy we adopt, before presenting our results in Section 5. We outline our conclusions in Section 6.

1.1 Related Literature

1.1.1 Information processing and linguistic complexity

An established literature exists on information processing, seeking to develop information-theoretic measures to quantify the 'mental effort' required to process information. In particular, work has sought to investigate the cognitive load conveyed by each word in a sentence. A particularly common operationalisation of a word's information content is its 'surprisal' - a theoretical measure of the extent to which a word came unexpected to a reader (**?**). Efforts to measure surprisal in probabilistic language models can be categorised into two strands. The first, characterised by difficulties in constructing accurate models, was able only to assign surprisal values to words' *part-of-speech*, rather than to the words themselves. As more sophisticated models developed, a second strand of work emerged able to capture the surprisal of *actual* words. Both strands find positive correlations between surprisal and word-reading times.

A more recent operationalisation of a word's information content is founded in the principle of (entropy-based) uncertainty reduction (Hale, 2003, 2006, 2011). Specifically, this is the idea that the degree to which a word informs an individual is reflected by the degree to which it reduces an individual's uncertainty about what is being communicated. Models have estimated the impact on word-reading time of entropy reduction, again defined over both *parts-of-speech*, finding this operationalisation also describes cognitive load, independently of 'surprisal'.

A nascent literature has emerged seeking to capture the difficulty of comprehending texts through *quantitative* measures. Specifically, work has sought to measure texts' linguistic complexity, with the primary aim of understanding its association with individuals' capacity to process information

?. The application of these techniques has become increasingly popular across various economic fields that are characterised by high levels of complexity, including regulatory economics,[2] corporate reporting,[3] and central bank communications.

However, complexity is seldom well defined (**?**). To date, the research literature on linguistic complexity has largely restricted attention to a single measure of linguistic complexity: the Flesch-Kincaid score. This is defined as an objective measure of 'readability', determining the number of years of education an individual would need to have, on average, to be able to understand a piece of text. However, as a measure of linguistic complexity, it is restrictive. It focuses only on *semantic* structure, and accounts only for texts' average sentence and word length. Linguistic complexity is far broader than this.

More recently, efforts have sought to broaden focus to a wider variety of dimensions of complexity. Beyond the Flesch-Kincaid score, or simple word count measures (**?**, **?**), **?** use measures of 'conditionality' and 'lexical diversity', to reflect how cognitive loads may increase as individuals are required to consider a greater number of possible states of the world or have to interpret a greater range of words, respectively. However, these measures still do not capture the *content* of the text itself. As the information-theory literature teaches us, the *actual* words used are an important predictor of the cognitive load of text.

It is this gap in the literature that we seek to bridge. In the same way that, as they became more sophisticated, probabilistic language models were better able to measure surprisal and entropy reduction of actual words, developments in semantic modelling and text analysis techniques provide the opportunity to capture the content of *actual* words and text in quantitative measures. We develop measures of what we refer to as 'conceptual' complexity to do exactly this, taking steps towards more effectively evaluating the dimensions of linguistic complexity that the information processing literature indicate matter most.

A related, but separate strand of recent work, largely motivated by **?**, has begun to focus on the role of narratives in reducing the costs of processing information,[4] utilising sophisticated Natural Language Processing (NLP) techniques to identify the specific topics discussed in text, model their relative importance within a text, and draw inferences on their association with information processing.[5] Our work differs from this strand in seeking to provide a quantitative, simple, measure.

1.1.2 Central bank communications and rational inattention

A rapidly growing and evolving literature on central bank (CB) communications has been at the vanguard of policy strategies adopted by CBs over the past two decades. Initial focus concerned the benefits of divulging more information and increasing transparency. As a consensus emerged that this yielded greater CB accountability and promoted the effectiveness of stabilisation policies, CBs all over the world began to move away from historical intentional opacity, instead communicating decisions more openly through quarterly publications, press conferences, and transparent inflation targets around the turn of the century. Andy **?** refers to this as a "revolution" in CB transparency. Epitomising this transition has been the adoption of Forward Guidance policies in the wake of the Financial Crisis. Communications have become a central part of CB policy, with broad academic and policymaker agreement on the importance of clear communications as a tool for anchoring expectations (**????**).

Following this "revolution", considerable empirical evidence suggests that professional forecasters' expectations are indeed well anchored by CBs.(**????**) However, despite growing recognition that

[2] **???**.
[3] **?????????**.
[4] **?**.
[5] See **????**.

CBs have a number of different target audiences,[6] ????, and ? note that relatively little academic attention has concerned the expectations of the general public. ? describes this as symbolising a "selective revolution" within the literature on CB transparency.

CBs' ability to minimise the impact of shocks on the economy in the long-term rests on their capacity to anchor expectations across the economy. Hence, it is surprising that the research literature has given little focus to the formation of household expectations, given that: (i) households account for the largest expenditure component of GDP, and (ii) a growing body of empirical evidence suggests that most retailers, wholesalers, price- and wage- setters, form expectations in a way which closely resemble that of households, rather than that of professional forecasters (???).

Indeed, preliminary evidence suggests that CB communications are *not* getting through to the general public. ? argue that public understanding of monetary policy has remained largely immune to the increased quantity of communication by CBs. Empirical evidence from the Netherlands (?), US (????), New Zealand (?), South Africa (?), UK (?), Japan (Bank of Japan, 2005), and Eurozone (European Commission, 2016) indicate low levels of informedness across both households and firms in relation to monetary policy. ? finds that US household inflation expectations are far less anchored than are those of professional forecasters, arguing that CBs face starkly different challenges in seeking to anchor expectations of different audiences. Whilst the anchoring of financial market expectations is an issue of *credibility*, anchoring household expectations is more a problem of *informedness*.

This lack of informedness may be explained by rational inattention (??). As first postulated by ???, economic agents are constrained in the amount of attention that they can devote to different sources of information. People choose to pay attention to things that are important to them. ? explains that "while financial market participants and professional forecasters are likely to be very attentive to even the smallest change in the policy statement, the effects may be very different regarding individuals." High perceived costs and low perceived benefits limit households' attentiveness to monetary policy communications (?). Consistent with this, Coibion et al. (2018) find that whilst economic agents in low inflation countries are "remarkably inattentive", those in high inflation and unstable environments, where benefits are greater, pay considerable attention to inflation.[7] On the cost side, ? explains that households face high costs from low economic literacy and difficulty in comprehending complex communications. In developing a simple theoretical model showing the importance of simplicity in central bank communications, we join an extensive list of work applying Sims' (2003) rational inattention framework to various economic settings such as optimal price setting (?), consumption saving problems (?), investment decisions (??), settings with elements of price stickiness (?), stochastic choice settings (?), dynamic games (?), global games (?), dynamic learning (?), and social learning (?).

Finally, a nascent but growing empirical literature has also focused on these costs to paying attention to CB communications, seeking to apply the semantic modelling techniques discussed above to measure linguistic complexity.(????????). ? finds that CB publications in the UK, Chile, Czech Republic, ECB, Poland and Sweden require between 14-18 years of schooling to be understood, whereas the average number of years of schooling needed to understand a political speech is about 8, and a broadsheet newspaper 12 (?). Indeed experimental studies have found indicative evidence that the linguistic complexity of CB communications has significant negative effects on the degree of informedness.(???????). Indeed, on the back of this work, and consistent with advice by ? that "there is an argument for guiding the simplification of the policy message," in order to reduce the error with which the public receive policy announcements, CBs have recently sought to reduce the complexity of their messages and communicate more effectively with the public.

[6] The ECB (2002) noted: transparency is "more than simply releasing information; but it must be communicated to different audience across different environments". ? three Fed audiences: political authorities, financial markets and the general public. The Swedish Riksbank identifies eight (?).

[7] Though we show that there is little evidence that the public in the developed world has been more attentive to central bank messages during 2022, despite heightened levels of macroeconomic volatility, uncertainty and inflation.

However, as with the broader linguistic complexity research literature, analysis has largely restricted attention to the Flesch-Kincaid score. Additionally, empirical work has focused primarily on the impact of complexity on the formation of inflation expectations. Our paper extends the research literature by broadening our understanding of the dimensions of complexity and expanding the set of measures used to capture them. We also provide causal evidence on which dimensions matter most, not just in relation to inflation expectations, but people's understanding of information more generally, as well as their sentiments towards the central bank.

2 A Theoretical Argument for Simplicity

If the central bank aims to align beliefs and expectations formed by economic actors more closely to CB forecasts and targets, communications are a potentially important tool. In this section, we formalise the argument for simplicity in communication by developing a simple rational inattention model, based on **?**, that embeds complexity of communication in an individual's optimal attention allocation decision. In a world in which attention is a scarce resource, economic actors will decide how much (if any!) attention to pay to CB communications based on the relative costs and benefits of doing so.

In the model, we focus on formalising the implications of complexity on the degree to which people choose to engage with messages communicated by the CB and, thus, process the information communicated. We show that linguistic complexity reduces the degree to which people pay attention to CB communications, which in turn reduces the accuracy of beliefs and expectations formed based on messages communicated by the CB.

There are two other possible channels through which complexity may also directly impact the accuracy of beliefs and expectations formed: (a) by reducing trust in the central bank (found by **?** to be linked to reduced attention to CB communications), and (b) by *directly* reducing the degree to which people are able to accurately process information, for a given level of attention. For simplicity, we do not also exogenously formalise the *additional* effects on expectations formation of complexity through its impact on trust or directly on information process; incorporating these channels would exacerbate the pervasiveness of the role of complexity on the formation of accurate expectations, as we show in the experimental study detailed in Sections 4 and 5. Additionally, we show in Appendix A.2 that complexity may continue to impact the accuracy of expectations formed by economic actors, even when that information is intermediated through highly trained journalists rather than being read directly from the CB.

2.1 Model Environment

There are households and also a central bank (CB). The CB is charged with minimising the impact of shocks on the economy. Its only tool for doing so is the anchoring of households' expectations to its long-run targets, through the publication of economic reports. We abstract from the direct use of other monetary policy tools, such as interest rate instruments or open market operations. We assume that the CB has perfect information about the state of the economy, while households are imperfectly informed. The CB transmits a message to households detailing the true state of the economy (containing both nowcasts and forecasts), and households optimally choose how much information to pay to this message.

Households, which we assume to be homogenous and characterised by a representative household, h, maximise expected utility, subject to a constraint, by optimising the amount of attention they pay to the signal that they receive in the message transmitted by the CB. Household h's utility function is given by:

$$u_h(x, \tilde{x}_h) = -b(x - \tilde{x}_h)^2 \tag{1}$$

where $\tilde{x}_h$ represents household h's posterior belief and $b > 0$ is the benefit derived from being well informed.[8] Deviations of $\tilde{x}_h$ from x reflect frictions caused by imperfect information which result in sub-optimal choices relative to the counterfactual case of perfect information and a fall in household utility (**?**). A well-informed household has smaller deviations of $\tilde{x}_h$, from the true message, x with b capturing the benefit of being informed about the state of the economy.

The constraint the household faces arises from the fact that attention is scarce and reflects the cost, c_h, associated with paying attention to the CB's message. Specifically, we follow **?** in modelling attention as an information flow and the constraint on attention is modelled as a bound on information flow. Household h's choice of how much attention to pay to the signal received via the message transmitted by the CB is characterised by their choice of how much information to process. The cost of paying attention is characterised by the cost of processing information. The cost is defined as:

$$c_h = \underbrace{(1+\mu)}_{\text{marginal cost}} \cdot \underbrace{\lambda_h}_{\text{quantity}} \tag{2}$$

where μ is the linguistic complexity of the message communicated by the CB, and λ_h is the quantity of information that household h processes. The marginal cost of processing information is assumed to be increasing in the linguistic complexity of the CB's message.

2.2 Household Information Processing

Although the model is static, it is convenient to think of the transmission and receipt of the CB's message and the resulting impact on the household's updated beliefs as consisting of three stages.

Stage 1. Household h has a prior belief $\bar{x}_h$ about the state of the economy.

The household is uncertain about the true state of the economy, but knows the mean and variance of the distribution from which the state of the economy is drawn:[9]

$$x \sim \mathcal{N}(0, \sigma_x^2)$$

where σ_x^2 reflects uncertainty about the state of the economy. The household's prior belief will be (optimally) set equal to the expected state of the economy, such that: $\bar{x}_h = E[x] = 0$.

Stage 2. The CB transmits a message, x, revealing the true state of the economy.[10]

Stage 3. Households receive a noisy signal depending on attention paid to the CB message:

$$s_h = x + \epsilon_h \tag{3}$$

where s_h is the signal received by household h, x is the true state of the economy as transmitted by the CB's message, and $\epsilon_h \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is noise within the signal that is interpreted as arising from household h's limited attention.

Stage 4. The household updates its beliefs about the state of the economy.

Household h uses the noisy signal received from the CB to update its prior, $\bar{x}_h$. Based on the utility function specified in equation (1), the utility maximising rationally inattentive household

[8] I follow **?** in modelling utility as quadratic. This is a necessary condition for Gaussian uncertainty to be the optimal choice of distribution (**?**). **??** and **?** consider more general utility functions.

[9] **?** explains that Gaussian uncertainty is optimal in this setting when agents face quadratic utility with a linear constraint. **?** and **?** corroborate this, finding that for low information flow (as is assumed in this setting), departures from Gaussianity are small with non-quadratic objective functions. In addition, **?** explains that RI models are easiest to handle and lead to particularly appealing and useful properties when random variables are normal.

[10] We assume that the CB wants to anchor households' beliefs and expectations and so the CB will always seek to communicate all information about the state of the economy, with no incentives to withhold information or purposefully limit transparency (see **?**).

will choose to set their posterior belief, $\tilde{x}_h$, equal to the expectation of the true state of the economy as communicated by the CB's message, x, given the signal received, s_h. That is:

$$\tilde{x}_h = E[x|s_h] = \xi_h(x + \epsilon_h) \tag{4}$$

where ξ_h is the weight that household h attaches to the signal.[11]

The weight, ξ_h, that household h attaches to the signal, s_h, characterises the degree to which it pays attention to the signal received from the CB. The greater the weight attached to the signal, the more attention is paid, the closer is its conditional expectation of x given s_h (and, thus its posterior belief) to the true state of the economy. In contrast, the lower the weight attached to the signal, the less attention paid to the signal, the closer h's posterior belief lies to its prior.

2.3 Optimal Choice of Attention

In order to model the household's optimal choice of attention, we must relate the quantity of information processed by household h, λ_h (in equation 2), to the weight attached to the signal received from the CB, ξ_h (in equation 4). We follow much of the literature in doing this by modelling each as reflecting the expected reduction in uncertainty due to the acquisition of the signal. Information processed is captured by: $\lambda_h \equiv H(x) - E[H(x|s_h)]$; where $H()$ is a Shannon entropy function (**????**). Following **?**, we define the weight attached to the signal, ξ_h, as:

$$\xi_h \equiv \left(1 - \frac{\sigma^2_{x|s}}{\sigma^2_x})\right) \tag{5}$$

where $\xi_h \in [0, 1]$. A greater weight, ξ_h, attached to the signal, s_h, by household h results in a greater reduction in uncertainty given the acquisition of the signal. We can interpret $\xi_h = 1$ as reflecting the scenario in which household h pays full attention to the signal and, thus, processes complete and perfect information about the state of the economy. In contrast, $\xi_h = 0$ reflects the scenario in which no attention is paid by household h to the signal and no information is processed.

By the Shannon entropy properties of a Gaussian variable, we can thus relate λ_h to ξ_h:

$$\lambda_h = \frac{1}{2}log\left(\frac{1}{1-\xi_h}\right) \tag{6}$$

Households seek to maximise their expected utility subject to their constraint on attention:

$$\max\left\{E[u_h(x, \tilde{x}_h)] - c_h\right\} \tag{7}$$

which yields optimal weight:[12]

$$\xi^*_h = max\left(0, 1 - \frac{(1+\mu)}{2b\sigma^2_x}\right) \tag{8}$$

We can see that the optimal signal weight, reflecting the optimal level of attention, increases with the benefit of paying attention, b, and the degree of uncertainty surrounding the state of the economy, σ^2_x. In contrast, attention decreases with the linguistic complexity of the CB's message, μ.

The deviation of the posterior belief from the true message communicated by the CB is given by:[13]

$$x - \tilde{x}_h = \frac{(1+\mu)x}{2b\sigma^2_x} - \eta_h \tag{9}$$

where $\eta_h \equiv \xi^*_h\epsilon_h \sim \mathcal{N}(0, \sigma^2_\eta)$ can be interpreted as resulting noise in actions.[14]

[11] Equation 4 follows from the fact that $E[x|s_h] = (1-\xi_h)\bar{x}_h + \xi_h s_h$ and $\bar{x}_h = E[x] = 0$

[12] Derivations are provided in Appendix A.1.

[13] Note that this deviation is zero in expectation by Law of Iterated Expectations.

[14] The variance of the noise in actions, $\sigma^2_\eta = (\xi^*_h)^2\sigma^2_\epsilon$ will be small as high attentiveness implies relatively high ξ^*_h, but relatively low σ^2_ϵ and vice versa. At each extreme, $\sigma^2_\eta = 0$ as $\sigma^2_\epsilon = 0$ in the full attention case, whilst $\xi^*_h = 0$ in the no attention case.

This simple model implies an under-reaction to shocks. This is a standard implication in rational inattention settings. In our setting, $(1+\mu) > 0$ results in *under-reaction* to messages from the CB, with households choosing a weight $\xi_h < 1$ and paying less attention to the message communicated by the CB than in a perfect information setting.

This simple model also produces a qualitative case for simplicity of CB communication. To the extent that linguistic complexity increases the perceived costs of paying attention, the deviation between expectations formed and the true message communicated by the CB (i.e. the 'inaccuracy' of expectations formed) is *increasing* in the degree of linguistic complexity: $\frac{\partial(x-\tilde{x}_h)}{\partial\mu} > 0$.

As mentioned already, we do not model an effect of complexity on trust in the CB and subsequently on expectations formation. We can think of trust as factoring (positively) into the perceived benefits to paying attention, b. Given that we show that the deviation of expectations from the true CB message is strictly decreasing in the benefit to paying attention ($\frac{\partial(x-\tilde{x}_h)}{\partial b} < 0$), complexity would also be associated with lower levels of attention through low trust. We also do not model the direct effect of complexity on information processing capacity *for a given level of attention.* This could be done by also incorporating μ directly in equation (9), exogenously from ξ_h.

One critique of this simple model is that, in the real world, most people do not pay any attention to CB communications and instead get their information about the state of the economy via the media, rather than directly from the CB (**????**). In Appendix A.2, we describe an extension of the model to incorporate a role for the media. The key assumption is that journalists, j, are rationally inattentive and receive the CB signal before transmitting it on to final agents. We show that complexity may continue to impact the accuracy of expectations formed by economic actors, even when that information is first received, simplified, and then transmitted to those actors by highly trained journalists. The implication is that simplified communication may benefit financial market participants (who are likely to get their information directly from the CB) as well as journalists (who then transmit these messages to the broader public).

While the model proposes a formal mechanism through which linguistic complexity affects public engagement with CB communications, we have not been explicit about what complexity is. That is, we have not distinguished between different dimensions of complexity. In fact, the main contribution of this paper is to explore the role that different types of complexity, 'semantic' and 'conceptual'. We turn now to being clear on that distinction.

3 Linguistic Complexity: Categories and Measurement

Complexity is seldom well defined (**?**). The literature typically assumes that textual complexity increases "information processing costs", but generally offers little indication as to what these processing costs are (**??**). Within the field of CB communications, definitions of complexity have been particularly narrow and restrictive. Most empirical work has focused exclusively on a single measure of linguistic complexity: the Flesch-Kincaid (FK) score. In this section, we discuss the limitations of the FK score, propose a categorisation of linguistic complexity by 'semantic' and 'conceptual' dimensions, construct novel measures to capture the latter, and apply these measures to quarterly Bank of England publications.

3.1 Beyond the Flesch-Kincaid Score

The FK score is an objective measure of 'readability'. It determines the number of years of education an individual would need to have, on average, to be able to understand a piece of text. The FK accommodates cross-country (**?**), cross-institution (**?**), and temporal (**?**) comparisons. However, as a measure of linguistic complexity, it is restrictive, focusing only on semantic structure, and accounting only for average sentence and word length of a piece of text. Linguistic complexity is far broader than this.

Significant interest has been expressed in expanding our understanding of the breadth and nature of linguistic complexity. Andy **?**, Chief Economist at the BoE, argues for a shift of focus towards behavioural aspects of information processing. He calls for more investigation on how the use of narratives and the expression of concepts, using terms with which readers can relate, might facilitate the processing of information and enhance engagement by wider audiences. **?** and **?** point to the use of "popular narratives" in other fields: Marketing,[15] Journalism,[16] Education,[17] Health Interventions,[18] and Philanthropy.[19] Despite this, relatively few studies within the field of CB communications have yet extended semantic modelling techniques to analyse measures of complexity beyond the FK score.

Our aim is to extend analysis in this field by developing a broader range of measures of linguistic complexity.

3.2 Categories of linguistic complexity

A number of taxonomies have been proposed to distinguish between forms of complexity. **?**, following **?**, distinguish between 'local' and 'global' complexity. Within the context of CB regulations, local complexity refers to difficulties in processing the language of individual provisions, whilst global complexity refers to the network of provisions. Given that CB communications are relatively self-contained (comprehension does not rely on references to other documents).[20] We restrict analysis to 'local' complexity, and from hereon refer to 'linguistic complexity' as synonymous with 'local' complexity. We further distinguish between two forms of linguistic complexity: 'semantic' and 'conceptual' linguistic complexity. Whilst the former captures the grammatical and semantic structure of a piece of text, the latter instead is determined by the complexity of the content of the text.

Semantic and conceptual forms of complexity capture very different determinants of complexity and imply that different pieces of text can be rendered linguistically complex for different reasons. A document that discusses high level quantum mechanics using very short words and phrases would be regarded as relatively simple by a measure of semantic complexity, yet, in reality, it is so conceptually complex and technical that it is likely incomprehensible for most people.

In the following sections, we detail the existing measures of semantic complexity and construct a novel McMahon-Naylor Conceptual Complexity (MNCC) index, before applying these measures to quarterly BoE publications.

3.3 Measurement of linguistic complexity

3.3.1 Measures of semantic complexity

The two most common measures of 'semantic' complexity are the FK Score and Word Count.

Flesch-Kincaid (FK) Score
First suggested by **?**, the FK score offers a picture of the overall level of semantic complexity of text. It is computed as a composite measure of the average word and sentence length of a piece of text. The formula is given by:

$$\text{Flesch Kincaid Score} = 0.39\frac{n(Words)}{n(Sentences)} + 11.8\frac{n(Syllables)}{n(Words)} - 15.59$$

[15] **?**.
[16] **?**.
[17] **?**.
[18] **?**.
[19] **?**.
[20] Of course, the macro economy more broadly is a complex web of interactions that underlie the MPR. Further research could extend measures of complexity to consider 'global' dimensions as well.

where $nWords$ refers to the total number of words in a piece of text. Analogous definitions hold for $nSentences$ and $nSyllables$.

Word Count
Word Count provides a measure of the length of a piece of text.[21] **?** compute the word count for Fed statements and show a substantial increase under the Yellen chairmanship. In contrast, **?** explains recent BoE strategies to produce shorter forms of communications.

3.3.2 Novel measures of conceptual complexity

Empirical studies have shown that simpler terminology can dramatically increase the readability of text. For instance, **?** explains that terms such as 'inflation', 'employment' and 'annuities' resonate less well with people than do their less technical counterparts, 'prices', 'jobs' and 'investment' (**?**). **?** similarly emphasise the effect that technical jargon might have on increasing the difficulty of understanding information. We address this in this section by constructing a sophisticated measure of conceptual complexity with specific application to macroeconomics and finance.

Simple Measure: Proportion of Jargon (PoJ)
The simplest way to measure a document's conceptual complexity is to capture the number of technical jargon terms used in it. Very simply, the more densely packed a document is with jargon terms, the harder it is likely to be to understand. A simple measure of the proportion of jargon (PoJ) within a document, d, would be:

$$\mathrm{PoJ}_d = \frac{\sum_{j=1}^{J} w_j}{\sum_{i=1}^{N} w_i} \equiv \frac{W_j}{W_i}$$

where w_j represents the number of instances that jargon term $j \in \{1, ..., J\}$ is mentioned, and w_i represents the number of instances that *any* word $i \in \{1, ..., N\}$ is mentioned. Thus, PoJ_d captures the total number of jargon words (W_j) as a fraction of the total number of words (W_i) in document d.[22] We apply this measure to BoE publications in Section **??**, defining 'jargon' terms based on a dictionary that we construct by merging published economic, business, and financial A-Z lists.[23]

However, there are a number of features of conceptual complexity that this simple PoJ measure is unable to capture. For instance, it does not reveal whether a document refers to lots of different jargon words, or simply the same ones repeatedly. A document in which 10% of the words are a single jargon term (e.g. 'growth') is likely to be less complex than a document in which 10% of words refer to different jargon terms (e.g. 'growth', 'GDP', 'activity', 'output'). Furthermore, if *different* jargon terms *are* mentioned, this simple measure does not reveal whether they refer to similar concepts or completely distinct topics. The abovelisted jargon terms each relate to a similar concept, but a document that also discusses a greater range of topics, such as monetary policy, inflation, growth, financial markets, etc. is likely to be more conceptually complex.

We construct a more sophisticated measure of conceptual complexity, which we term the McMahon-Naylor Conceptual Complexity (MNCC) index, that seeks to capture each of these characteristics.

McMahon-Naylor Conceptual Complexity (MNCC) index
The MNCC index has three key features. It increases in:

1. the proportion of jargon used;

[21] **?**.

[22] We drop the subscript d from the right-hand side of the equation throughout this section for notational simplicity.

[23] Sources: Economist, the Guardian and Investopedia. A list of the terms included in this dictionary is provided in Appendix B.1, along with a more detailed discussion of the methods used.

2. the breadth and dispersion of *distinct* jargon terms used *within* a given topic;
3. the number of topics covered.

Feature 1. We use the simple PoJ_d measure described above as the baseline for feature 1. We then augment this measure by incorporating features 2 and 3 in a manner that ensures the MNCC index is still comparable to the simple PoJ_d measure above.

We consider that there are T broad topics relating to the economy. Each jargon term or phrase, j, is mapped to a topic $t \in \{1, ..., T\}$.[24] It is not the case that the words have to be synonyms within the topic; a piece carefully distinguishing between the level and rate of change of prices should appropriately use separate price and inflation terms. Rather, we consider a document that distinguishes between these first and second derivatives to be more conceptually complex than one that just makes reference to one.

Feature 2. We measure the within-topic intensity of jargon using a version of the Herfindahl index of concentration. We create a weight, ψ, that adjusts the jargon count within topic t (in document d) based on the 'concentration' (or, conversely, breadth and dispersion) of *distinct* jargon terms used within that topic.[25] This weight is given by:

$$\psi_{t,d} = \sqrt{\sum_{j_t=1}^{J_t} s_{j,t}^2}$$

where $s_{j,t} \equiv \frac{w_{j,t}}{W_{j,t}}$ represents the share of references, $w_{j,t}$, to jargon term $j_t \in \{1, ..., J_t\}$ in topic t in the total count of references to *all* jargon terms, $W_{j,t} \equiv \sum_{j_t=1}^{J_t} w_{j,t}$, in that topic. The weight, $\psi_{t,d} \in [0, 1]$ is equal to 1 if only a single jargon term j_t is used within topic t. It falls towards zero as more jargon terms within the topic are used, and specifically they are used in a less concentrated (or, equivalently, more dispersed) manner. The weight also treats differentially the use of alternative jargon terms once versus many times, reflecting the diminishing impact on conceptual complexity of using the same jargon term multiple times.

We then use a transformation of this weight, reflecting the 'concentration' of jargon terms, to scale the within-topic t jargon count as follows:

$$W_{j,t,d}^* = \frac{W_{j,t}}{\Psi_t}$$

where $\Psi_t = 2^{log_{10}\psi_t}$. This transformation is chosen such that where $\psi_t = 1$ it is also the case that $\Psi_t = 1$ and the adjusted jargon count $W_{j,t,d}^*$ in topic t is equal to the baseline jargon count $W_{j,t,d}$. As ψ_t decreases, the jargon count is adjusted upwards to reflect the greater conceptual complexity arising from a lower within-topic concentration (or, equivalently, greater breadth and dispersion) of jargon terms. Specifically, as ψ_t reduces by a factor of 10, the transformation is such that Ψ_t reduces by a factor of 2, thereby doubling the within-topic conceptual complexity.[26]

Feature 3. We draw inspiration from the tf-idf (term frequency - inverse document frequency) weighting commonly used in natural language processing, to account for the number of different topics discussed in document d. The 'topic-coverage' weight is given by:

$$\Phi_d = \frac{\log_{10}(T + v)}{\log_{10}(T + v) - \log_{10} T_d}$$

where T is the total number of topics that we distinguish between, T_d is the number of topics covered in the particular document d, and v is a coefficient that allows the user to adjust how extra

[24] We detail this mapping in Appendix B.1.
[25] Again, we drop the notation d from the RHS of the equation for notational simplicity.
[26] This transformation is helpful to avoid exponential increases in $W_{j,t,d}^*$ as $\psi_t \to 0$.

topic coverage is penalised in the weighting. As detailed in Appendix B, we distinguish between 10 topics in total: monetary policy; inflation; output, production, and supply side; private demand (consumption and investment); fiscal policy (including government expenditure); open economy; labour market; fnancial market; financial stability and macroprudential policy; and 'other'). We then choose $v = 90$ such that covering all topics doubles the difficulty and, hence, the adjusted jargon count compared to a baseline of covering only one topic.[27]

Taking all of these adjustments into account, the McMahon-Naylor Conceptual Complexity (MNCC) index is given by:

$$\mathrm{MNCC}_d = \frac{\left(\sum_{t=1}^{T} W^*_{j,t,d}\right) \times \Phi_d}{W_i}$$

That is, the MNCC index for a document d is increasing in the sum of the jargon counts (feature 1), across all topics $t_d \in \{1, ..., T_d\}$ covered in d, adjusted for the breadth and dispersion of *distinct* jargon terms within each topic t, $W^*_{j,t,d}$ (feature 2), and the range of topics covered, given by the topic-coverage weight Φ_d (feature 3). The index is then given as a proportion of the total number of words in the document W_i. $\mathrm{MNCC}_d = \mathrm{PoJ}_d$ if document d refers only to a single jargon term in a single topic. Otherwise, $\mathrm{MNCC}_d > \mathrm{PoJ}_d$ and this difference is increasing the breadth and dispersion of jargon terms used within each topic discussed, and the number of topics of topics covered in document d.

3.4 Application: Complexity of Bank of England publications

CBs communicate in a variety of forms, from press conferences and speeches to quarterly publications. In the UK, the Bank of England (BoE) releases three quarterly publications: the Monetary Policy Report (MPR, formerly 'Inflation Report'), the Monetary Policy Summary (MPS), and the Visual Summary (VS). The MPR, introduced in 1993, is the primary publication, detailing the state of the economy and monetary policy decisions. The MPS is a brief, but technical, summary of the MPR. The VS is a recent innovation, introduced in 2017 Q4, with the objective of conveying communications more simply and targeting broader audiences.[28]

In this section, we apply the complexity measures described and constructed in Section 3.3 to text from the three abovementioned quarterly BoE publications. In order to do so, we first construct a text mining algorithm to generate a novel set of cleaned text data for each of these publications between Q3 2015 and Q3 2023, and combine this with MPR text data mined and shared with us by Hansen & McMahon (2016) to produce a dataset with text for 71 MPRs (Q4 2005 - Q3 2023), 37 MP Summaries (Q3 2015 - Q3 2023), and 28 Visual Summaries (Q4 2017 - Q3 2023). The mining of this data and construction of this dataset is no simple task and a useful contribution in and of itself.[29]

We show that, consistent with (i) active BoE efforts to simplify its communications; and (ii) the literature's focus on traditional measures of complexity such as the Flesch-Kincaid score, *semantic* complexity has followed a clear trend-decline since the early 2010s. However, we show that, in contrast, dimensions of *conceptual* complexity, as captured by our novel measures, have *not* followed the same trend. They have evolved with much greater volatility, in particular for the MPR and MPS, and have, if anything, *increased* over this same period.

3.4.1 Semantic complexity

The varying aims and objectives of CB communications over time are well reflected in the semantic complexity measures of BoE publications, particularly the FK score. Since the 1990s, the

[27] Setting $v = 990$ would mean that covering all topics adds 50% to the adjusted jargon count.

[28] See Appendix B.2 for examples of each publication.

[29] Details of the steps taken to do so are provided in Appendix B.2, and we intend to make this text mining and cleaning algorithm, which is replicable for future BoE publications releases, public on our Github project site, for use in future research.

BoE (like many other CBs) has sought to increase its transparency and placed greater weight on communications. This is reflected by the increasing length of MPR publications between 2005 and 2014 (Figure 1bi), rising from 16,350 words in 2006 Q1 to 23,587 words in 2014 Q1. Contemporaneously, Figure 1bii shows that the FK score also increased consistently during this period, estimating that an individual would require, on average, approximately 15 years of schooling to be able to understand the MPRs published in 2013, up from 12 years in 2005. Hence, not only did the length of MPR publications increase, but so did the semantic and structural complexity more generally.[30]

Around the early/mid 2010s, CBs began seeking to reduce the complexity of their communications in order to engage with broader audiences. This was an explicit objective of the BoE following the appointment of Mark Carney as Governor in 2013. It implemented the 'Vision 2020 strategy' which, in part, aimed to increase accessibility of communications, culminating in the introduction of the Visual Summary. Indeed, these efforts are depicted clearly in Figures 1bi and 1bii, with a reversal of the trend increase across both semantic complexity measures. The length of the MPR fell to below 16,000 words in 2019, and, by the FK score, an individual required fewer than 10 years of schooling to understand the 2020 Q1 MPR.

Most recently, following a sequence of significant macroeconomic shocks, we see sharp spikes in the length of the MPR (in 2020 Q2 and Q3, after the onset of Covid-19, and 2021 Q4, after UK inflation had risen above the BoE's 2% target for the first time in nearly a decade), and a higher baseline length of the text, likely reflecting the broader macroeconomic uncertainty. Yet, despite this, the FK score remains relatively stable and low (hovering at around 12 since 2021), reflecting again the BoE's efforts to maintain lower levels of (semantic) complexity.

Finally, on the other two shorter publications, the VS has a significantly lower FK score (mean of 6.30), reflecting its aim of being more accessible, while the MPS is even more complex, by the FK metric, than the MPR (though this too has followed an, albeit more modest, downward trend).[31]

Taken together, focusing only on these measures of complexity, as much of the literature has generally done, one might conclude that efforts by the BoE to simplify its communications have been broadly successful. However, as we show below, the story is somewhat different across other dimensions of complexity.

3.4.2 Conceptual complexity

Moving on from measures of *semantic* complexity, Figures 1di and 1dii show the evolution of *conceptual* complexity, as captured by our two novel measures: the Proportion of Jargon (PoJ) and the McMahon-Naylor (MNCC) index. There are a few things to note.

First, and perhaps most importantly, we do *not* observe the same trend-decline in complexity since the early/mid 2010s across either metric as we saw for the semantic complexity measures. Since 2013, the MPR has maintained its PoJ (i.e. proportion of words that are jargon) consistently around 5%-6%. Its MNCC index (i.e. adjusting the PoJ to account for the breadth of jargon terms used and range of topics discussed), has fluctuated a little over the years, but has not exhibited the same trend-decline observed for the FK score. In contrast, its MNCC index significantly *higher* in recent periods (around 20%) than it had been in 2013 (around 13%). The MPS has also fluctuated significantly across both metrics, with again no clear trend-decline and significantly higher levels of complexity in more recent periods (over 25%) than in 2015 (when it was closer to 20%). Finally, the VS has also fluctuated (increasing between 2017 and 2021, before falling more recently), but again this hasn't followed the trend-decline observed with the FK score, and remains more complex than when it was first introduced.

[30] Summary statistics are presented in Table 4 in Appendix B.2.3.
[31] These observations are consistent with previous analyses (**?**)

Second, we observe the value of exploiting the additional information that the more sophisticated MNCC index is able to make use of, both in comparing (a) the relative levels of the three texts, and (b) their relative volatility. In relation to (a), focusing first on the simple PoJ measure in Figure 1di, we see that, although the MPS typically has a higher score than each of the MPR and the VS, there have been periods in which the VS is similarly as complex as the MPS, and is consistently as, or more, complex than the MPR. Using only this simple metric, one might conclude that this is a rather concerning observation for the VS, which has been introduced specifically with the aim of engaging broader audiences. However, Figure 1dii shows us that, once we adjust the simple PoJ measure to take into account the range of topics that these reports cover, and the breadth of jargon used within these topics, encouragingly, the VS is indeed less complex than each of the other publications. Its MNCC index hovers around the 10% mark, while the adjustment for each of the MPR and MPS reports is much greater. Indeed, the MNCC index for the MPS peaked at 34% in 2018, and more recently lies around 25%, while the MPR lies around 20%.

In relation to (b), we see that while the MPS is highly volatile across both metrics, the MPR is more stable across both, perhaps partly due to its greater length. In contrast to each of these publications, the PoJ and the MNCC index each tell *different* stories for the volatility of the VS. While the PoJ metric depicts a rather volatile picture, the MNCC index presents a more stable one. This tells us that, while the proportion of words that are jargon in the VS fluctuates significantly, the range of jargon terms used and the topics discussed does not.

Figures 2a and Figure 2c depict the observations made above on the relative (a) levels, and (b) volatility of the respective publications. Figure 2a presents the greater breadth of jargon terms and topics discussed in the MPS relative to the VS. We see much a greater range of jargon terms used in the MPS (identified by the coloured words), both within the same topic (e.g. ‘gdp’, ‘growth’, and ‘activity’) as well as across topics (covering monetary policy, inflation, output, financial markets, open economy, labour market, etc.). In contrast, the VS uses a much narrower range of terms and covers far fewer topics; talking mainly about ‘interest rates’ and ‘inflation’ (signified by the size of these terms). Meanwhile, Figure 2c depicts the volatility of MPS reports, captured by the conceptual complexity measures; showing visible and significant differences in the quantity and range of jargon used in different reports (2019 Q4 vs 2023 Q2).[32]

A final point worth noting is that, unlike the simple PoJ metric, the MNCC index seems to capture periods of heightened uncertainty, particularly in the MPR. The index captures visible peaks after the GFC between 2009 and 2010, as well as around the Brexit referendum in 2016 Q3. This potentially reflects the fact that, in seeking to communicate this heightened uncertainty, CBs tend to cover a broader *range* of topics, and use a greater *breadth* of jargon terms; rather than necessarily using *more* jargon.

To summarise, we see that active efforts by the BoE to simplify its communications have been largely successful if we focus (as much empirical work in this field has done) exclusively on *semantic* dimensions of complexity: both Word Count and the Flesch-Kincaid Score have systematically fallen across BoE publications. However, we do *not* see the same trend-declines across measures of *conceptual* complexity. In fact, we see both much greater volatility in these dimensions as well as, in particular for the MPR and MPS, complexity having *increased* across this dimension.

These observations motivate the following question: which dimensions of complexity matter *more*? This is what we seek to answer in the next section.

[32]We also provide the word cloud for the MPR across the entire sample period (2005-2023) in Appendix B.2.4.

Figure 1a: Traditional Semantic Complexity Measures

(i) Word Count (ii) Flesch-Kincaid Score

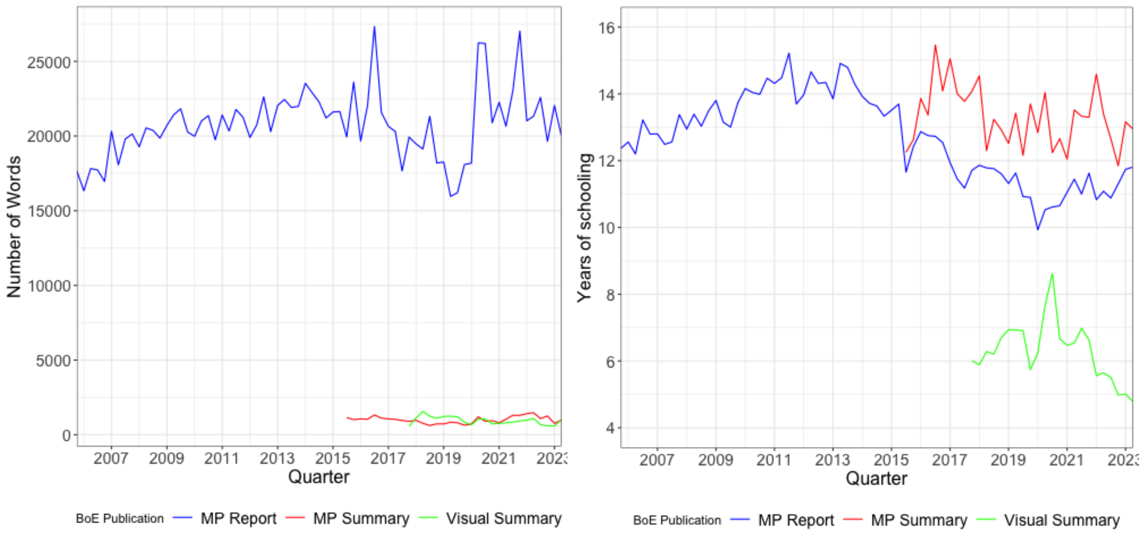


Figure 1c: Novel Conceptual Complexity Measures

(i) Proportion of Jargon (PoJ) (ii) McMahon-Naylor (MNCC) index

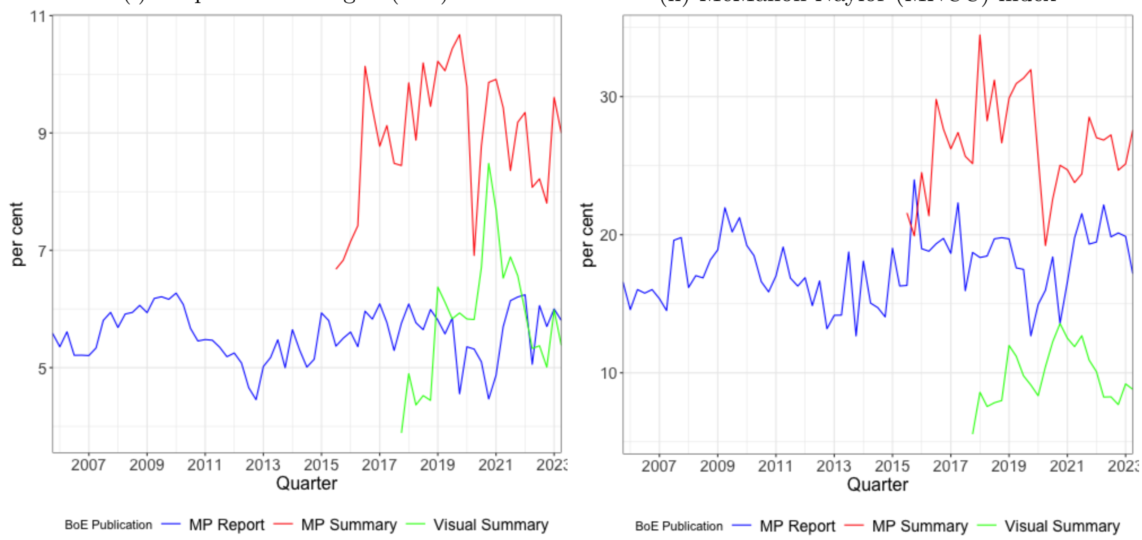
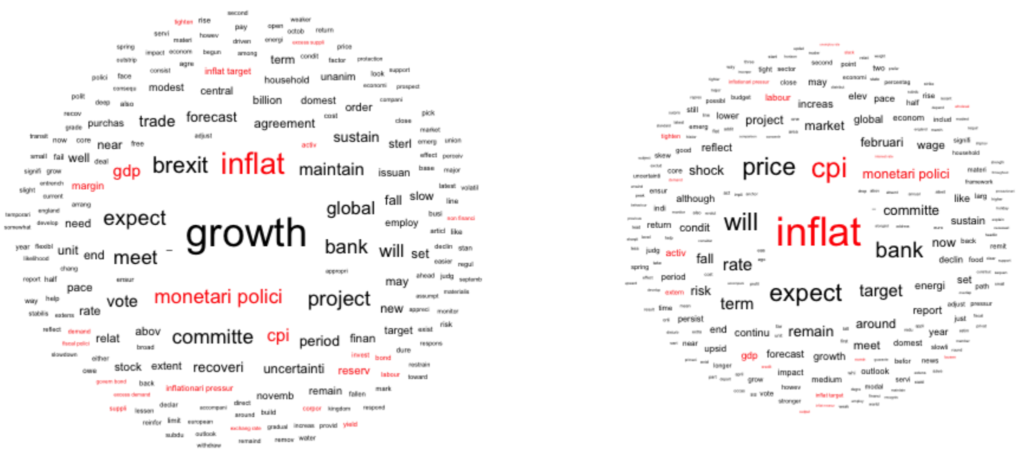


Figure 2a: Wordcloud of jargon terms

(i) MP Summary (2015-2023) (ii) Visual Summary (2017-2023)

Figure 2c: Wordcloud of jargon terms

(i) MP Summary 2019 Q4 (ii) Visual Summary 2023 Q2



4 Experimental Study and Empirical Strategy

4.1 Experimental Study

We run an experimental study to test which dimensions of complexity are most important for effective communication.

Our study consists of 1,800 representative members of the UK public and was conducted in June 2021.[33] We ask participants a set of baseline questions relating to demographics (such as age, region, income, occupation, level of education, and country of birth), level of interest in economic affairs, level of informedness about the state of the economy, and attitudes towards public institutions (including the government, the legal system, and the central bank).

We then randomly assign participants to one of 6 treatments. Each treatment consists of reading a report from a hypothetical central bank. The reports communicate the same underlying information, but vary in their degree of complexity across semantic and conceptual dimensions, with texts categorised as either 'Low', 'Medium', or 'High' complexity across each. On the semantic complexity side, all texts are approximately 1,000 words long, so variation is based only on the FK score. On the conceptual complexity side, variation is based on various combinations of the novel measures we construct in Section 3: the (basic) PoJ and the (more sophisticated) McMahon-Naylor Conceptual Complexity (MNCC) index.[34] The respective FK, PoJ, and MNCC scores associated with each category is specified in Table 1. The texts themselves are pasted in Appendix C.

Table 1: Texts vary across different dimensions of complexity

(a) Variation in complexity

		Semantic		
		Low	Medium	High
Conceptual	Low	Text 1	Text 2	
	Medium	Text 3	Text 4	
	High		Text 5	Text 6

(b) Degree of complexity

Complexity	**Semantic**	**Conceptual**	
	FK	PoJ	MNCC
Low	6	5	10
Medium	10.5	10	15
High	14.5	10	30

The variation in complexity across the texts is derived from actual BoE publications. For instance, text 1, which has 'Low' complexity across both dimensions, reflects the degree of complexity of the 2018 Q1 VS, text 3 ('Low' semantic, 'Medium' conceptual) reflects that of the 2019 Q4 VS, and text 6 ('High' semantic and conceptual) that of the 2018 Q1 MPS.

Having read the report, we then ask participants a set of post-treatment questions. We seek to draw insights on the degree to which these dimensions of complexity impact (i) respondents' *perceived* understanding of the report they read, (ii) their *actual* understanding, and (iii) their sentiments towards the CB.[35]

[33]The study is run through X website, where participants are subscribed members who receive X amount of money per survey they complete.

[34]Note that the variation from 'Medium' to 'High' conceptual complexity comes from an increase in MNCC, while holding PoJ fixed.

[35]We follow the recommendations of Roth et. al., (2020) on information provision experiment design in order

4.1.1 Testing the impact on understanding

Perceived understanding of the report
Before detailing the specific questions we ask, it is worth contextualising what we can and cannot test from a survey experiment such as ours.

In the real world, where attention is a scarce resource, there are a huge number of factors that may affect whether someone decides to pay attention to a message communicated by the central bank (not least, the current level of inflation, and the degree of uncertainty or volatility). For the average individual, there are likely to be much more interesting things to devote attention to. In our study, respondents are paid to complete the survey and read the reports therein. Thus, we cannot directly simulate the conditions in which people decide how and where to optimally allocate attention. However, we *can* glean insights into a number of factors that are likely to *impact* the relative costs and benefits of paying attention to central bank communications, and thus *influence* people's attention allocation decision (on the margin, at least).

One such factor is the degree to which people *believe* they have understood (or would understand) a piece of text. Thus, we first test people's self-reported *perceived* understanding of the central bank report. We ask:

Q30. '*To what extent are you able to understand the content and messages of the material you just read?*'. '*None or nearly none of it*', '*A small amount of it (less than half)*', '*About half of it*', '*A lot of it (more than half)*', '*All or nearly all of it*'.

Respondents' perceived understanding is likely to impact both perceived *costs* of paying attention to central bank reports (trying to understand the information communicated) as well as the perceived *benefits* of doing so (which are small if they don't think they would understand much).

Actual understanding of the report
The survey design we implement allows us to *directly* test the degree to which respondents have accurately processed the information communicated in the report, and how this varies across different dimensions and levels of complexity. We ask respondents various questions relating to descriptions by the hypothetical central bank about the current state of the economy, such as:

Q32. '*What is the current inflation rate in the economy described?*'
And their expectations for how the economy will evolve:

Q32.a. '*What do you think is the probability that the inflation rate in the hypothetical economy over the coming years will be in each of the following intervals? These should sum to 100.*' '*Less than 1%*', '*between 1% and 3%*', '*between 3% and 5%*', '*between 5% and 10%*', '*greater than 10%*'

We also ask respondents questions relating to more tangible day-to-day variables, such as pay:

Q34. '*What do you expect to happen to pay: Rise, fall, or stay the same?*'.

The relative degree to which respondents are able to accurately process this information and accurately update priors across different treatments, has important implications not only for the effective communication of central bank messages but also the broader information processing literature.

4.1.2 Testing the impact on sentiments towards the CB

We can also directly ask respondents questions about their sentiments towards the CB in light of the report they have just read. Specifically, after having asked questions about their understanding,

to reduce measurement error and experimenter demand effects.

we tell respondents that the text they read was based on a report published by the BoE, and ask about their resulting sentiments towards it. With a choice of options between '*Disagree Strongly*', '*Disagree*', '*Neither Agree nor Disagree*', '*Agree*', and '*Agree Strongly*', we ask questions such as:

Q41. *To what extent do you agree with each of the following statements:*

i *Having read the document, I now have a better understanding of the role of the Bank of England.*

ii *Having read the document, I now have more trust in the Bank of England as an institution.*

iii *Having read the document, I am now more likely to pay attention to future documents published by the Bank of England.*

As with respondents' *perceived* understanding of the information they read, their sentiments towards the institution are also likely to impact their desire to pay attention to and engage with CB communications. Indeed, we ask this directly in (iii), and indirectly through (i) and (ii), on the basis that people are unlikely to use up scarce attention on CB messages if they don't know it's function or fundamentally have no trust in it.

4.2 Empirical Strategy

We draw inferences on the effect of increasing complexity across conceptual and semantic dimensions on (i) perceived understanding, (ii) actual understanding, and (iii) sentiments towards the CB, by comparing post-treatment responses across each of these variables of interest, Y_i, across treatments. Our baseline regression specification is given by:

$$\begin{aligned} \mathrm{Y}_i = \beta_0 &+ \beta_1 \text{Conceptual Medium}_i + \beta_2 \text{Conceptual High}_i \\ &+ \gamma_1 \text{Semantic Medium}_i + \gamma_2 \text{Semantic High}_i \\ &+ \delta X_i + \epsilon_i \end{aligned} \tag{10}$$

Conceptual Medium$_i$, Conceptual High$_i$, Semantic Medium$_i$, and Semantic High$_i$ are dummy variables that take the value of 1 if respondent i is treated with a text of that respective level of complexity, and 0 otherwise. X_i represents a set of conditioning demographic factors (such as income, age, education, and country of birth) as well as levels of pre-treatment interest and informedness specific to individual i. ϵ_i is the error term.

Each of the β and γ coefficients can be interpreted as the estimated *casual* marginal effect on Y_i of reading a text of that specific degree of complexity across that specific dimension, rather than the baseline text that is 'low' complexity across both dimensions (text 1), conditioning on the level of complexity across the *other* dimension and on demographic factors and pre-treatment informedness. For example, β_1 is interpreted as the estimated causal marginal effect on Y_i of reading a CB report of 'medium' *conceptual* complexity (i.e. texts 3 or 4) rather than text 1, accounting for the effect of any increase in *semantic* complexity (and demographic factors). That is, we are able to identify the causal marginal effect of increasing complexity across one dimension, disentangling any effect arising from a change in the other dimension.

In theory, by nature of our experiment where participants are *randomly* allocated to a treatment, with a sufficiently large sample size, we would not need to condition on individuals' demographic characteristics in order to identify causal effects. Observations should be identically and independently distributed (IID) and unconditional differences in average responses between texts would be interpreted as the local average treatment (i.e. causal) effect of reading a certain text rather than another. However, with 1,800 participants split across the 6 treatments, there is a question about whether the IID assumption would *necessarily* hold. We reduce our reliance on this assumption by explicitly conditioning on the abovementioned demographic factors, which would contaminate our findings if the demographic constitution of respondents randomly assigned to texts were, by chance, imbalanced. Specifically, we account for those factors that appear to be associated with our dependent variables of interest. We show these relationships in Appendix D. By conditioning on these characteristics, we are confident that we are able to identify causal effects of differences in conceptual and semantic complexity.

5 Results

We derive five main findings. First, complexity, broadly defined, reduces perceived understanding, actual understanding, and sentiments towards the central bank. Second, distinguishing between different dimensions of complexity, we find that it is *conceptual*, not *semantic*, complexity that drives these effects. Third, the impact on understanding and sentiments is explained entirely by the McMahon-Naylor (MNCC) index, and not a simple Proportion of Jargon (PoJ) metric, giving credence to the more sophisticated measure we construct. Fourth, each of these results hold even once we focus on a sub-sample of respondents who studied economics at university, with important implications for communicating effectively not just with the general public, but also other actors in the economy. Fifth, we find some evidence of a potential 'goldilocks' level of conceptual complexity whereby processing certain information becomes more accurate as we increase complexity from 'Low' to 'Medium' but then less accurate again raising this further to 'High'.

5.1 Broadly defined complexity

To begin with, we simply test whether complexity, broadly defined, affects our dependent variables of interest. We find that it does.

We split texts into 'Low', 'Medium', and 'High' complexity across *both* dimensions. The 'Low' category comprises just text 1, which has low semantic and conceptual complexity. The 'Medium' group comprises texts 2, 3, and 4, which each have at least one dimension at medium complexity, but none that are high. The 'High' group comprises texts 5 and 6, where at least one of the dimensions is highly complex.

Figure 3 shows the average, unconditional, degree of *perceived* understanding across respondents exposed to texts of 'Low', 'Medium', and 'High' complexity. We see that perceived understanding falls as complexity increases, and particularly so from 'Medium' to 'High'.

We find that this result holds once we condition on demographic factors in a simplified version of the regression described by equation 10 in Section 4.2, and find similar results across factors capturing *actual* understanding as well as sentiments towards the central bank. We report each of these results in full in Appendix D.[36] Taken together, we draw our first main finding:

Result 1: Complexity, broadly defined, reduces perceived and actual understanding, and sentiments towards the central bank.

5.2 Semantic vs Conceptual Complexity

5.2.1 Conceptual complexity matters more

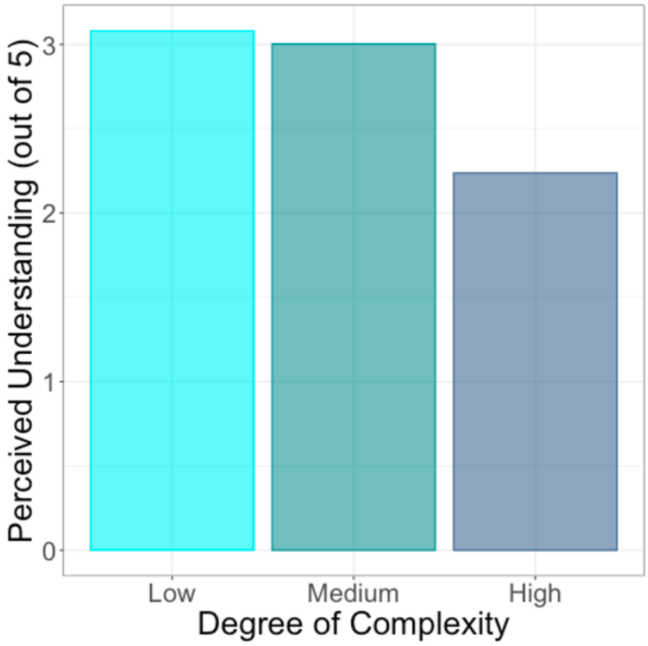
We now go one step further, in seeking to disentangle the effects arising from conceptual and semantic dimensions of complexity, on each of dependent variables of interest.

Perceived understanding of the report
Figure 4 shows the average, unconditional, degree of perceived understanding across all treatments. The colours represent the degree of semantic complexity: green is 'Low', blue is 'Medium', red is 'High'. The patterns represent the degree of conceptual complexity: no pattern is 'Low', striped is 'Medium', and cross-hatch is 'High'.

We see a modest decline in perceived understanding as we increase conceptual complexity from 'Low' to 'Medium', holding semantic complexity fixed at 'Low'. Similarly, there seems to be little

[36] Placeholder: to insert these results into the appendix.

Figure 3: Perceived understanding by degree of complexity



Q30. '*To what extent are you able to understand the content and messages of the material you just read?*'.

'(1) *None or nearly none of it*', '(2) *A small amount of it (less than half)*', '(3) *About half of it*', '(4) *A lot of it (more than half)*', '(5) *All or nearly all of it*'.

change as we increase semantic complexity from 'Low' to 'Medium', holding conceptual complexity fixed at 'Low'. The real action happens at text 5. That is, as conceptual complexity increases from 'Medium' to 'High', holding semantic complexity fixed at 'Medium', we see a *significant* reduction in the degree of perceived understanding. Yet, we don't see a further fall at text 6. That is, once semantic complexity increases from 'Medium' to 'High', while holding conceptual complexity fixed, there is no further material reduction in perceived understanding.

Column (1) of Table 2 confirms that this unconditional observation holds once we condition on demographic characteristics and pre-treatment levels of informedness. We see that the only statistically significant marginal effect on perceived understanding arises from being assigned to a report that is of 'High' conceptual complexity. That is, conditioning on demographic factors and pre-treatment levels of informedness, as well as changes in the degree of semantic complexity, an increase in conceptual complexity from low to high materially reduces the degree of perceived informedness at the 1% level of statistical significance. In contrast, increasing semantic complexity from low to high, while conditioning on the level of conceptual complexity, has no significant effect on perceived understanding.

Furthermore, we see from the magnitude of the coefficient on 'High' conceptual complexity term (-0.802), that the reduction in perceived understanding of the report outweighs the positive effect from each of demographic characteristics we condition on: age (0.004), being born in the UK (0.045), income quintile (0.167), having studied economics at university (0.450) (which we return to below); as well as having had a high pre-treatment level of informedness (0.518).[37]

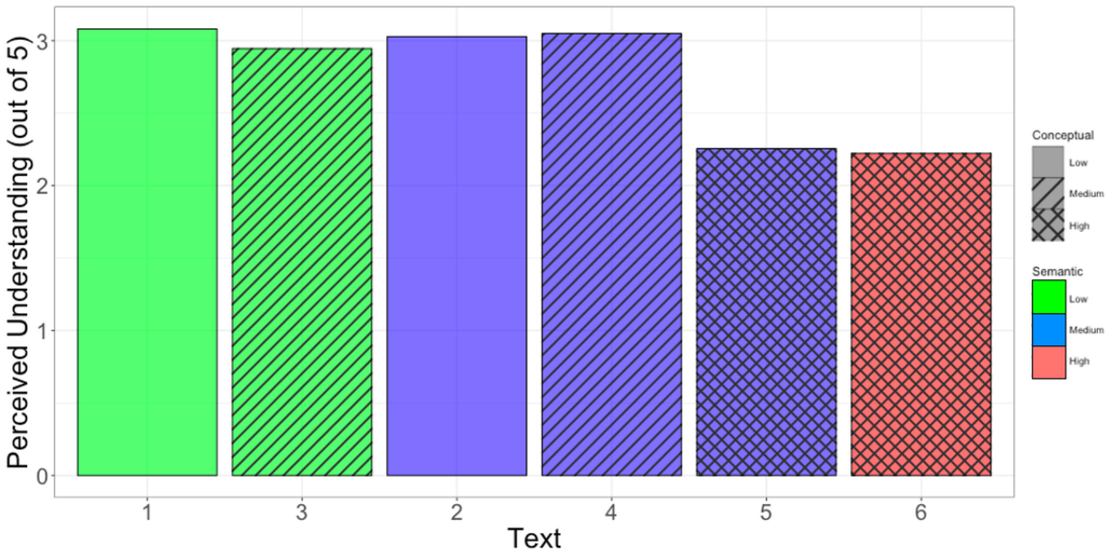
[37] We capture pre-treatment informedness by the accuracy of responses to a question about the BoE's inflation target.

Table 2: Baseline Regression Results

	Perceived	Actual Understanding			Sentiments towards CB		
	Understanding	Inf(t)	i(t)	Exp Pay	Trust	Attention	BoE Role
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Conceptual							
Medium	−0.062	−0.014	0.028	0.006	−0.035	−0.046	−0.090
	(0.054)	(0.028)	(0.028)	(0.027)	(0.052)	(0.063)	(0.060)
High	−0.802***	−0.080*	−0.195***	−0.134***	−0.197**	−0.322***	−0.542***
	(0.083)	(0.043)	(0.043)	(0.042)	(0.080)	(0.097)	(0.092)
Semantic							
Medium	0.007	−0.044	−0.003	−0.049*	0.032	−0.017	0.062
	(0.055)	(0.028)	(0.028)	(0.028)	(0.053)	(0.064)	(0.061)
High	−0.017	−0.004	0.0001	−0.123**	−0.016	−0.135	0.052
	(0.105)	(0.054)	(0.054)	(0.053)	(0.101)	(0.123)	(0.117)
Controls							
Age	0.004*	−0.001	0.003***	−0.001	0.007***	0.003	0.0003
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
UK country of birth	0.045	−0.001	−0.008	−0.013	−0.106*	−0.235***	−0.039
	(0.059)	(0.030)	(0.030)	(0.030)	(0.056)	(0.069)	(0.065)
Income	0.167***	0.012	0.026**	0.017	0.055***	0.032	0.072***
	(0.022)	(0.011)	(0.011)	(0.011)	(0.021)	(0.026)	(0.025)
Econ at Uni	0.448***	−0.033	0.020	−0.048*	0.116**	0.222***	0.253***
	(0.051)	(0.026)	(0.026)	(0.026)	(0.049)	(0.059)	(0.056)
Pre-anchored Exps	0.518***	0.233***	0.174***	0.093***	0.146***	0.122**	0.321***
	(0.047)	(0.024)	(0.024)	(0.024)	(0.045)	(0.055)	(0.052)
Constant	2.173***	0.350***	0.292***	0.676***	1.455***	2.178***	1.737***
	(0.091)	(0.046)	(0.047)	(0.046)	(0.087)	(0.106)	(0.100)
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,745	1,745	1,745	1,745	1,742	1,743	1,745
R^2	0.267	0.063	0.089	0.050	0.046	0.051	0.090

Note: *p<0.1; **p<0.05; ***p<0.01

Figure 4: Perceived understanding by text



Q30. '*To what extent are you able to understand the content and messages of the material you just read?*'.

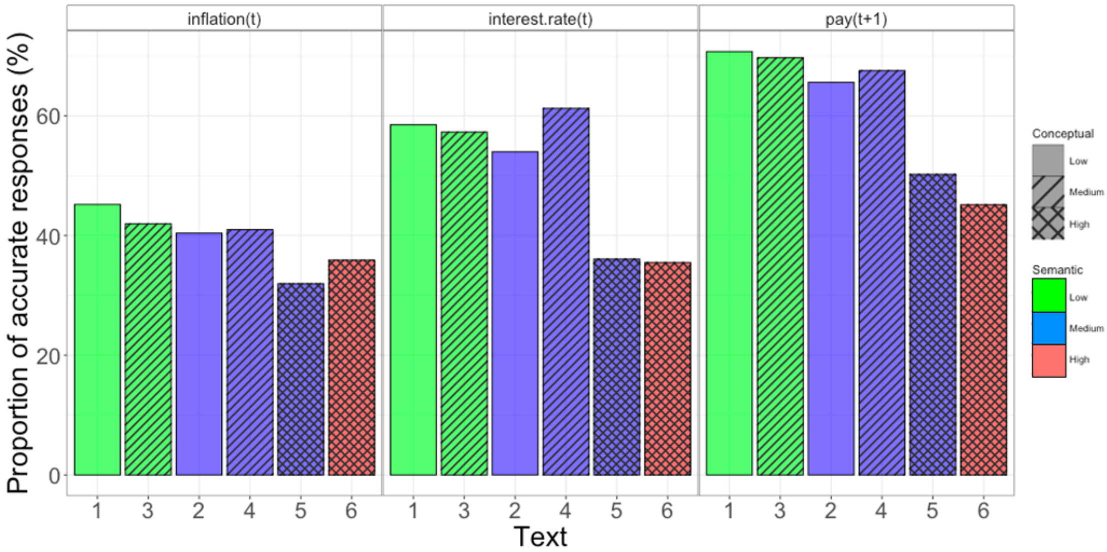
'(1) *None or nearly none of it*', '(2) *A small amount of it (less than half)*', '(3) *About half of it*', '(4) *A lot of it (more than half)*', '(5) *All or nearly all of it*'.

Actual understanding of the report
We see similar results in relation to respondents' *actual* understanding of the information provided in the central bank report. Figure 5 shows the average unconditional proportion of respondents who correctly answered questions related to the current level of inflation, current interest rates, and expected evolution of pay, across the six texts. Again, we observe no significant differences between texts 1-4, and a material reduction from text 4 to 5, as conceptual complexity is increased from 'Medium' to 'High'. Furthermore, we again see little evidence of a reduction in understanding as a result of increasing semantic complexity. Columns 2-4 in Table 2 corroborates these observations, conditioning on the same demographic factors and level of pre-treatment informedness as described above.

Sentiments towards the central bank
Finally, we see the same story also in relation to questions capturing respondents' sentiments towards the central bank. Figure 6 shows the average unconditional responses to questions asking about the degree to which, having read the report, respondents' had a better understanding of the role of BoE, were more likely to pay attention to future documents published by the BoE, and had more trust in the BoE as an institution. Again, across the board, we see the most material fall from text 4 to 5, at 'High' conceptual complexity. Columns 5-7 corroborate these unconditional findings.

Direct feedback on what would make the report simpler
Finally, to corroborate our findings, we directly ask respondents what they think would have made the text they read easier to understand. We ask them to select any of the following: shorter words, shorter sentences, fewer technical words, and fewer technical concepts. If the Flesch-Kincaid score captures the dimensions of complexity (i.e. semantic complexity) that matter most for understanding, then we would expect most respondents to respond with either shorter words or sentences. However, as shown in Figure 7, this is not what we see. We see more than 50% of respondents

Figure 5: Proportion who correctly responded to questions regarding the current state of the hypothetical economy: (i) Inflation(t), (ii) Interest rate(t), and (iii) Expected pay



Q32. *What is the current inflation rate in the hypothetical economy?*

Q35. *What is the current interest rate in the hypothetical economy?*

Q33. *What do you expect to happen to pay? Select one: 'Rise', 'Fall', or 'Stay the Same'*

identified technical concepts and technical words as the greatest barriers to understanding. While 40% of respondents also identified shorter sentences, fewer than 10% pointed to shorter words.

All taken together, we come to our second main finding:

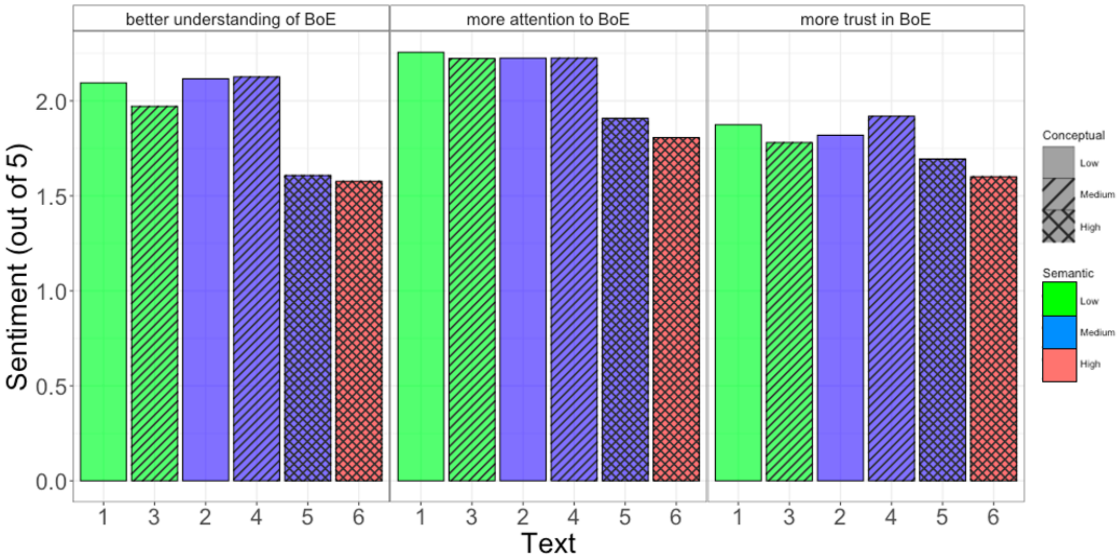
Result 2: Conceptual complexity matters more than semantic complexity in reducing people's perceived and actual understanding of information provided, as well as their sentiments towards the central bank.

5.2.2 The MNCC index exclusively explains the observed effects

The results reported in Section 5.2.1 for perceived understanding, actual understanding, and sentiments show that conceptual complexity matters more than semantic complexity. They also each had one further thing in common: the significant effect arises at 'High' levels of conceptual complexity. The difference between the 'Medium' and 'High' conceptual complexity texts is exclusively captured by the McMahon-Naylor (MNCC) index, and not the simple Proportion of Jargon (PoJ) measure. That is, the sheer quantity of jargon used as a proportion of the total word count in the 'Medium' conceptual complexity texts (3 and 4) and in the 'High' conceptual complexity texts (5 and 6) is the same. The difference between the texts lies in the *range* of topics covered in the respective reports, and the *breadth* of jargon used within these topics.
Thus, the results not only demonstrate the importance of conceptual dimensions of complexity, they also give credence to the sophisticated MNCC index as a way of capturing the key features of conceptual complexity that matter.

Result 3: The impact on understanding and sentiments is entirely explained by the McMahon-Naylor (MNCC) index, and not a simple Proportion of Jargon (PoJ) metric.

Figure 6: Sentiments towards the BoE: (i) understanding of its role, (ii) likelihood of paying attention to future documents it publishes, and (iii) trust in it as an institution



Q41. *To what extent do you agree with each of the following statements:*

i *Having read the document, I now have a better understanding of the role of the Bank of England.*

ii *Having read the document, I am now more likely to pay attention to future documents published by the Bank of England.*

iii *Having read the document, I now have more trust in the Bank of England as an institution.*

'(1) *Disagree Strongly*', '(2) *Disagree*', '(3) *Neither Agree nor Disagree*', '(4) *Agree*', and '(5) *Agree Strongly*'

5.2.3 The results hold for people who studied economics at university

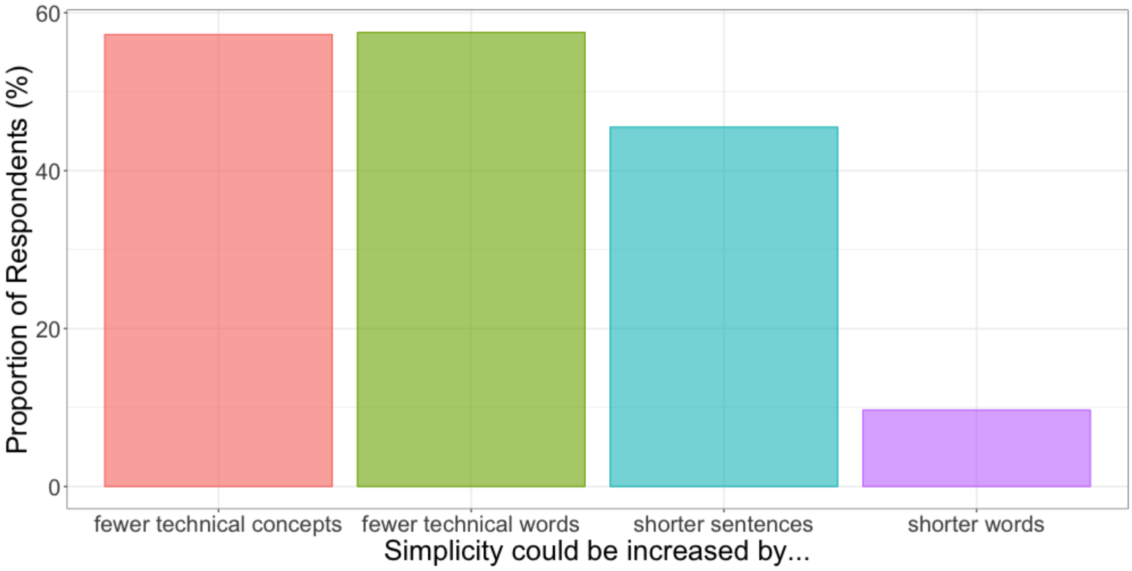
We mentioned in Section 5.2.1 that the magnitude of the negative marginal effect of being assigned to a 'High' conceptual complexity report, reported in Table 2, outweighed the positive effects of various demographic factors we condition on. Most strikingly, perhaps, this was true also for the coefficient capturing whether the respondent had studied economics at university. We dig into this observation by repeating the analysis presented above, but focusing only on the subset of respondents who studied economics at university. Our results are reported in Table 3.

We see that, nearly across the board, the negative marginal effect of having been assigned to a 'High' conceptual complexity report remains significant even across this sub-sample of respondents who are not only highly educated but also specifically trained in economics. Of course, the size of this sample is, naturally, much smaller so these results are potentially less statistically robust than the baseline results reported in 5.2.1. However, the fact that we nevertheless observe statistically significant results is striking.

The implications of these findings are potentially significant. The results suggest that conceptually complex language may not only impact the broad general public, but could also possibly be an important factor when communicating with technically trained audiences, such as journalists and professional forecasters. This is our fourth main finding:

Result 4: High conceptual complexity reduces perceived and actual understanding, as well as sentiments towards the central bank *even* amongst respondents who have

Figure 7: What would make the text simpler to read?



Q42. *Which of the following do you think would have made the text easier to understand? (Please select any that apply)*

Shorter sentences', '*Shorter words* , '*Less reference to technical concepts*', '*Fewer technical words*'

studied economics at university.

Table 3: Sub-Sample: Economics at University

	Perceived	Actual Understanding			Sentiments towards CB		
	Understanding	Inf(t)	i(t)	Exp Pay	Trust	Attention	BoE Role
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
High Conceptual	−0.781***	−0.053	−0.195**	−0.206**	−0.342**	−0.411**	−0.463***
	(0.182)	(0.090)	(0.088)	(0.089)	(0.151)	(0.179)	(0.171)
High Semantic	0.093	−0.041	−0.087	−0.025	0.234	0.030	0.187
	(0.238)	(0.117)	(0.115)	(0.116)	(0.197)	(0.234)	(0.223)
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Econ	Econ	Econ	Econ	Econ	Econ	Econ
Observations	288	288	288	288	288	288	288
R^2	0.201	0.067	0.121	0.068	0.049	0.046	0.042

Note: *p<0.1; **p<0.05; ***p<0.01

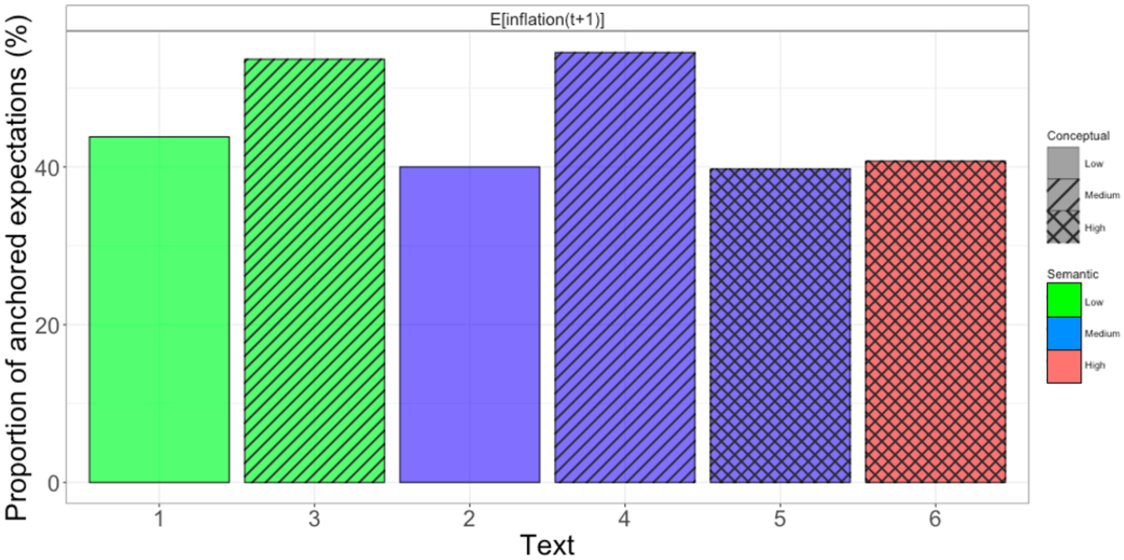
5.2.4 Possible goldilocks zone?

Finally, we report one additional and particularly interesting observation we found in relation to *expectations* for future inflation in the hypothetical economy described in the report. Figure 8 shows the proportion of respondents who developed anchored expectations for inflation in the hypothetical economy (defined as expecting inflation to be between 1% and 3% over the coming years). We see an interesting dynamic: the proportion of respondents who form well-anchored inflation expectations *increases* as conceptual complexity increases from 'Low' to 'Medium'. This then falls again as conceptual complexity increases from 'Medium' to 'High'. This observation points to the possible existence of a 'goldilocks' level of conceptual complexity below which content is oversimplified to such a degree that respondents are unable to link terms to macroeconomic dynamics, and above which the content is too complex to understand. This observation holds when we condition on the usual demographic factors mentioned above, reported in Appendix D.

This brings to our fifth and final main finding:

Result 5: We find evidence of a potential 'goldilocks' levels of conceptual complexity for the formation of accurate expectations.

Figure 8: Proportion who formed anchored expectations about the future state of the hypothetical economy



Q32.a. *What do you think is the probability that the inflation rate in the hypothetical economy over the coming years will be in each of the following intervals? The percentage chance (%) must be a number between 0 and 100 and the sum of your answers must add to 100.*

'Less than 1%', 'Between 1% and 3%', 'Between 3% and 5%', 'Between 5% and 10%', 'Greater than 10%'

6 Conclusions

Central banks and other policy institutions often have to communicate inherently complex messages to a range of different audiences. But *do* they 'get through'? Evidence suggests that, for the majority of the general public, the answer is 'no' (**?**). *Can* they 'get through'? We show in this paper that they can *if* messages are communicated using simpler language. Moreover, we show that this would not only benefit the general public, but also more technically trained audiences, with broader implications for the effectiveness of communications.

We draw these conclusions through three primary contributions. First, we formalise the argument for simplicity in communications by developing a simple rational inattention model, based on **?**, that embeds complexity of communications in an individual's optimal attention allocation decision. We show that linguistic complexity reduces the degree to which people are willing to pay attention to messages communicated by the central bank, and thus, the degree to which they form accurate beliefs about the economy and well-anchored expectations.

Second, we seek to broaden our understanding of what linguistic complexity actually is, by constructing novel quantitative measures of it. To date, both the research literature and policy institutions have focused primarily on rather restrictive measures that capture only narrow 'semantic' dimensions of complexity (e.g. average word and sentence length), such as the Flesch-Kincaid score. These measures provide no insight into how complex the *content* of the text may be. Yet, an extensive theoretical literature on information processing stresses the importance of new information conveyed by *specific* words (**?**). We construct the McMahon-Naylor (MNCC) index that seeks to measure the 'conceptual' complexity of text. Utilising a dictionary of economic and financial jargon terms, the MNCC index captures the quantity and breadth of jargon used in a text, as well as the range of different technical topics covered. Focusing on quarterly Bank of England publications, we show that efforts to simplify language have been successful if one focuses only on *semantic* dimensions of complexity (e.g. the FK score). However, the *conceptual* complexity of its reports has not followed the same trend-decline, instead having *increased* over the same period for certain publications and demonstrated far greater volatility.

Finally, we test the relative importance of these dimensions of complexity in an experimental study with 1,800 representative members of the public. We randomly assign respondents to hypothetical central bank reports that vary in complexity across 'semantic' and 'conceptual' dimensions. We find that '*conceptual*' complexity, captured by the MNCC index, matters more than 'semantic'. It reduces: (i) respondents' *perceived* understanding of the report they read, (ii) their *actual* understanding of the information conveyed, and (iii) their sentiments towards the central bank (such as trust), with some evidence of a potential 'goldilocks' level of complexity. Moreover, each of these results hold focusing on a sub-sample of highly educated respondents who studied economics at university, with potentially important implications for the effectiveness of communications with a *range* of actors in the economy, not just the general public.

Our findings have important and clear policy implications. If central banks and other policy institutions wish to communicate complex messages effectively and 'get through' to their range of audiences, they should pay close consideration to the complexity of the language they use. Specifically, *conceptual* dimensions of complexity are particularly important, as captured by the novel MNCC index we construct, not only for effective communications with the general public but potentially also for *all* economic agents.

Appendix A Theoretical Model: Derivations and Extensions

A.1 Derivations for the Simple Model

A.1.1 Relating λ_h to ξ_h

Information Processed, λ_h
The entropy $H()$ of a normally distributed random variable with variance σ^2 is: $\frac{1}{2}log(2\pi e\sigma^2)$. Thus, the quantity of information chosen by household h can be expressed as:

$$\begin{aligned}\lambda_h &= H(x) - E[H(x|s_h)]\\ &= \frac{1}{2}log(2\pi e\sigma_x^2) - \frac{1}{2}log(2\pi e\sigma_{x|s_h}^2)\\ &= \frac{1}{2}log\left(\frac{\sigma_x^2}{\sigma_{x|s_h}^2}\right)\end{aligned} \tag{11}$$

where $\sigma_{x|s_h}^2$ is the posterior uncertainty that household h has about the true message, x, given the signal, s_h.[38] The reduction in uncertainty about x given the signal, s_h, is characterised by the ratio $\frac{\sigma_x^2}{\sigma_{x|s}^2}$. The smaller is $\sigma_{x|s}^2$ relative to σ_x^2, the greater is the reduction in uncertainty given the acquisition of the signal and, thus, the greater is the quantity of information processed. The entropy function $H()$ can be thought of as describing 'disorder' associated with x. Hence, the more we seek to reduce the uncertainty about x given the signal, s_h, (that is, the smaller is $\sigma_{x|s}^2$), the more we *reduce* the expected 'disorder' around x given s_h (that is, the smaller is $E[H(x|s_h)]$), the greater the quantity of information processed, λ_h, by household h.

Weight attached to the signal received, ξ_h
Just as the quantity of information processed, λ_h, by household h is a characteristic of the degree to which it pays attention to the signal received from the CB, similarly, the weight, ξ_h, that household h attaches to the signal, s_h, is also a feature of the degree to which it pays attention to the signal received from the CB. Hence, household h's optimal choice of attention determines *both* its optimal choice of how much information to process, λ_h, *and* its optimal choice of weight, ξ_h, to attach to the signal, s_h.

Given this, it is convenient to likewise define the weight, ξ_h, in terms of the reduction in uncertainty about the true state of the economy as a result of acquiring the signal, s_h:

$$\xi_h \equiv \left(1 - \frac{\sigma_{x|s}^2}{\sigma_x^2})\right) \tag{12}$$

where $\xi_h \in [0, 1]$. The posterior distribution $x|s$ is derived as follows.

We have a prior distribution: $x \sim \mathcal{N}(0, \sigma_x^2)$. The signal, $s = x + \epsilon$ where $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ contains noisy information about x. Using s, we can form a posterior distribution of $x|s$ given by:

$$x|s \sim \mathcal{N}(E[x|s], \sigma_{x|s}^2)$$

We can derive $E[x|s]$ and $\sigma_{x|s}$ using Bayes' Rule. Firstly, given that each of x and ϵ are normally distributed, we know that a linear addition of the two is also normally distributed such that:

$$s \sim \mathcal{N}(0, \sigma_s^2)$$

In addition, we know that:

$$s|x \sim \mathcal{N}(E[s|x], \sigma_{s|x}^2) \tag{13}$$

[38]Henceforth, $\sigma_{x|s_h}^2 \equiv \sigma_{x|s}^2$ for simplicity of notation.

where

$$\begin{aligned} E[s|x] &= E[x+\epsilon|x] \\ &= x + E[\epsilon|x] \\ &= x + E[\epsilon] \\ &= x \end{aligned}$$

given that $E[\epsilon|x] = E[\epsilon] = 0$ follows from the independence of ϵ from x. The variance is given by:

$$\begin{aligned} \sigma^2_{s|x} &= Var[s|x] \\ &= E[s^2|x] - E[s|x]^2 \\ &= E[(x+\epsilon)^2|x] - E[(x+\epsilon)|x]^2 \\ &= E[x^2 + 2x\epsilon + \epsilon^2|x] - x^2 - 2xE[\epsilon|x] - E[\epsilon|x]^2 \\ &= E[\epsilon^2|x] - E[\epsilon|x]^2 \\ &= Var(\epsilon) \\ &= \sigma^2_\epsilon \end{aligned}$$

again this holds by independence of ϵ from x. The distribution of $s|x$ can be re-written as:

$$s|x \sim \mathcal{N}(x, \sigma^2_\epsilon) \tag{14}$$

With this, we can write the normal distributions of x and $s|x$ as:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left\{-\frac{(x-E[x])^2}{2\sigma_x^2}\right\} = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left\{-\frac{x^2}{2\sigma_x^2}\right\} \tag{15}$$

$$f(s|x) = \frac{1}{\sqrt{2\pi\sigma_{s|x}^2}} \exp\left\{-\frac{(s-x)^2}{2\sigma_{s|x}^2}\right\} = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left\{-\frac{(s-x)^2}{2\sigma_\epsilon^2}\right\} \tag{16}$$

By Bayes' Rule:

$$\begin{aligned} f(x|s) &= \frac{f(x)f(s|x)}{f(s)} \\ &\propto f(x)f(s|x) \\ &= \frac{1}{\sqrt{2\pi\sigma_x^2}} \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left\{-\frac{x^2}{2\sigma_x^2} - \frac{(s-x)^2}{2\sigma_\epsilon^2}\right\} \\ &= constant * \exp\left\{\frac{-x^2\sigma_\epsilon^2 - s^2\sigma_x^2 + 2sx\sigma_x^2 - x^2\sigma_x^2}{2\sigma_x^2\sigma_\epsilon^2}\right\} \\ &\propto \exp\left\{\frac{-x^2(\sigma_x^2+\sigma_\epsilon^2) + 2sx\sigma_x^2 - s^2\sigma_x^2}{2\sigma_x^2\sigma_\epsilon^2}\right\} \\ &= \exp\left\{\frac{-x^2 + 2x\frac{s\sigma_x^2}{\sigma_x^2+\sigma_\epsilon^2} - \frac{s^2\sigma_x^2}{\sigma_x^2+\sigma_\epsilon^2}}{2\frac{\sigma_x^2\sigma_\epsilon^2}{\sigma_x^2+\sigma_\epsilon^2}}\right\} \\ &\propto \exp\left\{\frac{-x^2 + 2x\frac{s\sigma_x^2}{\sigma_x^2+\sigma_\epsilon^2} - \frac{s^2\sigma_x^2}{\sigma_x^2+\sigma_\epsilon^2}}{2\frac{\sigma_x^2\sigma_\epsilon^2}{\sigma_x^2+\sigma_\epsilon^2}}\right\} \exp\left\{\frac{-(\frac{s\sigma_x^2}{\sigma_x^2+\sigma_\epsilon^2})^2}{2\frac{\sigma_x^2\sigma_\epsilon^2}{\sigma_x^2+\sigma_\epsilon^2}}\right\} \\ &= \exp\left\{\frac{-x^2 + 2x\frac{s\sigma_x^2}{\sigma_x^2+\sigma_\epsilon^2} - (\frac{s\sigma_x^2}{\sigma_x^2+\sigma_\epsilon^2})^2}{2\frac{\sigma_x^2\sigma_\epsilon^2}{\sigma_x^2+\sigma_\epsilon^2}}\right\} \exp\left\{\frac{-\frac{s^2\sigma_x^2}{\sigma_x^2+\sigma_\epsilon^2}}{2\frac{\sigma_x^2\sigma_\epsilon^2}{\sigma_x^2+\sigma_\epsilon^2}}\right\} \\ &\propto \exp\left\{\frac{-(x - \frac{s\sigma_x^2}{\sigma_x^2+\sigma_\epsilon^2})^2}{2\frac{\sigma_x^2\sigma_\epsilon^2}{\sigma_x^2+\sigma_\epsilon^2}}\right\} \\ &= \exp\left\{-\frac{(x-E[x|s])^2}{2\sigma_{x|s}^2}\right\} \end{aligned}$$

where

$$E[x|s] = \frac{s\sigma_x^2}{\sigma_x^2 + \sigma_\epsilon^2} \tag{17}$$

$$\sigma_{x|s}^2 = \frac{\sigma_x^2\sigma_\epsilon^2}{\sigma_x^2 + \sigma_\epsilon^2} \tag{18}$$

A density must integrate to unity such that:

$$f[x|s] = \frac{1}{\sqrt{2\pi\sigma_{x|s}^2}} \exp\left\{-\frac{(x - E[x|s])^2}{2\sigma_{x|s}^2}\right\} \tag{19}$$

and the posterior distribution is given by:

$$x|s \sim \mathcal{N}(E[x|s], \sigma_{x|s}^2) \tag{20}$$

where we can sub in from equations (17) and (18) above.

Ultimately, we obtain that: $\sigma_{x|s}^2 = \frac{\sigma_x^2\sigma_\epsilon^2}{\sigma_x^2+\sigma_\epsilon^2}$ and $\xi_h = \frac{\sigma_x^2}{\sigma_x^2+\sigma_\epsilon^2}$. In the case of perfect information where household h faces no constraints on attention, then $\sigma_\epsilon = 0$ and, thus, $\xi_h = 1$.

Combining together

Rearranging and substituting equation (12) into equation (11), we can re-write the quantity of information chosen by household h in terms of the weight, ξ_h, they attach to the signal:

$$\lambda_h = \frac{1}{2} log\left(\frac{1}{1 - \xi_h}\right) \tag{21}$$

A.1.2 Household Maximisation Problem

Households seek to maximise their expected utility subject to their constraint on attention. Their problem is described by:

$$\max \{E[u_h(x, \tilde{x}_h)] - c_h\} \tag{22}$$

Household h's expected utility function is defined as:

$$E[u_h(x, \tilde{x}_h)] = E_x[E_s[u_h(x, \tilde{x}_h)]] \tag{23}$$

$$= E_x[E_s[-b(x - \tilde{x}_h)^2]] \tag{24}$$

$$= E_x[E_s[-b(x - E[x|s_h])^2]] \tag{25}$$

where equations (24) and (25) follow by substitution from (1) and (4). Notice that $E[x-E[x|s_h]]^2 \equiv \sigma_{x|s}^2$ by definition. Hence, we are able to define the expected utility function in terms of the posterior uncertainty about x, given the acquisition of the signal, s_h:

$$E[u_h(x, \tilde{x}_h)] = -b\sigma_{x|s}^2$$

Finally from (12), we can write the expected utility function in terms of the weight, ξ_h, that household h attaches to the signal:

$$E[u_h(x, \tilde{x}_h)] = -b(1 - \xi_h)\sigma_x^2 \tag{26}$$

Now both the expected utility function and the cost function are defined in terms of the weight, ξ_h. Hence, we can specify the household maximisation problem wholly in terms of the exogenous parameters and the choice variable, ξ_h, representing the weight that a representative rationally inattentive household h attaches to the signal received from the CB about the state of the economy. Substituting in from equations (2), (11) and (26), we can rewrite the household's problem described in equation (22) as:

$$\max_{\xi_h \in [0,1]} \left\{-b(1 - \xi_h)\sigma_x^2 - \frac{(1 + \mu)}{2} log(\frac{1}{1 - \xi_h})\right\} \tag{27}$$

The first order condition is:

$$\frac{\partial max}{\partial \xi_h} = b\sigma_x^2 - \frac{(1+\mu)}{2}\frac{1}{1-\xi_h}$$

which yields optimal weight:

$$\xi_h^* = max\left(0, 1 - \frac{(1+\mu)}{2b\sigma_x^2}\right) \tag{28}$$

We can see that the optimal signal weight, reflecting the optimal level of attention, increases with the benefit of paying attention, b, and the degree of uncertainty surrounding the state of the economy, σ_x^2. In contrast, attention decreases with the linguistic complexity of the CB's message, μ.

A.1.3 Bayesian updating of beliefs

Having determined the optimal weight to attach to the signal, we substitute equation (8) into (4) to obtain the corresponding posterior belief is given by:

$$\begin{aligned} \tilde{x}_h &= E[x|s_h] \\ &= \xi_h^*(x + \epsilon_h) \\ &= \left(1 - \frac{(1+\mu)}{2b\sigma_x^2}\right)x + \eta_h \end{aligned}$$

where $\eta_h \equiv \xi_h^* \epsilon_h \sim \mathcal{N}(0, \sigma_\eta^2)$ can be interpreted as resulting noise in actions.[39] Ultimately, the deviation of the posterior belief from the true message communicated by the CB is given by:[40]

$$x - \tilde{x}_h = \frac{(1+\mu)x}{2b\sigma_x^2} - \eta_h \tag{29}$$

A.2 Extension: Role of Journalists

In the real world, most people get their information about the state of the economy via the media, rather than directly from the CB (????). In this section, we draw on the empirical finding presented in Section 5.2.3 (coined 'Result 4') that complexity also impacts highly educated individuals with university degrees in economics, to describe how complexity may still play a pervasive role in reducing the accuracy of expectations formed by economic actors, in a setting in which the media first receives, simplifies, and transmits the CB's message. Specifically, we extend the simple model presented in Section 2 to incorporate a role for the media.

A.2.1 Setup and Assumptions

We assume that journalists, j, are also rationally inattentive and receive the CB signal before transmitting it on to final agents, f. The action choice of the rationally inattentive journalist is assumed to be to pay that level of attention to the CB signal which maximises a constrained expected utility function, where utility depends on the difference between the true state of the economy and its posterior belief. The implicit assumption is that a longer-run objective of maximising public engagement (for example, subscriptions) translates via reputation effects associated with the onward reporting of the CB message with a reasonable degree of accuracy. Another assumption imposed is that the media best achieves this objective of maximising public engagement by minimising the cost that final agents face to paying attention to the message that they transmit. This is modelled as an assumption that the message transmitted to final agents by the media is no longer linguistically complex: $\mu = 0$.

Media journalists receive a *noisy* signal from the CB about the true state of the economy. The media optimally choose how much attention to pay to this signal and form a posterior belief

[39] The variance of the noise in actions, $\sigma_\eta^2 = (\xi_h^*)^2\sigma_\epsilon^2$ will be small as high attentiveness implies relatively high ξ_h^*, but relatively low σ_ϵ^2 and vice versa. At each extreme, $\sigma_\eta^2 = 0$ as $\sigma_\epsilon^2 = 0$ in the full attention case, whilst $\xi_h^* = 0$ in the no attention case.

[40] Note that this deviation is zero in expectation: $E[x - \tilde{x}_h] = E[x] - E[E[x|s_h]] = E[x] - E[x] = 0$ by Law of Iterated Expectations.

about the state of the economy. Before transmitting a signal of their posterior beliefs to final agents, media journalists simplify the language of the original message; so as to achieve their objective of maximising engagement. They then transmit this simplified signal to final agents, who optimally choose how much attention to pay to this. The setup can be summarised by the following propositions.

Proposition 1: Media journalists face exactly the same problem as do final agents in *Scenario 1*. That is, they receive a *noisy* signal about the state of the economy from a message communicated by the CB, and optimally allocate attention to this subject to utility and cost functions, $u_m(x, \tilde{x}_m)$ and c_m, that are analogous to those described by equations (1) and (2) respectively.

Proposition 2: Media journalists transmit a signal of their posterior beliefs about the state of the economy to final agents given by:

$$s_f = \tilde{x}_m + \epsilon_f \tag{30}$$

such that, unlike in *Scenario 1*, final agents no longer receive a signal of the true message, x. Instead, they receive a signal of media journalist m's posterior belief, $\tilde{x}_m$. Given that $\tilde{x}_m$ is itself a function of x which is normally distributed, it also holds that $\tilde{x}_m \sim \mathcal{N}(0, \sigma^2_{\tilde{x}_m})$.

Proposition 3: Final agents' utility is exactly the same as that described in Section 2 for the Direct Signal case:

$$u_f(x, \tilde{x}_f) = -b_f(x - \tilde{x}_f)^2 \tag{31}$$

However, now, final agents do not receive a direct signal of the true state of the economy, x. Instead, the best that final agents can do is to use $\tilde{x}_m$ as a proxy for x. Hence, their utility function can be written as:

$$u_f(\tilde{x}_m, \tilde{x}_f) = -b_f(\tilde{x}_m - \tilde{x}_f)^2 \tag{32}$$

such that a final agent, f, seeks to maximise expected utility by minimising the deviation of its own posterior belief, $\tilde{x}_f$, from that of the media journalist, $\tilde{x}_m$. Note that $u_f(\tilde{x}_m, \tilde{x}_f)$ is a good approximation of $u_f(x, \tilde{x}_f)$ if and only if $x \approx \tilde{x}_m$. Nevertheless, the information that final agents acquire about $\tilde{x}_m$ via the signal, s_f, is the only information that they receive that contains any information about x. Hence, seeking to minimise the distance between their posterior belief, $\tilde{x}_f$, and that of the media journalist, $\tilde{x}_m$, is the best that final agents can do in seeking to minimise the distance from the true state of the economy, x.

Proposition 4: The cost to a final agent, f, of paying attention to the linguistically simplified ($\mu = 0$) signal it receives from the media is given by:

$$c_f = \lambda_f \tag{33}$$

where the cost is no longer a function of final agent f's ability, a_f.

Proposition 5: Final agents optimally allocate attention to s_f to maximise the expectation of utility, described by equation (32) subject to the costs, described by equation (33).

A.2.2 Results

By Proposition 1, the weight attached by the media to the signal it receives from the CB is analogous to that described by equation (8), such that the deviation of their posterior belief from the true state of the economy is:

$$x - \tilde{x}_m = \frac{x}{2b_m\sigma_x^2}[1 + (1 - a_m)\mu\sigma_x^2] - \eta_m \tag{34}$$

Final agents solve their maximisation problem of paying optimal attention to the signal received from the media, yielding a deviation of final agent f's posterior belief from media journalist m's posterior belief given by:

$$\tilde{x}_m - \tilde{x}_f = \frac{\tilde{x}_m}{2b_f\sigma_{\tilde{x}^2_m}} + \eta_m - \eta_f \tag{35}$$

Ultimately, the deviation of final agent f's posterior belief from the true message, x, communicated by the CB is arrived at by summing equations (34) and (35):

$$x - \tilde{x}_f = \frac{x}{2b_m\sigma_x^2}\psi + \frac{x}{2b_f\sigma^2_{\tilde{x}_m}}\left(1 - \frac{}{2b_m\sigma_x^2}\psi\right) - \eta_f \tag{36}$$

where $\psi = [1 + (1 - a_m)\mu\sigma_x^2]$.

A.2.3 Discussion

Each of the implications drawn from the simple model presented in Section 2 hold here. Crucially, the model describes how linguistic complexity continues to play a pervasive role in reducing the degree to which final agents form posterior beliefs that lie close to the true state of the economy as communicated by the CB's message, even in a setting in which the media acts as an intermediary.

For simplicity, we have assumed in this model that there is no *additional* noise generated in the process of simplifying the message received from the CB for purposes of disseminating this to the public. In reality, the more complex the original message communicated, the greater the likelihood of some 'lost in translation' noise arising, and the even more pervasive the role of complexity.

Appendix B Jargon Dictionary and Text Analysis

B.1 Jargon Dictionary

We construct a jargon dictionary based on A-Z lists of economics, business, and financial terms published by the Economist, the Guardian, and Investopedia. Our dictionary contains 350 jargon terms in total. We then manually categorise these into 10 topics. The list of terms in our dictionary, by category, is presented Table B.1.

	Monetary.Policy	Inflation	Output..Production..and.Supply.side	Private.Demand..C.I.	Fiscal.Policy..incl.G.	Open.Economy	Labour	Financial.Markets	Financial.Stability.Macroprudential.Policy	Other
1	Base rate	CPI	Activity	Capital Expenditure	Budget Deficit	Balance of payments	Discouraged workers	30 Year Treasury	Balance Sheet	Adverse Selection
2	Basis point	Demand pull inflation	Aggregate demand	Consumption	Debt Management Office	Capital account	Economically inactive	Bond	Banking System	Animal spirits
3	Discount Rate	Disinflation	Boom	Corporates	DMO	Commodity cycle	Frictional unemployment	Capital markets	Capital Buffers	Annual equivalent rate
4	Effective Lower Bound	Hyperinflation	Business Cycle	Infrastructure	Fiscal drag	Comparative Advantage	Full employment	Carry trade	Credit	Annual Percentage Rate
5	ELB	IMF	Demand	Investment	Fiscal Policy	Currency peg	Furlough	Commercial paper	Credit crunch	APR
6	Interest on reserves	Inflation	Economic Growth	Marginal propensity to consume	Income tax	Devaluation	Gig economy	Credit Default Swap	Disintermediation	Arbitrage
7	Interest Rate	Inflation measures	Excess Demand	Reinvestment	Indirect taxation	Exchange Rate	Job Retention Scheme	Debt Ratio	Financial Stability	Coronavirus
8	Loosen	Inflation targeting	Excess Supply	Retail sales	Laffer curve	Exports	Labour	Debt Service Coverage Ratio	Financial System	Covid
9	Market Implied	Inflationary	GDP	Small and Medium Sized Enterprises	Marginal tax rate	Floating exchange rate	Labour market flexibility	Debt to Equity	Fixed rate mortgage	Covid 19
10	Monetarism	Inflationary pressures	GNI	SME	Nash Equilibrium	Imports	Lump of labour fallacy	Derivatives	Flexible mortgage	Crowding out
11	Monetary Policy	Phillips Curve	GNP	Term Funding Scheme	National debt	Japanese Yen	Nairu	Diversification	Gross Profit	Demand Curve
12	Money supply	Price elasticity	Gross Domestic Product	TFSME	Regressive tax	Tradable	Natural rate of unemployment	Dividend Yield	Gross Profit Margin	Demand Elasticity
13	MPC	Reflation	Gross National Income		Regressive taxes	Trade Deficit	Non Accelerating Inflation Rate of Unemployment	Equity	Interest only mortgage	Duopoly
14	Negative interest rates	RPI	Gross National Product		Seigniorage	Trade Surplus	Reservation wage	Equity derivatives	Investment banks	Econometrics
15	Nominal interest rates	RPIX	Industrial output		Tax		Structural unemployment	Finance	Leverage	Economies of Scale
16	Open Market Operations	Stagflation	Lean manufacturing		Taxation		Unemployment	Financial markets	Leverage Ratio	Elasticity
17	QE	Sticky prices	Macroeconomics				Unemployment Rate	Fixed Income	Liability	Endogenous growth theory
18	QT	Wage price spiral	Macroeconomy				Wage price spiral	Fixed Income Security	Liquid asset	Equilibrium
19	Quantitative Easing		Manufacturing output					Foreign direct investment	Liquidity	Externality
20	Quantitative tightening		Marginal cost					Foreign exchange	Liquidity Coverage Ratio	Free rider
21	R*		National income					Foreign exchange market	Liquidity Ratio	Game Theory
22	Real interest rate		Output					Foreign exchange reserves	Loan to value	Geometric Mean
23	Reserves		Output gap					Forex	Loan To Value Ratio	Gini coefficient
24	Taylor rule		PMI					Forward exchange rate	Macroprudential regulation	Gini Index
25	Tighten		Purchasing Managers' Index					Futures Curves	Margin	Gold standard
26	Zero interest rates		Real Gross Domestic Product					Futures Prices	Net Income	Industrial policy
27	Zero lower bound		Slack					Gilts	Net Operating Income	Inequality
28	ZLB		Stagnation					Government Bond	Net Present Value	Informal economy
29			Supply					Hedge	Net Profit Margin	Invisible Hand
30			Supply shock					Hedging	Net Worth	J Curve
31			Supply side economics					Inverted Yield Curve	Non Financial	Joint Probability
32			Value Added					Investment trust	Non Financial Institution	Keynesian Economics
33								Issuance	NPV	Knightian uncertainty
34								Japanese Government Bond	Regulatory Capital Buffers	Kurtosis
35								Junk Bond	Secondary market	Lagged effect
36								Libor rate	Securities	Lagging indicators
37								Margin Call	Securitisation	Laissez Faire
38								Mark to market	Security	Law of Demand
39								Maturity	Solvency Ratio	Law of Supply
40								Money markets	Sovereign risk	Law of Supply and Demand
41								Over The Counter	Sovereign wealth funds	Leading indicators
42								Over The Counter Market	Sub prime loans	Life cycle hypothesis
43								Present Value	Subprime mortgages	Liquidity trap
44								Price earnings ratio	Systematic risk	Loss aversion
45								Price to Earnings Ratio	Systemic risk	Marginal utility
46								Primary balance	Tangible assets	Market failure
47								Remittances	Value at Risk	Median
48								Repo	VaR	Medium of exchange
49								Repurchase agreement	VAR	Microeconomics
50								Risk adjusted return	Variable rate mortgage	Milton Friedman
51								Risk free return	Volcker Rule	Minsky moment
52								Risk premium	Working Capital	Models
53								S P 500		Money illusion
54								Share index		Monopolistic Competition
55								Share options		Moral hazard
56								Sonia		Multiplier effect
57								Spot market		Nominal
58								Spot price		Normal Distribution
59								Spread		Null Hypothesis
60								Standard Poor's 500		Office for National Statistics
61								Swaps		Office of Fair Trading
62								T Bills		OFT
63								Treasury		ONS
64								Treasury Bills		Opportunity Cost
65								Treasury bonds		Optimal currency area
66								Treasury Inflation Protected Security		Precautionary motive
67								VIX		Public goods
68								VIX CBOE		R Squared
69								Yield		Real terms
70								Yield Basis		Real time indicators
71								Yield Curve		Real values
72								Yield Spread		Regression
73								Yield to Maturity		Ricardian equivalence
74								Zero coupon bond		Risk averse
75								Zero Coupon Inflation Swap		Say's law
76										Seasonal adjustment
77										Standard Deviation
78										Structural adjustment
79										Substitution effect
80										Sunk costs
81										Supply Curve
82										T Test
83										Technical Analysis
84										Volume
85										Wholesale
86										Z Score

B.2 Text Mining

In this section, we explain the specific methods that we used to mine the text from each of the BoE's Monetary Policy Report, Monetary Policy Summary, and Visual Summary publications. The benefit of doing so is that future work might be able to replicate these methods. In addition, this is a fairly arduous task and future research may benefit from a discussion of the methods used to go alongside the code that we will make available.

The Monetary Policy Report (MPR) and Monetary Policy Summaries (MPS) are each found within a single PDF document, whilst the Visual Summary (VS) is available on a BoE web page. Each entail different methods to correctly mine the text. we start with the methods used to mine theMPRand MPS text.

B.2.1 Monetary Policy Report and Monetary Policy Summary

The first step to mining text is to import text in its raw form. The text for the BoE's Inflation Reports and Monetary Policy Summaries are available to download from the quarterly Monetary Policy Report PDF. Raw text imported from a PDF document doesn't distinguish between text that is within graphs, tables, charts and figures from that that actually makes up the main text. Only the latter is desired. This poses a challenge because, up until 2019 Q3, the structure of theMPRis such that each page of the PDF is split into two columns. On the left hand side (LHS) of each page are the graphs, tables, charts and figures, whilst the main text is located on the right hand side (RHS). Unfortunately, importing the raw PDF text combines the graph/chart text on a line on the LHS with the main text on the same line on the RHS. That is, each raw text line/string is made up of the LHS graph text transitioning seamlessly into the RHS main text. Thus, one major challenge of scraping the document for the relevant main text is distinguishing between the LHS and the RHS text.

In addition, the structure of the text within each document is fairly inconsistent. Specifically, a major challenge is to differentiate between different types of pages, each requiring a different scraper to mine the raw text cleanly. Up to 2019 Q3 there are five primary different types of page within each MPR PDF that have to be distinguished between.[41]

1. Title, Contents, Index, and Glossary pages
2. Monetary Policy Summary (MPS) pages
3. MPR 'main' text pages
4. MPR Box text pages
5. MPR 'main' text and Box text combined pages (only present pre-February 2019)[42]

Having downloaded the PDF and imported the raw text into the software used to mine the text (I use have used R), the first task is to split the document text into more manageable chunks, for instance, by page. For the purposes of this analysis, we remove the title, contents, index and glossary pages from each document, focusing on the text for each of the MPS, MPR 'main' text and MPR Box text pages respectively. we now discuss the methods used to mine each of these types of pages. Figures 9, 10, 11, and 12 show examples of each type of page as well as a snapshot of the cleaned text version of that page from the set of text data constructed.

MPS Text Cleaning

We identify a page as being the first page of the MPS if: the second line on the page contains Monetary Policy Summary (lower case pre-2015 Q4); or if the subsequent page is not a Section Title MPR main text page. Unlike MPR main text pages, the MPS text is not split into two columns and contains no graphs, figures, charts or tables and so is relatively straightforward to mine. We omit the title of the MPS from the text string itself across each document, as we also

[41] There has been a structural change in the, renamed, 'Monetary Policy Report' in 2019 Q4 which we discuss later on.

[42] Henceforth, we drop the single inverted comma when referring to 'main' text pages.

do for the MPR text pages and VS documents.

MPR Main Text Page Cleaning
An MPR main text page is identified as any page that contains any one of the following: a section title; an MPR main text specific chart, table or figure; no MPR Box text specific chart, table or figure; or 5 or more strings in a row that begin with more than one space character (indicating a LHS and RHS column split specific to MPR 'main' text pages). Within each MPR main text page there are a number of further distinctions to be made between different types of pages: section title pages, non-section title pages and pages solely containing a full page table or chart. The latter we remove. Section title pages contain the following: the section title in the second line (sometimes extending to the third), a summary paragraph (sometimes in bullet point format) which spans the width of the page, followed by the bulk of the text which is then split into the LHS and RHS columns. we identify the location of the final line of the summary paragraph, splitting the subsequent lines on the page into a LHS and RHS. In order to be able split lines to reflect the break between the LHS and RHS columns one must first identify the length of the LHS string. This is done by identifying the location of the last character preceding a set of at least 15 space characters. This set of at least 15 space characters represents the break between the LHS and RHS column. The length of the LHS column is different on each page so this process must be characterised generally to work for all pages.

I have kept only the narrative text where possible. Thus, we discard each of the following: text on the LHS column of each page (associated with charts, tables, graphs or figures, headers at the top of each page, footnotes at the bottom of each page and sources. We also remove the actual section title and subsection titles from the text string. We do this for consistency with previous sets of such text data (Haldane and McMahon, 2018). In addition, we remove the bullet points from each of the section 'summaries' and add full stops at the end of each to ensure that they are regarded as separate sentences when it comes to applying the measures of linguistic complexity.

MPR Box Text Page Cleaning
Firstly, we identify a Box page as any page that contains: box specific charts, tables, figures, or graphs (these have a slightly different title structure to those in the main text pages); or a lines containing a '?' (subheadings phrased as questions are unique to boxes). Similarly to the main text pages, the Box text pages are split into a LHS and RHS column. we identify the length of the LHS string for each Box page using the same method as above. Unlike main text pages, however, there is no such systematic split of charts and figures on the LHS and text on the RHS. Instead, charts, figures, and text are all intermingled. Hence, for each box page we separate the chart, figure and table lines from the text by identifying and removing lines that contain the following: box specific chart, figure, graph or table title and all of the subsequent 5 lines, begin with more than 2 space characters, begin with 'Source' and all of the subsequent 4 lines, begin with '(X)' where X is a number (reflecting a specific source is being given for that chart, table graph or figure), or begin with 'Key Judgment'. Similarly to the main text pages, we remove the title of the Box, again consistent with previous work by Haldane and McMahon (2018), and also remove any footers at the bottom of the RHS. This leaves just the Box text of interest.

Mixed MPR main and Box Text Page Cleaning
Up to the February 2019 MPR, Boxes are not exclusively confined to their own pages and no longer located exclusively at the end of a section. Instead, they are commonly intermingled with 'main' text appearing in the middle of a section as opposed to the end. We identify a page as being a mixed MPR main and Box text page if it contains any of the following: both a main page specific and a Box page specific chart, table, figure or graph; or any of the other conditions unique to both main and Box text pages respectively.

One then needs to separate the Box text from the main text and and run the appropriate code (described above) on the respective sections of the page. Thankfully, Box text always comes at the top of the page with the 'main' text below it. Thus, one can identify the first line of the main text by identifying the first line on the page to contain either: a 'main' text specific chart, table, graph or figure; or at least 3 consecutive lines of no text on the LHS column. Having identified where the Box ends and the 'main' text begins one can direct the different sections to the relevant parts of the code that scrape that style of page.

Looped script for all documents
Having tailored my code to correctly scrape each page for all the different possible formats of text structure of the MPR publications, we loop this code to run on all pages of each MPR document from August 2019 back to August 2015. The above methods are exhaustive and work for every page in every document during this time period.

There was a systematic change in the text structure of the MPR in November 2019 (or rather, 'Monetary Policy Report' as it is known from this quarter onwards). Specifically, the document no longer splits main and Box text pages into a LHS and RHS column. This makes the document substantially more simple to mine for the relevant text data, but requires a different scraper to do so. Hence, for November 2019 (the last quarter in my sample), we have created a different algorithm to mine the text for. However, we have made this as general as possible and will be replicable for all future Inflation Reports ('Monetary Policy Reports').

All of this text data has been saved in a single data frame, readily available for future analysis.

Novel Data Set
Having mined the relevant text from the MPR publications as described above, we have collated this data into a single data frame. This data frame contains separate columns for each of: MPR main text, MPR Box text, and MPS text. This segmentation allows for analysis to be conducted on each either separately or jointly however is preferred. Specifically, each cell within any one of the columns stated above contains a page of text in a single character string. Columns stating the document, page number, section, subsection, or box to which this text data belongs run alongside the cells containing the text data for chronological purposes.

B.2.2 Visual Summaries

The Visual Summary text is significantly more straight forward to scrape. The Visual Summaries are each available on a URL web page, as opposed to solely in PDF form as is the case for the Inflation Reports. Using the in-built Google Chrome tool 'Selector Gadget' one can select the sections of the web page containing the relevant text to import. Thus, by the time the raw text has been imported into R, it is fairly clean already. The relevant text that we restrict my analysis to is: the interest rate rise, the four key summary points at the top of the page, the subsequent sub-headings and corresponding text. That is, we omit chart text, chart titles, related links, sources and the title (for consistency with the MPS and MPR text data).
The text data for the Visual Summaries lies in a separate data frame, with each cell in the text column containing all of the (cleaned) text from a single VS page in a single character string. Figure 13 shows an example of a VS page and the text mined from it.

Inflation Report August 2019 Monetary Policy Summary i

Monetary Policy Summary

The Bank of England's Monetary Policy Committee (MPC) sets monetary policy to meet the 2% inflation target, and in a way that helps to sustain growth and employment. At its meeting ending on 31 July 2019, the MPC voted unanimously to maintain Bank Rate at 0.75%. The Committee voted unanimously to maintain the stock of sterling non-financial investment-grade corporate bond purchases, financed by the issuance of central bank reserves, at £10 billion. The Committee also voted unanimously to maintain the stock of UK government bond purchases, financed by the issuance of central bank reserves, at £435 billion.

Since May, global trade tensions have intensified and global activity has remained soft. This has led to a substantial decline in advanced economies' forward interest rates and a material loosening in financial conditions, including in the United Kingdom. An increase in the perceived likelihood of a no-deal Brexit has further lowered UK interest rates and led to a marked depreciation of the sterling exchange rate.

Brexit-related developments, such as stockbuilding ahead of previous deadlines, are making UK data volatile. After growing by 0.5% in 2019 Q1, GDP is expected to have been flat in Q2, slightly weaker than anticipated in May. Looking through recent volatility, underlying growth appears to have slowed since 2018 to a rate below potential, reflecting both the impact of intensifying Brexit-related uncertainties on business investment and weaker global growth on net trade. Evidence from companies, up to the middle of July, suggests that uncertainty over the United Kingdom's future trading relationship with the European Union has become more entrenched. The labour market remains tight. Annual pay growth has been relatively strong. Consumer spending has remained resilient. CPI inflation was 2.0% in June and core CPI inflation was 1.8%.

The Committee's updated projections are set out in the accompanying August *Inflation Report*. They continue to assume a smooth adjustment to the average of a range of possible outcomes for the United Kingdom's eventual trading relationship with the European Union. In the central projection, conditioned on prevailing asset prices, underlying output growth is subdued in the near term, reflecting more entrenched Brexit uncertainties. This means that a margin of excess supply persists over the first year of the projection. Thereafter, GDP is projected to accelerate to robust growth rates, reflecting a gradual recovery in global growth and firming UK domestic demand growth, driven in large part by a recovery in investment growth as uncertainties dissipate in line with the Brexit conditioning assumption. The acceleration in GDP results in a significant build-up of excess demand, to around 1¾% of potential GDP by the end of the forecast period. After falling in the near term, CPI inflation is projected to rise above the 2% target, as building excess demand leads to firmer domestic inflationary pressures. Conditioned on prevailing asset prices, CPI inflation reaches 2.4% by the end of the three-year forecast period.

These projections are affected by an inconsistency between the smooth Brexit conditioning assumption underpinning the forecast and the prevailing market asset prices on which the forecasts are also conditioned. These asset prices reflect market participants' perceptions of the likelihood and consequences of a no-deal Brexit. If, as assumed, Brexit proceeds smoothly to some form of deal, market interest rates would likely rise and the sterling exchange rate would likely appreciate. A more consistent forecast would therefore have somewhat lower paths for GDP growth and CPI inflation.

(i) Example: MPS page from 2019 Q3

	Quarter	Page_Number	Monetary_Policy_Summary
1	2019_Q3	2	
2	2019_Q3	3	
4	2019_Q3	4	The Bank of England's Monetary Policy Committee (MPC) sets monetary policy to meet the 2% inflation target, and in a way that helps to sustain growth and employment. At its meeting ending on 31 July 2019, the MPC voted unanimously to maintain Bank Rate at 0.75%. The Committee voted unanimously to maintain the stock of sterling non-financial investment-grade corporate bond purchases, financed by the issuance of central bank reserves, at £10 billion. The Committee also voted unanimously to maintain the stock of UK government bond purchases, financed by the issuance of central bank reserves, at £435 billion. Since May, global trade tensions have intensified and global activity has remained soft. This has led to a substantial decline in advanced economies' forward interest rates and a material loosening in financial conditions, including in the United Kingdom. An increase in the perceived likelihood of a no-deal Brexit has further lowered UK interest rates and led to a marked depreciation of the sterling exchange rate. Brexit-related developments, such as stockbuilding ahead of previous deadlines, are making UK data volatile. After growing by 0.5% in 2019 Q1, GDP is expected to have been flat in Q2, slightly weaker than anticipated in May. Looking through recent volatility, underlying growth appears to have slowed since 2018 to a rate below potential, reflecting both the impact of intensifying Brexit-related uncertainties on business investment and weaker global growth on net trade. Evidence from companies, up to the middle of July, suggests that uncertainty over the United Kingdom's future trading relationship with the European Union has become more entrenched. The labour market remains tight. Annual pay growth has been relatively strong. Consumer spending has remained resilient. CPI inflation was 2.0% in June and core CPI inflation was 1.8%. The Committee's updated projections are set out in the accompanying August Inflation Report. They continue to assume a smooth adjustment to the average of a range of possible outcomes for the United Kingdom's eventual trading relationship with the European Union. In the central projection, conditioned on prevailing asset prices, underlying output growth is subdued in the near term, reflecting more entrenched Brexit uncertainties. This means that a margin of excess supply persists over the first year of the projection. Thereafter, GDP is projected to accelerate to robust growth rates, reflecting a gradual recovery in global growth and firming UK domestic demand growth, driven in large part by a recovery in investment growth as uncertainties dissipate in line with the Brexit conditioning assumption. The acceleration in GDP results in a significant build-up of excess demand, to around 1¾% of potential GDP by the end of the forecast period. After falling in the near term, CPI inflation is projected to rise above the 2% target, as building excess demand leads to firmer domestic inflationary pressures. Conditioned on prevailing asset prices, CPI inflation reaches 2.4% by the end of the three-year forecast period. These projections are affected by an inconsistency between the smooth Brexit conditioning assumption underpinning the forecast and the prevailing market asset prices on which the forecasts are also conditioned. These asset prices reflect market participants' perceptions of the likelihood and consequences of a no-deal Brexit. If, as assumed, Brexit proceeds smoothly to some form of deal, market interest rates would likely rise and the sterling exchange rate would likely appreciate. A more consistent forecast would therefore have somewhat lower paths for GDP growth and CPI inflation. Increased uncertainty about the nature of EU withdrawal means that the economy could follow a wide range of paths over coming years. The appropriate path of monetary policy will depend on the balance of the effects of Brexit on

(ii) Snapshot of Cleaned MPS Text

Figure 9: Monetary Policy Summary example page and cleaned text data

1 Global developments and domestic financial conditions

- The outlook for global growth has deteriorated a little, in part reflecting escalating trade tensions.
- The market path for interest rates has fallen further in the UK since May, as in other advanced economies.
- The probability market participants attach to a no-deal Brexit has increased. This has contributed to the lower path for UK interest rates and the 4% depreciation of sterling.

Table 1.A Global GDP growth appears to have slowed slightly in 2019 Q2

GDP in selected countries and regions[(a)]

Percentage changes on a quarter earlier

	Quarterly averages		2018				2019	
	1998–2007	2010–17	Q1	Q2	Q3	Q4	Q1	Q2
United Kingdom	0.7	0.5	0.1	0.4	0.7	0.2	0.5	n.a.
Euro area (39%)	0.6	0.3	0.4	0.4	0.2	0.2	0.4	0.2
United States (18%)	0.7	0.6	0.6	0.9	0.7	0.3	0.8	0.5
China (4%)[(b)]	2.5	1.9	1.5	1.7	1.6	1.5	1.4	1.6
Japan (2%)	0.3	0.3	-0.1	0.6	-0.6	0.5	0.6	n.a.
India (1%)	1.8	1.8	2.0	1.3	1.4	1.7	1.2	n.a.
Russia (1%)[(c)]	1.9	0.4	1.7	0.7	0.3	0.1	-0.4	n.a.
Brazil (1%)	0.8	0.2	0.5	0.0	0.5	0.1	-0.2	n.a.
UK-weighted world GDP[(d)]	0.7	0.6	0.6	0.6	0.4	0.4	0.6	0.4

Sources: Eikon from Refinitiv, IMF *World Economic Outlook* (WEO), National Bureau of Statistics of China, OECD, ONS and Bank calculations.

(a) Real GDP measures. Figures in parentheses are shares in UK exports in 2017.
(b) Estimates from 2010 Q4 onwards are from the National Bureau of Statistics of China. Earlier estimates are based on OECD data.
(c) The earliest observation for Russia is 2003 Q2.
(d) Constructed using data for real GDP growth rates for 180 countries weighted according to their shares in UK exports. Figure for 2019 Q2 is a Bank staff projection. The latest US and euro-area GDP data for 2019 Q2 have not been incorporated into this projection.

Chart 1.1 Survey indicators of global output growth have fallen, particularly in the manufacturing sector

Global purchasing managers' indices[(a)]

Sources: Eikon from Refinitiv, IHS Markit and JPMorgan.

(a) Measures of current monthly composite (services and manufacturing) output, manufacturing output, and manufacturing export orders growth based on the results of surveys in 44 countries. Together these countries account for an estimated 89% of global GDP.

1.1 Global economic developments

Since May, the outlook for global growth has deteriorated a little. In 2019 Q2, UK-weighted world GDP growth appears to have slowed slightly to 0.4% (**Table 1.A**), slightly lower than expected in May. US and euro-area GDP growth both slowed, following surprising strength in Q1, to 0.5% and 0.2% respectively. Growth in emerging markets has been weaker than projected in the May *Report*, having slowed in the past year reflecting a previous tightening in financial conditions.

Higher-frequency indicators further suggest that global output growth may have weakened in recent months. Global PMIs have continued to fall since May, particularly in the manufacturing sector, where the output index has dipped below 50 (**Chart 1.1**). Forward-looking surveys suggest that growth is likely to stabilise in the near term. For example, the manufacturing export orders index has remained at a similar level over the past three months (**Chart 1.1**).

Softer global growth — particularly in the manufacturing sector — is likely at least in part to reflect the impact of trade tensions, which have increased over the past year and intensified further since May. The US and China both implemented higher tariffs over 2018, with the US applying tariffs to US$250 billion of imports from China, and China reciprocating with tariffs on US$110 billion of imports from the US. Tariffs were due to increase in 2019, but at the time of the May *Report*, those were not assumed to be implemented, given that trade talks between the two countries appeared to be progressing positively. Trade talks subsequently broke down, however, and tariffs were increased.

As well as the tariffs implemented so far, other developments have added to concerns about trade protectionism. For example, the US announced plans to impose further tariffs on all remaining imports from China, although in June both parties agreed to continue talks. The US administration is also

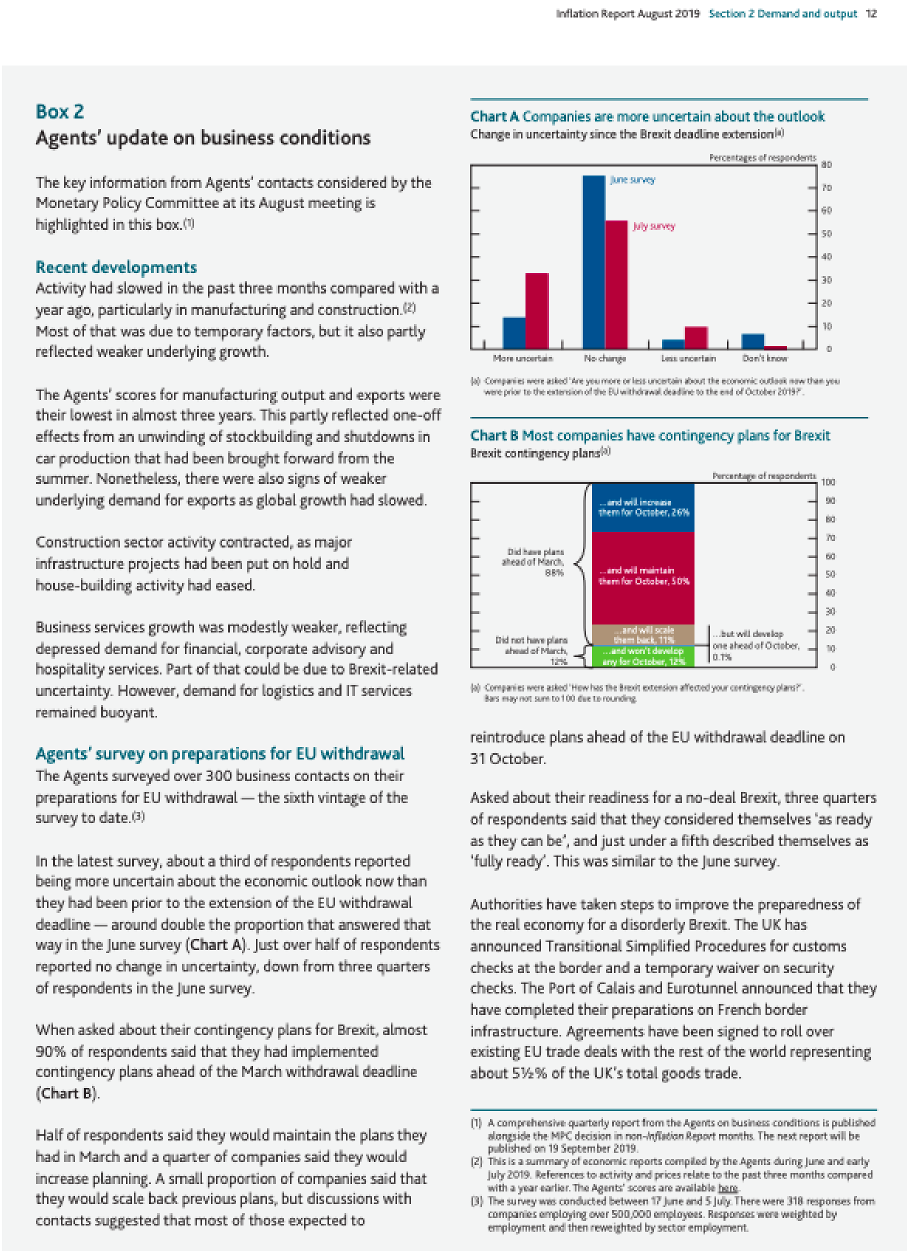
(i) Example: An MPR main text page from 2019 Q3

Quarter	Page_Number	IR_Section	IR_Sub_Section	Inflation_Report
2019_Q3	6	1 Global developments and domestic financial conditions.	1.1 Global economic developments.	The outlook for global growth has deteriorated a little, in part reflecting escalating trade tensions. The market path for interest rates has fallen further in the UK since May, as in other advanced economies. The probability market participants attach to a no-deal Brexit has increased. This has contributed to the lower path for UK interest rates and the 4% depreciation of sterling.Since May, the outlook for global growth has deteriorated a little. In 2019 Q2, UK-weighted world GDP growth appears to have slowed slightly to 0.4% (Table 1.A), slightly lower than expected in May. US and euro-area GDP growth both slowed, following surprising strength in Q1, to 0.5% and 0.2% respectively. Growth in emerging markets has been weaker than projected in the May Report, having slowed in the past year reflecting a previous tightening in financial conditions. Higher-frequency indicators further suggest that global output growth may have weakened in recent months. Global PMIs have continued to fall since May, particularly in the manufacturing sector, where the output index has dipped below 50 (Chart 1.1). Forward-looking surveys suggest that growth is likely to stabilise in the near term. For example, the manufacturing export orders index has remained at a similar level over the past three months (Chart 1.1). Softer global growth — particularly in the manufacturing sector — is likely at least in part to reflect the impact of trade tensions, which have increased over the past year and intensified further since May. The US and China both implemented higher tariffs over 2018, with the US applying tariffs to US$250 billion of imports from China, and China reciprocating with tariffs on US$110 billion of imports from the US. Tariffs were due to increase in 2019, but at the time of the May Report, those were not assumed to be implemented, given that trade talks between the two countries appeared to be progressing positively. Trade talks subsequently broke down, however, and tariffs were increased. As well as the tariffs implemented so far, other developments have added to concerns about trade protectionism. For example, the US announced plans to impose further tariffs on all remaining imports from China, although in June both parties agreed to continue talks. The US administration is also

(ii) Snapshot of cleaned MPR text

Figure 10: Monetary Policy Report main text example page and cleaned text data

B.2.3 Summary Statistics for BoE Publications



Inflation Report August 2019 Section 2 Demand and output 12

Box 2
Agents' update on business conditions

The key information from Agents' contacts considered by the Monetary Policy Committee at its August meeting is highlighted in this box.(1)

Recent developments
Activity had slowed in the past three months compared with a year ago, particularly in manufacturing and construction.(2) Most of that was due to temporary factors, but it also partly reflected weaker underlying growth.

The Agents' scores for manufacturing output and exports were their lowest in almost three years. This partly reflected one-off effects from an unwinding of stockbuilding and shutdowns in car production that had been brought forward from the summer. Nonetheless, there were also signs of weaker underlying demand for exports as global growth had slowed.

Construction sector activity contracted, as major infrastructure projects had been put on hold and house-building activity had eased.

Business services growth was modestly weaker, reflecting depressed demand for financial, corporate advisory and hospitality services. Part of that could be due to Brexit-related uncertainty. However, demand for logistics and IT services remained buoyant.

Agents' survey on preparations for EU withdrawal
The Agents surveyed over 300 business contacts on their preparations for EU withdrawal — the sixth vintage of the survey to date.(3)

In the latest survey, about a third of respondents reported being more uncertain about the economic outlook now than they had been prior to the extension of the EU withdrawal deadline — around double the proportion that answered that way in the June survey (**Chart A**). Just over half of respondents reported no change in uncertainty, down from three quarters of respondents in the June survey.

When asked about their contingency plans for Brexit, almost 90% of respondents said that they had implemented contingency plans ahead of the March withdrawal deadline (**Chart B**).

Half of respondents said they would maintain the plans they had in March and a quarter of companies said they would increase planning. A small proportion of companies said that they would scale back previous plans, but discussions with contacts suggested that most of those expected to

Chart A Companies are more uncertain about the outlook
Change in uncertainty since the Brexit deadline extension(a)

(a) Companies were asked 'Are you more or less uncertain about the economic outlook now than you were prior to the extension of the EU withdrawal deadline to the end of October 2019?'.

Chart B Most companies have contingency plans for Brexit
Brexit contingency plans(a)

(a) Companies were asked 'How has the Brexit extension affected your contingency plans?'. Bars may not sum to 100 due to rounding.

reintroduce plans ahead of the EU withdrawal deadline on 31 October.

Asked about their readiness for a no-deal Brexit, three quarters of respondents said that they considered themselves 'as ready as they can be', and just under a fifth described themselves as 'fully ready'. This was similar to the June survey.

Authorities have taken steps to improve the preparedness of the real economy for a disorderly Brexit. The UK has announced Transitional Simplified Procedures for customs checks at the border and a temporary waiver on security checks. The Port of Calais and Eurotunnel announced that they have completed their preparations on French border infrastructure. Agreements have been signed to roll over existing EU trade deals with the rest of the world representing about 5½% of the UK's total goods trade.

(1) A comprehensive quarterly report from the Agents on business conditions is published alongside the MPC decision in non-*Inflation Report* months. The next report will be published on 19 September 2019.
(2) This is a summary of economic reports compiled by the Agents during June and early July 2019. References to activity and prices relate to the past three months compared with a year earlier. The Agents' scores are available here.
(3) The survey was conducted between 17 June and 5 July. There were 318 responses from companies employing over 500,000 employees. Responses were weighted by employment and then reweighted by sector employment.

(i) Example: Box 2 page from the 2019 Q3 MPR

Quarter	IR_Box_Number	IR_Boxes
2019_Q3	Box 2.	Agents' update on business conditions. The key information from Agents' contacts considered by the Monetary Policy Committee at its August meeting is Recent developments Activity had slowed in the past three months compared with a Most of that was due to temporary factors, but it also partly reflected weaker underlying growth. The Agents' scores for manufacturing output and exports were their lowest in almost three years. This partly reflected one-off effects from an unwinding of stockbuilding and shutdowns in car production that had been brought forward from the summer. Nonetheless, there were also signs of weaker underlying demand for exports as global growth had slowed. Construction sector activity contracted, as major infrastructure projects had been put on hold and house-building activity had eased. Business services growth was modestly weaker, reflecting depressed demand for financial, corporate advisory and hospitality services. Part of that could be due to Brexit-related uncertainty. However, demand for logistics and IT services remained buoyant. Agents' survey on preparations for EU withdrawal The Agents surveyed over 300 business contacts on their preparations for EU withdrawal — the sixth vintage of the survey to date. In the latest survey, about a third of respondents reported being more uncertain about the economic outlook now than they had been prior to the extension of the EU withdrawal deadline — around double the proportion that answered that way in the June survey (Chart A). Just over half of respondents reported no change in uncertainty, down from three quarters of respondents in the June survey. When asked about their contingency plans for Brexit, almost 90% of respondents said that they had implemented contingency plans ahead of the March withdrawal deadline (Chart B). Half of respondents said they would maintain the plans they had in March and a quarter of companies said they would increase planning. A small proportion of companies said that they would scale back previous plans, but discussions with contacts suggested that most of those expected to reintroduce plans ahead of the EU withdrawal deadline on 31 October. Asked about their readiness for a no-deal Brexit, three quarters of respondents said that they considered themselves 'as ready as they can be', and just under a fifth described themselves as 'fully ready'. This was similar to the June survey. Authorities have taken steps to improve the preparedness of the real economy for a disorderly Brexit. The UK has announced Transitional Simplified Procedures for customs checks at the border and a temporary waiver on security checks. The Port of Calais and Eurotunnel announced that they have completed their preparations on French border infrastructure. Agreements have been signed to roll over existing EU trade deals with the rest of the world representing about 5½% of the UK's total goods trade.

(ii) Snapshot of cleaned Box text

Figure 11: Example of Monetary Policy Report Box text page and cleaned text data

Box 1

Monetary policy since the February *Report*

In the MPC's central projection in the February *Report*, GDP was expected to grow by around 1¾% per year on average over the forecast period. While modest by historical standards, that growth rate was expected to exceed the diminished rate of supply growth of the economy, which was projected to be around 1½% per year. As a result, a small margin of excess demand was projected to emerge by early 2020 and build thereafter. That would support domestic cost growth, although CPI inflation was projected to fall back gradually as the effects of sterling's past depreciation faded. Conditional on the path for Bank Rate implied by market interest rates prevailing at the time, inflation remained above the 2% target in the second and third years of the MPC's central projection.

At its meeting ending on 21 March 2018, the MPC noted that recent data releases had been broadly consistent with the view of the medium-term outlook set out in the February *Report*. The prospects for global GDP growth remained strong, and financial conditions continued to be accommodative, with little persistent effect from the recent financial market volatility.

CPI inflation fell from 3.0% in January to 2.7% in February. Inflation was expected to ease further, although to remain above the 2% target in the short term. A range of measures of domestically generated inflation had picked up, although they generally remained below target-consistent levels. Wage growth had also increased steadily, as expected. The unemployment rate had remained low and other indicators suggested that the margin of spare capacity within the labour market was limited. Taken together with the increase in job-to-job flow rates, which could strengthen workers' bargaining power, this provided increasing confidence that growth in wages and unit labour costs would pick up to rates consistent with the inflation target.

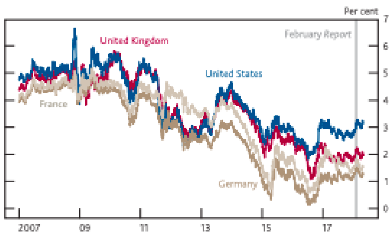
As in February, the best collective judgement of the MPC remained that, given the prospect of excess demand, an ongoing tightening of monetary policy over the forecast period would be appropriate to return inflation sustainably to its target at a more conventional horizon. All members agreed that any future increases in Bank Rate were likely to be at a gradual pace and to a limited extent.

For seven members, however, that did not require an increase in Bank Rate at this meeting. There had been few surprises in recent economic data and the February *Inflation Report* projections, conditioned on a gently rising path of Bank Rate, had appeared broadly on track. The May forecast round would enable the Committee to undertake a fuller assessment of the underlying momentum in the economy, the degree of slack remaining and the extent of domestic inflationary pressures.

Two members favoured a rise in Bank Rate by 25 basis points. They noted the widespread evidence that slack was largely used up and that pay growth was picking up, presenting upside risks to inflation in the medium term. A modest tightening of monetary policy at this meeting could mitigate the risks from a more sustained period of above-target inflation that might ultimately necessitate a more abrupt change in policy.

Chart 1.10 Longer-term interest rates have risen slightly in the US since the run-up to the February *Report*

Five-year, five-year forward nominal interest rates[a]



Sources: Bloomberg Finance L.P. and Bank calculations.

between lenders in the face of subdued demand for credit in the UK.

Borrowing costs for households and companies are projected to rise gradually further in coming years. The market-implied path for Bank Rate remains upward sloping and bank funding costs, over and above those market rates, have risen in recent months, unwinding some of the compression since 2016. This will feed through gradually to higher retail interest rates. Nevertheless, credit conditions are likely to remain accommodative, and this will support the outlook for consumption and business investment (Section 5).

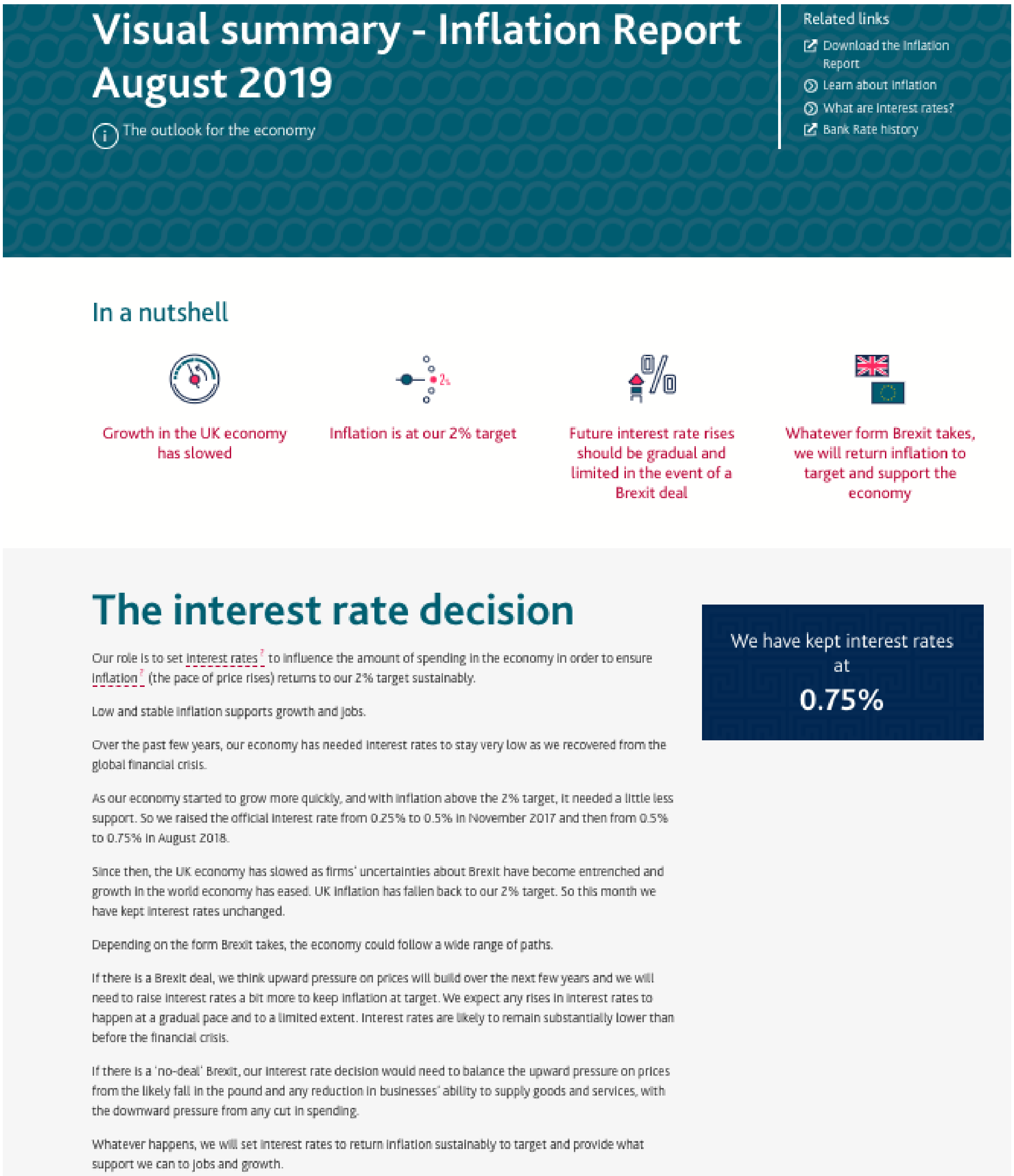
Bank funding costs and retail deposit rates

(i) Example: Page with Box 1 and main text from 2018 Q2 MPR

Quarter	Inflation_Report	IR_Box_Number	IR_Boxes
2018_Q2	between lenders in the face of subdued demand for credit in the UK. Borrowing costs for households and companies are projected to rise gradually further in coming years. The market-implied path for Bank Rate remains upward sloping and bank funding costs, over and above those market rates, have risen in recent months, unwinding some of the compression since 2016. This will feed through gradually to higher retail interest rates. Nevertheless, credit conditions are likely to remain accommodative, and this will support the outlook for consumption and business investment (Section 5). Bank funding costs and retail deposit rates As explained in Box 2, an important component of the interest rates facing households and companies is the cost of bank funding. During the financial crisis, the spread between reference rates and bank funding costs rose sharply as investors demanded higher compensation for the risks associated with providing funding to banks (Chart 1.11). These	Box 1.	Monetary policy since the February Report. In the MPC's central projection in the February Report, GDP was expected to grow by around 1¾% per year on average over the forecast period. While modest by historical standards, that growth rate was expected to exceed the diminished rate of supply growth of the economy, which was projected to be around 1½% per year. As a result, a small margin of excess demand was projected to emerge by early 2020 and build thereafter. That would support domestic cost growth, although CPI inflation was projected to fall back gradually as the effects of sterling's past depreciation faded. Conditional on the path for Bank Rate implied by market interest rates prevailing at the time, inflation remained above the 2% target in the second and third years of the MPC's central projection. At its meeting ending on 21 March 2018, the MPC noted that recent data releases had been broadly consistent with the view of the medium-term outlook set out in the February Report. The prospects for global GDP growth remained strong, and financial conditions continued to be accommodative, with little persistent effect from the recent financial market volatility. CPI inflation fell from 3.0% in January to 2.7% in February. Inflation was expected to ease further, although to remain above the 2% target in the short term. A range of measures of domestically generated inflation had picked up, although they generally remained below target-consistent levels. Wage growth had also increased steadily, as expected. The unemployment rate had remained low and other indicators suggested that the margin of spare capacity within the labour market was limited. Taken together with the increase in job-to-job flow rates, which could strengthen workers' bargaining power, this provided increasing confidence that growth in wages and unit labour costs would pick up to rates consistent with the inflation target. As in February, the best collective judgement of the MPC remained that, given the prospect of excess demand, an ongoing tightening of monetary policy over the forecast period would be appropriate to return inflation sustainably to its target at a more conventional horizon. All members agreed that any future increases in Bank Rate were likely to be at a gradual pace and to a limited extent. For seven members, however, that did not require an increase in Bank Rate at this meeting. There had been few surprises in recent economic data and the February Inflation Report projections, conditioned on a gently rising path of Bank Rate, had appeared broadly on track. The May forecast round would enable the Committee to undertake a fuller assessment of the underlying momentum in the economy, the degree of slack remaining and the extent of domestic inflationary pressures. Two members favoured a rise in Bank Rate by 25 basis points. They noted the widespread evidence that slack was largely used up and that pay growth was picking up, presenting upside risks to inflation in the medium term. A modest tightening of monetary policy at this meeting could mitigate the risks from a more sustained period of above-target inflation that might ultimately necessitate a more abrupt change in policy.

(ii) Snapshot of cleaned MPR main and Box text

Figure 12: Example of Monetary Policy Report combined 'main' and Box text page and cleaned text data



Visual summary - Inflation Report August 2019

The outlook for the economy

Related links

- Download the Inflation Report
- Learn about inflation
- What are interest rates?
- Bank Rate history

In a nutshell

Growth in the UK economy has slowed

Inflation is at our 2% target

Future interest rate rises should be gradual and limited in the event of a Brexit deal

Whatever form Brexit takes, we will return inflation to target and support the economy

The interest rate decision

We have kept interest rates at 0.75%

Our role is to set interest rates to influence the amount of spending in the economy in order to ensure inflation (the pace of price rises) returns to our 2% target sustainably.

Low and stable inflation supports growth and jobs.

Over the past few years, our economy has needed interest rates to stay very low as we recovered from the global financial crisis.

As our economy started to grow more quickly, and with inflation above the 2% target, it needed a little less support. So we raised the official interest rate from 0.25% to 0.5% in November 2017 and then from 0.5% to 0.75% in August 2018.

Since then, the UK economy has slowed as firms' uncertainties about Brexit have become entrenched and growth in the world economy has eased. UK inflation has fallen back to our 2% target. So this month we have kept interest rates unchanged.

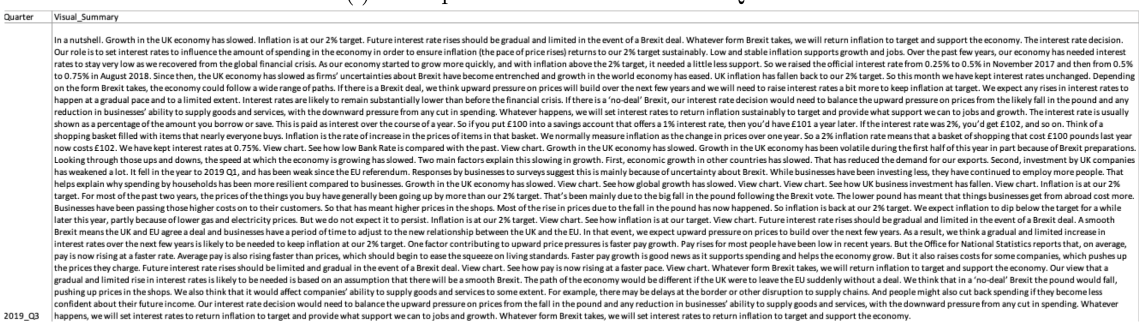
Depending on the form Brexit takes, the economy could follow a wide range of paths.

If there is a Brexit deal, we think upward pressure on prices will build over the next few years and we will need to raise interest rates a bit more to keep inflation at target. We expect any rises in interest rates to happen at a gradual pace and to a limited extent. Interest rates are likely to remain substantially lower than before the financial crisis.

If there is a 'no-deal' Brexit, our interest rate decision would need to balance the upward pressure on prices from the likely fall in the pound and any reduction in businesses' ability to supply goods and services, with the downward pressure from any cut in spending.

Whatever happens, we will set interest rates to return inflation sustainably to target and provide what support we can to jobs and growth.

(i) Example: A section of the 2019 Q3 VS



Quarter	Visual_Summary
2019_Q3	In a nutshell. Growth in the UK economy has slowed. Inflation is at our 2% target. Future interest rate rises should be gradual and limited in the event of a Brexit deal. Whatever form Brexit takes, we will return inflation to target and support the economy. The interest rate decision. Our role is to set interest rates to influence the amount of spending in the economy in order to ensure inflation (the pace of price rises) returns to our 2% target sustainably. Low and stable inflation supports growth and jobs. Over the past few years, our economy has needed interest rates to stay very low as we recovered from the global financial crisis. As our economy started to grow more quickly, and with inflation above the 2% target, it needed a little less support. So we raised the official interest rate from 0.25% to 0.5% in November 2017 and then from 0.5% to 0.75% in August 2018. Since then, the UK economy has slowed as firms' uncertainties about Brexit have become entrenched and growth in the world economy has eased. UK inflation has fallen back to our 2% target. So this month we have kept interest rates unchanged. Depending on the form Brexit takes, the economy could follow a wide range of paths. If there is a Brexit deal, we think upward pressure on prices will build over the next few years and we will need to raise interest rates a bit more to keep inflation at target. We expect any rises in interest rates to happen at a gradual pace and to a limited extent. Interest rates are likely to remain substantially lower than before the financial crisis. If there is a 'no-deal' Brexit, our interest rate decision would need to balance the upward pressure on prices from the likely fall in the pound and any reduction in businesses' ability to supply goods and services, with the downward pressure from any cut in spending. Whatever happens, we will set interest rates to return inflation sustainably to target and provide what support we can to jobs and growth. The interest rate is usually shown as a percentage of the amount you borrow or save. This is paid as interest over the course of a year. So if you put £100 into a savings account that offers a 1% interest rate, then you'd have £101 a year later. If the interest rate was 2%, you'd get £102, and so on. Think of a shopping basket filled with items that nearly everyone buys. Inflation is the rate of increase in the prices of items in that basket. We normally measure inflation as the change in prices over one year. So a 2% inflation rate means that a basket of shopping that cost £100 pounds last year now costs £102. We have kept interest rates at 0.75%. View chart. See how low Bank Rate is compared with the past. View chart. Growth in the UK economy has slowed. Growth in the UK economy has been volatile during the first half of this year in part because of Brexit preparations. Looking through those ups and downs, the speed at which the economy is growing has slowed. Two main factors explain this slowing in growth. First, economic growth in other countries has slowed. That has reduced the demand for our exports. Second, investment by UK companies has weakened a lot. It fell in the year to 2019 Q1, and has been weak since the EU referendum. Responses by businesses to surveys suggest this is mainly because of uncertainty about Brexit. While businesses have been investing less, they have continued to employ more people. That helps explain why spending by households has been more resilient compared to businesses. Growth in the UK economy has slowed. View chart. See how global growth has slowed. View chart. View chart. See how UK business investment has fallen. View chart. Inflation is at our 2% target. For most of the past two years, the prices of the things you buy have generally been going up by more than our 2% target. That's been mainly due to the big fall in the pound following the Brexit vote. The lower pound has meant that things businesses get from abroad cost more. Businesses have been passing those higher costs on to their customers. So that has meant higher prices in the shops. Most of the rise in prices due to the fall in the pound has now happened. So inflation is back at our 2% target. We expect inflation to dip below the target for a while later this year, partly because of lower gas and electricity prices. But we do not expect it to persist. Inflation is at our 2% target. View chart. See how inflation is at our target. View chart. Future interest rate rises should be gradual and limited in the event of a Brexit deal. A smooth Brexit means the UK and EU agree a deal and businesses have a period of time to adjust to the new relationship between the UK and the EU. In that event, we expect upward pressure on prices to build over the next few years. As a result, we think a gradual and limited increase in interest rates over the next few years is likely to be needed to keep inflation at our 2% target. One factor contributing to upward price pressures is faster pay growth. Pay rises for most people have been low in recent years. But the Office for National Statistics reports that, on average, pay is now rising at a faster rate. Average pay is also rising faster than prices, which should begin to ease the squeeze on living standards. Faster pay growth is good news as it supports spending and helps the economy grow. But it also raises costs for some companies, which pushes up the prices they charge. Future interest rate rises should be limited and gradual in the event of a Brexit deal. View chart. See how pay is now rising at a faster pace. View chart. Whatever form Brexit takes, we will return inflation to target and support the economy. Our view that a gradual and limited rise in interest rates is likely to be needed is based on an assumption that there will be a smooth Brexit. The path of the economy would be different if the UK were to leave the EU suddenly without a deal. We think that in a 'no-deal' Brexit the pound would fall, pushing up prices in the shops. We also think that it would affect companies' ability to supply goods and services to some extent. For example, there may be delays at the border or other disruption to supply chains. And people might also cut back spending if they become less confident about their future income. Our interest rate decision would need to balance the upward pressure on prices from the fall in the pound and any reduction in businesses' ability to supply goods and services, with the downward pressure from any cut in spending. Whatever happens, we will set interest rates to return inflation to target and provide what support we can to jobs and growth. Whatever form Brexit takes, we will set interest rates to return inflation to target and support the economy.

(ii) Snapshot of cleaned VS text

Figure 13: Example of Visual Summary page and cleaned text data

Table 4: Summary statistics of the linguistic complexity measures

Linguistic Complexity Measure	**Type***	**Obs**	**Mean**	**Std. Dev**	**Min**	**Max**
Word Count†	MPR	71	20,730	2,273	15,963	27,336
	MPS	32	997	226	618	1,474
	VS	23	950	258	566	1,558
Flesch-Kincaid	MPR	71	12.67	1.30	9.93	15.22
	MPS	32	13.31	0.89	11.84	15.46
	VS	23	6.30	0.88	4.79	8.62
Proportion of Jargon (%)	MPR	71	5.56	0.44	4.45	6.27
	MPS	32	8.95	1.12	6.68	10.68
	VS	23	5.82	1.08	3.89	8.48
McMahon-Naylor (MNCC) index (%)	MPR	71	17.64	2.44	12.67	23.98
	MPS	32	26.44	3.52	19.21	34.46
	VS	23	9.76	2.06	5.56	13.56

*Type represents Bank of England Publication Type. † Statistics reported to 0 decimal places. The remaining statistics are reported to 2 decimal places.

B.2.4 Word Cloud for MPR

Figure 14: Word Cloud for MPR (2005-2023)

Appendix C Survey Treatments: CB reports from a hypothetical economy

Figure 15: Text 1: Low Semantic, Low Conceptual Complexity

In a nutshell. Interest rates kept at 0.5%. The fall in the value of the pound has led to higher prices. The world economy is growing strongly. The squeeze in living standards is easing. Inflation will fall back towards our 2% target.

The economy now needs a little less support. We cut interest rates to exceptionally low levels during the supply crisis to support spending and to reduce the number of people out of work. Over the past few years our economy has needed interest rates to stay very low as we recovered from the global supply crisis. But things are changing. The world economy is now growing strongly. In this country, the share of people without a job is at its lowest level for over 40 years, and businesses are finding it hard to recruit people. Our economy is probably growing about as fast as it can without overheating (at 1.6%). And inflation is above our 2% target (at 3.0%). That means the economy needs a little less support than before. So last quarter, we raised the official interest rate we set, known as Bank Rate, from 0.25% to 0.5%. In this quarter, we have kept it at 0.5%.

Our job is to meet the 2% inflation target. Inflation is currently above that target, because of the big fall in the value of the pound resulting from the trade war. The weaker pound has meant that things businesses get from abroad cost more. Businesses will need to pass those rising costs on to their customers. So that has meant higher prices in the shops. The fall in the value of the pound happened around 18 months ago. Its effect on inflation doesn't last forever. And in the next few months inflation is going to start to fall back gradually towards our target.

Just like at home, the world economy had been quite weak following the supply crisis. But across Europe, in the US and many other countries the economy is now growing strongly. Stronger growth abroad will benefit our economy by increasing demand for our exports. And it should encourage companies to invest and recruit more staff to meet this extra demand.

Over the past year, prices have been rising faster than wages. That means people have not been able to afford as much. We think that is changing. The share of people out of work is now at its lowest level since 1975. And there are a lot of job vacancies. This means that companies are having to compete hard with each other to recruit and retain workers. One way they do that is by offering higher wages – so we expect bigger pay rises over the next few years. We think that pay will rise faster than prices this year, easing the squeeze on living standards.

To make sure inflation falls back to our 2% target, we need to set interest rates (the cost of borrowing) so that the amount of spending in the economy isn't too low or too high. If we set interest rates too low, then growth in the economy will be too fast, and inflation will stay above our target. But if we set interest rates too high or raise them too rapidly then growth will be too slow, and inflation will fall below our target. Put another way, we need to keep the economy growing at its speed limit. The speed limit for the economy is determined by two things: how many people are in work; and how productive the businesses they work for are. A few years ago many more people were out of work. So there was scope for the economy to grow quite quickly as a lot of those people found jobs. Now, with a record number of people in work, there isn't much more economic growth that can come from unemployed people finding work. Instead, it will mostly need to come from higher productivity - our ability to produce more with the people already in work and the resources that we have. But productivity has barely risen over the past decade. And we think that productivity will probably grow relatively slowly in coming years, too. We think that for inflation to settle back at the 2% target, the economy probably needs to grow at around 1.5% in coming years.

Inflation has been above our 2% target over the past year because of the sharp fall in the pound triggered by the trade war. During that time, we had to balance how quickly we take inflation back to the target with the support we give to jobs and activity. To ensure a sustainable return of inflation to the target, we need to keep economic growth around its new, lower, speed limit. With a strengthening world economy and more people in work, our economy now needs a little less support from us. We raised interest rates last quarter from 0.25% to 0.5%. If the economy continues to perform as expected, we think we will need to raise them further, reducing the amount of support we are providing to the economy. We expect any further rises in interest rates to happen at a gradual pace and to a limited extent. Interest rates are likely to remain substantially lower than a decade ago.

[PLACEHOLDER: ADD OTHER TEXTS]

Figure 16: Text 6: High Semantic, High Conceptual Complexity

At its meeting ending on 1st of this month, the MPC voted unanimously to maintain Bank Rate at 0.5%. The Committee voted unanimously to maintain the stock of sterling non-financial investment-grade corporate bond purchases, financed by the issuance of central bank reserves, at £10 billion. The Committee also voted unanimously to maintain the stock of government bond purchases, financed by the issuance of central bank reserves, at £435 billion. More detail regarding the MPC's latest projections for output and inflation is available on Central Bank's website.

The global economy is growing at its fastest pace in five years. The expansion is becoming increasingly broad-based and investment driven. Notwithstanding recent volatility in financial markets, global financial conditions remain supportive. Net trade is benefiting from robust global demand and the past depreciation of sterling. Along with high rates of profitability, the low cost of capital and limited spare capacity, strong global activity is supporting business investment, although it remains restrained by trade war-related uncertainties. Household consumption growth is expected to remain relatively subdued, reflecting weak real income growth. GDP growth is currently at 1.6% and is expected to average around 1.75% over the forecast, a slightly faster pace than was projected last quarter despite the updated projections being conditioned on the higher market-implied path for interest rates and stronger exchange rate prevailing in financial markets at the time of the forecast. While modest by historical standards, that rate of growth is still expected to exceed the diminished rate of supply growth.

Following its annual assessment of the supply side of the economy, the MPC judges that the economy has only a very limited degree of slack and that its supply capacity will grow only modestly over the forecast, averaging around 1.5% per year. This reflects lower growth in labour supply and rates of productivity growth that are around half of their pre-crisis average. As growth in demand outpaces that of supply, a small margin of excess demand emerges in 24 months' time and builds thereafter. CPI inflation fell from 3.1% last month to 3.0% this month. Inflation is expected to remain around 3% in the short term, reflecting recent higher oil prices. More generally, sustained above-target inflation remains almost entirely due to the effects of higher import prices following sterling's past depreciation. These external forces slowly dissipate over the forecast, while domestic inflationary pressures are expected to rise. The firming of shorter-term measures of wage growth in recent quarters, and a range of survey indicators that suggests pay growth will rise further in response to the tightening labour market, give increasing confidence that growth in wages and unit labour costs will pick up to target-consistent rates. On balance, CPI inflation is projected to fall back gradually over the forecast but remain above the 2% target in the second and third years of the MPC's central projection.

As in previous Reports, the MPC's projections are conditioned on the average of a range of possible outcomes for the economy's eventual trading relationship with its major trading partner. The projections also assume that, in the interim, households and companies base their decisions on the expectation of a smooth adjustment to a recovered trading relationship. Developments regarding the economy's trade war with its major trading partner — and in particular the reaction of households, businesses and asset prices to them — remain the most significant influence on, and source of uncertainty about, the economic outlook. In such exceptional circumstances, the MPC's remit specifies that the Committee must balance any trade-off between the speed at which it intends to return inflation sustainably to the target and the support that monetary policy provides to jobs and activity.

Over the past year, a steady absorption of slack has reduced the degree to which it was appropriate for the MPC to accommodate an extended period of inflation above the target. Consequently, at its last meeting, the Committee tightened modestly the stance of monetary policy (raising the cost of borrowing) in order to return inflation sustainably to the target. Since then, the prospect of a greater degree of excess demand over the forecast period and the expectation that inflation would remain above the target have further diminished the trade-off that the MPC is required to balance. It is therefore appropriate to set monetary policy so that inflation returns sustainably to its target at a more conventional horizon.

The Committee judges that, were the economy to evolve broadly in line with this period's Inflation Report projections, monetary policy would need to be tightened somewhat earlier and by a somewhat greater extent over the forecast period than anticipated at the time of last period's Report, in order to return inflation sustainably to the target. In light of these considerations, all members thought that the current policy stance remained appropriate to balance the demands of the MPC's remit. Any future increases in Bank Rate are expected to be at a gradual pace and to a limited extent. The Committee will monitor closely the incoming evidence on the evolving economic outlook, and stands ready to respond to developments as they unfold to ensure a sustainable return of inflation to the 2% target.

Appendix D Further results

Table 5: Baseline Results for broadly defined complexity

	Perceived	**Actual Understanding**			**Sentiments towards CB**		
	Understanding	Inf(t)	i(t)	Exp Pay	Trust	Attention	BoE Role
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Medium Complexity	−0.067	−0.043	−0.001	−0.030	−0.038	−0.042	−0.041
	(0.051)	(0.026)	(0.026)	(0.026)	(0.050)	(0.060)	(0.057)
High Complexity	−0.824***	−0.106***	−0.206***	−0.224***	−0.210***	−0.403***	−0.496***
	(0.059)	(0.030)	(0.030)	(0.030)	(0.056)	(0.069)	(0.065)
Age	0.004*	−0.001	0.003***	−0.001	0.007***	0.003	0.0004
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
UK country of birth	0.048	−0.002	−0.009	−0.015	−0.103*	−0.234***	−0.033
	(0.059)	(0.030)	(0.030)	(0.030)	(0.056)	(0.068)	(0.065)
Income	0.167***	0.012	0.026**	0.017	0.055**	0.032	0.072***
	(0.022)	(0.011)	(0.011)	(0.011)	(0.021)	(0.026)	(0.025)
Econ at Uni	0.449***	−0.033	0.021	−0.049*	0.117**	0.222***	0.254***
	(0.051)	(0.026)	(0.026)	(0.026)	(0.049)	(0.059)	(0.056)
Pre-anchored Exps	0.517***	0.233***	0.174***	0.093***	0.146***	0.121**	0.320***
	(0.047)	(0.024)	(0.024)	(0.024)	(0.045)	(0.055)	(0.052)
Constant	2.187***	0.356***	0.303***	0.678***	1.471***	2.175***	1.741***
	(0.091)	(0.047)	(0.047)	(0.046)	(0.088)	(0.107)	(0.101)
Observations	1,745	1,745	1,745	1,745	1,742	1,743	1,745
R^2	0.267	0.062	0.088	0.048	0.046	0.050	0.089

Note: *p<0.1; **p<0.05; ***p<0.01