A Macroeconomy with Intuitive Thinkers

Preliminary draft

Maren Bartels* Johannes Binswanger† Albert Flak‡
Manuel Oechslin§

February 29, 2024

Abstract

A growing body of evidence shows that, when forming expectations, households, firms, and even experts often deviate from rational expectations, adhering to intuitive models about macroeconomic relationships that conflict with actual experience. The "stagflationary" intuitive model—high output comes with low inflation—is a prominent example. Starting with a linear difference model à la Blanchard and Kahn (1980), we develop a generic macroeconomic framework in which expectations emerge from an interplay of two mental systems. A rigorous thinking system forms expectations corresponding to standard rational expectations. An intuitive system forms expectations based on associative memory, perturbing rational expectations. As a result, households behave as if they were subject to cognitive discounting, and autonomous innovations in intuitive expectations are a source of macroeconomic fluctuations. We illustrate the tractability of the framework by applying it to stagflationary expectations and a New Keynesian model. Our analysis also informs empirical analysis concerning identification strategies about shocks to expectations.

JEL classification: E71, E31

Keywords: Behavioral macro, dual system processing, associative memory, expectations

*University of Lucerne, Department of Economics, Switzerland. Email: maren.bartels@unilu.ch.
†University of St. Gallen, Department of Economics, Switzerland. Email: johannes.binswanger@unisg.ch.
‡University of St. Gallen, Department of Economics, Switzerland. Email: albert.flak@unisg.ch.
§University of Lucerne, Department of Economics, Switzerland. Email: manuel.oechslin@unilu.ch.

1 Introduction

Survey evidence increasingly shows that macroeconomic expectations of households, managers in firms, and—to some degree—even professional forecasters systematically deviate from the full-information rational expectations (FIRE) benchmark. For example, when examining inflation, Coibion and Gorodnichenko (2012, 2015a) observe for the US that survey-measured expectations tend to underreact to innovations in fundamentals, a pattern that aligns well with theories of informational frictions. Another recent strand of the literature focuses on deviations involving *relationships* among multiple macroeconomic variables. Dräger et al. (2016) document that in a sample of US consumers, a large majority holds inflation and unemployment expectations that are inconsistent with the Phillips curve. Andre et al. (2022) show that, in the context of monetary policy shocks, US consumers tend to form expectations for inflation and unemployment as if they were using the subjective model of a "cost channel", a behavior that sets them apart from experts and possibly reflects selective retrieval from memory of directly experienced partial-equilibrium effects.[1] Bhandari et al. (2023) find, among other things, that the deviations from FIRE fluctuate over time and are wider during recessions. Occasionally, there are even moments when consumer economic sentiment seems largely divorced from the facts.[2]

In this paper, we primarily contribute to the recent literature on expectations about macroeconomic relationships, from both an empirical and a theoretical perspective. On the empirical side, we add to the existing findings by estimating autonomous innovations—i.e., shifts not driven by innovations in fundamentals—in consumers' expectations for inflation and output, using survey data that have been purified of supply-side shocks. Importantly, the autonomous innovations show a clear relationship: a favorable innovation to inflation expectations on average comes with a favorable innovation to output expectations, and vice versa. This reflects what we term an "all good/bad in one" intuitive model in the minds of consumers. We further provide suggestive evidence that those autonomous innovations in expectations may impact actual outcomes for inflation and output.

On the theoretical side, we incorporate the cognitive pattern of dual-system processing

[1] Similarly, in their examination of narratives as causal accounts of the US inflation surge in 2021 and 2022, Andre et al. (2023) observe marked disparities between experts and consumers, with the narratives embraced by the latter often showing a more selective (mostly supply-side) perspective.

[2] One such moment came in the US after the post-covid inflation, when the average US consumer had a very negative view of a fundamentally strong economy (The Economist, Jan 20, 2024).

into expectation formation; in the psychological literature, this pattern has been found of key importance for difficult cognitive tasks. Forming expectations—or making forecasts—about macroeconomic variables is a cognitively challenging task. It requires screening a wide field of variables for potential signals, assessing the informational content of each signal and hence weighing it appropriately. What is more, by the dynamic nature of economic decisions, expectations often have a recursive structure, such that the expectation of macroeconomic variables for next period depends on next period's expectations about values of those variables further ahead in the future. Arguably, at least for all those who are not professional forecasters (if not even for the latter), this is challenging indeed. In fact, many decisions that are of key importance at the macroeconomic level are made by households and firms. As forcefully argued by Coibion and Gorodnichenko (2015b) and Candia et al. (2021), firms' expectations differ substantially from those of professional forecasters and are often more closely aligned with household expectations.

Decades of research in psychology have shown that when faced with cognitively challenging tasks, individuals partly resort to *intuitive thinking*, rather than fully engaging in *rigorous thinking*. This pair of cognitive systems gives rise to what is known in psychology as the dual processing model of cognition (Kahneman 2011, Evans and Stanovich 2013, Ilut and Valchev 2023). A key aspect of our theoretical framework is that it encompasses two versions of expectations: those based on rigorous thinking and those based on intuitive thinking. In our framework, rigorous expectations are consistent with the data-generating processes of the fundamentals as it would materialize in a FIRE world, while the intuitive expectations may not be. Instead, they emerge from *associative memory* (Bordalo et al. 2023, Andre et al. 2022). We also endogenously derive the pattern of innovations in inflation and output expectations from a model of associative memory based on Bordalo et al. (2023), and we show that this endogenously leads to the "all good/bad in one" heuristic. The two versions of expectations are combined into one, which we refer to as a dual-system expectation. As a concrete application, we introduce dual-system expectations into a textbook version of the New Keynesian (NK) model and provide quantitative results, using a parametrization informed by our empirical analysis.

To empirically estimate autonomous innovations in (consumer) expectations, which we understand as driven by intuitive thinking, we use data from the University of Michigan's Survey of Consumers. This survey, representative of all US households, spans the years from 1960 to 2022. The survey allows us to construct proxies for the average US household's expectations regarding output changes and inflation over the next 12 months. Looking at the raw data, we find that positive expectations regarding output changes are often accompanied by low expectations for inflation. Such a negative correlation is not entirely unexpected: it reflects a finding by Kamdar (2019), who, also using the Michigan survey, documents a positive correlation between unemployment and inflation expectations.[3] Nor does such a negative correlation in the minds of US households necessarily collide with reality, as it could reflect the frequent experience of total factor productivity (TFP) shocks. However, when we purify the survey data of variation coming from TFP shocks, the negative correlation remains. Moreover, when we employ vector autoregressive (VAR) models to additionally filter for autocorrelation patterns to obtain a measure of autonomous innovations in expectations, those innovations still show the negative "stagflationary" correlation. This robustness of results suggests a deeply ingrained pattern in the intuitions of US consumers, corresponding to the "all good/bad in one" intuitive model. Beyond that, we use local projections to provide suggestive evidence that a structural form of the autonomous innovations in expectations, estimated by a structural VAR, impact US inflation and output. For example, we find a significantly positive and sizable effect of autonomous innovations in inflation expectations on actual inflation.

The starting point for our theoretical model is to introduce intuitive expectations into a generic linear difference model in the form of Blanchard and Kahn (1980). Intuitive expectations are introduced in a parsimonious way. Our representative agent agrees with the FIRE counterpart on the primitive building blocks of the economy and their structural relationships, as long as expressions involving expectations are not "solved". In other words, our representative agent agrees with a FIRE agent on all equations of the model, as long as expectation terms are left as such. Intuitive expectations only enter in an autonomous way when the agent must find a solution for the model in quantitative terms and thus "solve" expectation terms At this point, intuitive expectations perturb the rigorous ones, which follow the reasoning of a FIRE agent. The integration of the two thinking modes to form dual-system expectations is governed by a single weighting parameter. As this parameter approaches one, our framework reverts to the rational expectations benchmark.

As a consequence of dual-system expectations, households behave as if they were subject to cognitive discounting in the sense of Gabaix (2020). This leads to dampened impulse

[3] Working with belief wedges, which are defined as the deviation of survey responses from the corresponding rational expectations benchmark, Bhandari et al. (2023) also find that a positive (negative) belief wedge in unemployment is associated with a positive (negative) belief wedge in inflation.

responses for standard fundamental shocks, such as TFP shocks. However, unlike in the case of Gabaix (2020), the dampening does not lead to a shrinkage towards the steady state but rather a shrinkage towards the autonomous perturbations introduced by intuitive thinking based on associative memory. In our model, these perturbations present an additional source of variation for macroeconomic variables. Our framework retains the recursive structure of expectations, determined by the forward solution of a linear stochastic difference equation. This recursive structure is a cornerstone of forward-looking economic reasoning. However, because of dual processing, only part of the expectation formation process enters the recursion—the part associated with rigorous thinking. It is due to the fact that expectations are only partially recursive that cognitive discounting emerges.

Our model also provides a guideline for empirical work on expectation or belief shocks. It helps assessing about when it is meaningful to decompose innovations in expectations—in the form of correlated VAR innovations—into mutually independent structural shocks. Due to the mechanics of associative memory, intuitive expectations for different variables—such as output and inflation— may be "packaged together". As such, they can be meaningfully be represented as correlated VAR innovations, but it may not always be meaningful to decompose them into independent structural components. In this respect, expectation shocks, as far as they concern intuitive expectations, are unlike fundamental shocks (such as TFP shocks) where a structural decomposition is generally meaningful. We illustrate this by applying the model of associative memory to (simultaneous) intuitive expectation formation about output and inflation.

The building blocks of our model are modularized. The interplay between the rigorous and intuitive mental system does not depend on the specific mechanics of how these intuitions emerge. In fact, if a researcher desires to abstract from the inner workings of the intuitive system, it can be represented by a set of exogenous state variables, the dynamics of which are guided by appropriate assumptions. Our microfoundation of intuitive thinking based on associative memory serve the purpose of endogenously deriving what these assumptions should be. Our framework is also very flexible. In particular, our approach to integrating autonomous innovations in expectations works for any linear difference model, including many variants of the Dynamic Stochastic General Equilibrium (DSGE) model. As a concrete application, this paper shows how it works with the textbook New Keynesian (NK) model of Galì (2015).

The rest of this paper is organized as follows. The next section discusses the related literature. Section 3 presents motivating evidence, with a focus on documenting the existence of autonomous innovations in expectations that show a "stagflationary" pattern. Section 4 presents our model in general terms. It first discusses the interplay of rigorous and intuitive thinking in the realm of a linear difference model à la Blanchard and Kahn (1980). It then presents a model of the inner workings of intuitive thinking based on associative memory. Section 5 applies the previous section's theoretical framework to a textbook version of the NK model, also offering a quantitative analysis. Section 6, finally, concludes.

2 Literature

Deviations from the full-information rational expectations benchmark are widespread. People generally underappreciate publicly available information when forming expectations (Bianchi et al. 2022). Even experts deviate from FIRE when making forecasts, either under-reacting (Coibion and Gorodnichenko 2012, 2015a) or over-reacting (Bordalo et al. 2020b) to news about fundamentals. Households, in turn, often hold expectations for macroeconomic variables, and the relationships among them, that conflict with experts' expectations and the macroeconomic record, as a growing body of survey-based research shows (e.g., Dräger et al. 2016; Kamdar 2019; Candia et al. 2020; Bhandari et al. 2023). Sometimes, as Candia et al. (2020) note, "it's as if the Phillips curve perceived by households is upward-sloping." (p. 9). Focusing on autonomous innovations in expectations estimated via a VAR, we find a similar "stagflationary view" among US households, even after taking account of TFP, oil price, and IST (investment-specific technology) shocks.

Ilut and Valchev (2023) provide a seminal application of dual-system processing for consumption and savings decisions. In their model, the agent initially only has a vague prior about the policy function relating optimal consumption levels to income states. The intuitive System 1 provides a default decision modeled as a prior belief that is available based on experience from the past without any effortful thinking. If this prior features high uncertainty, the agent finds it worthwhile to engage in rigorous thinking and so obtaining a posterior belief that is more informative. These dynamics have the effect that if the agent encounters a state that they have already experienced multiple times, they feel familiar with that state and stop engaging in rigirous thinking. This can happen even if the consumption choice suggested by System 1 deviates strongly from the optimal choice of a FIRE agent. Ilut and Valchev (2023) refer to this outcome as a learning trap. Our model of the interplay between the two mental system is less sophisticated. In our case, the "problem solving share" of rigorous thinking is constant and hence does not depend on history. We choose this simpler approach since, unlike Ilut and Valchev (2023), we consider a general-equilibrium setting with multiple state variables. Moreover, we want to preserve scope for a more detailed modelling of the inner workings of intuitive thinking based on associative memory. For this, we benefit from the reduced complexity resulting from the assumption of constant shares for the two mental systems.

Our modelling of the inner working of intuitive thinking builds upon Bordalo et al. (2020a) and, in particular, on Bordalo et al. (2023). These seminal studies introduce associative memory in a tractable way from the fields of cognitive psychology and neuroscience. They show how expectations and probability assessments can be heavily influenced by contextual cues that trigger the appearance of certain memories. Importantly, memories pertaining to a small and homogeneous class of experiences, a class with high "self-similarity", tend to come to mind much more easily than those pertaining to large and diverse classes of experiences. Our own model is more geared towards macroeconomic applications; the agent forms expectations about macroeconomic variables rather than forming norms, as in Bordalo et al. (2020a), or assessing probabilities, as in Bordalo et al. (2023). We also use a somewhat different formalism. Among other things, to model (dis)similarity within a group of experiences, we use entropy, which has also been linked to the mechanics of associative memory in neuroscience (Pineda et al. 2021, Pineda and Morales 2023).

Kamdar (2019) uses the rational inattention framework to offer a potential explanation for how completely rational consumers could arrive at such a stagflationary view. The paper demonstrates that when there are information costs of intermediate size, rationally inattentive consumers optimally gather information in such a way that the covariance of the posterior beliefs about labor market slackness and the price level is positive, even though the true covariance is modeled as zero. This result is very intriguing; however, it is obtained for a rather specific combination of technical assumptions regarding the nature of the signals that can be obtained, the nature of the shocks to labor market slackness and the price level, and the utility function of the consumers. Moreover, the desirable feature that the rational inattention model approaches the FIRE-benchmark as information costs go to zero is only obtained if there is a correspondence between the number of choice variables and the number of unknown states. It is thus unclear whether the rational inattention model is robust enough to generate a stagflationary view (and to have the FIRE-benchmark as a limiting case) across a wide range of linear difference models à la Blanchard and Kahn (1980). Broad applicability is an inherent feature of our approach, in which fundamental psychological concepts such as dual-system processing and associative memory form the basis of the stagflationary intuitive model in the mind of the average consumer.

Documenting a stagflationary (or “all good/bad in one”) intuitive model in the mind of the average US consumer also connects us with recent research on narratives, which are understood as causal models mapping actions to consequences (Eliaz and Spiegler 2020) or causal accounts for why an economic event occurred (Andre et al. 2023). Most notably, combining survey-based measurement of narratives with experimental treatments, Andre et al. (2023) explore narratives prevalent among US households regarding the surge in inflation during 2021-22 period. While there is substantial heterogeneity, narratives typically center around supply-side factors, thereby neglecting the demand side. However, narratives seem to go beyond helping consumers make sense of their macroeconomic environment. Shiller (2017, 2019) describes them as causative innovations that may have an impact on the economy, thereby driving economic fluctuations. Our finding, though suggestive in nature, that autonomous innovations in households’ expectations affect inflation and output is consistent with this view.[4]

Our findings regarding inflation and output are also in line with recent empirical results in the literature on possible effects of belief shocks on macroeconomic variables. Enders et al. (2021) extract belief shocks from nowcast errors by professional forecasters. They show that a favorable (adverse) belief shock regarding future output growth causes a subsequent rise (decline) in economic activity. Earlier contributions to the literature on the effects of belief shocks produced mixed results. Using a DSGE model for estimation, Barsky and Sims (2012) do not find that noise in beliefs about future productivity growth—labeled “animal spirits”—has a sizable impact on macroeconomic variables. This contrasts with Lorenzoni (2009), a theoretical study that takes a somewhat different approach to modeling noise in beliefs. According to the simulations presented there, belief shocks, by affecting aggregate demand, can generate sizable short-run effects when it comes to output, employment, and inflation.

Finally, this paper also relates to more recent theoretical work on behavioral elements in macroeconomic models, in particular work on higher-order uncertainty (Angeletos et al. 2018a) and cognitive discounting (Gabaix 2020). Angeletos et al. (2018a) introduce autonomous innovations in higher-order beliefs into variants of the DSGE model, a modification that allows waves of optimism or pessimism to produce autonomous deviations from FIRE in consumers’ and firms’ expectations. Gabaix (2020) posits that agents are myopic about potential devia-

[4]Focusing on US public firms rather than US households, Flynn and Sastry (2022) find that hiring decisions are significantly affected by narratives. Through this channel, fluctuations in narratives are estimated to contribute significantly to fluctuations in output.

tions from the steady state when it comes to the more distant future, resulting in a significantly higher discounting of future variable realizations compared to FIRE. Both autonomous innovations in expectations and heavier than rational discounting also arise in our framework—but as a consequence of dual-system processing (Kahneman 2003, 2011; Bordalo et al. 2020a). Our approach is to introduce dual-system processing with a reduced-form component, a choice influenced by our macroeconomic setting. Invoking dual-system processing also sets us apart from Bhandari et al. (2023), who obtain deviations from FIRE by allowing agents to have concerns about model misspecification. Finally, in a paper studying monetary policy in an experimental setting, Hommes et al. (2019) consider expectation formation based on heuristics that allow for learning from past mistakes. The aim of their framework is not to introduce an autonomous source of variation in expectations that can drive macroeconomic fluctuations.

3 Motivating Evidence

In this section, we provide motivating evidence for the fact that household expectations may exhibit patterns of intuitive thinking that would be absent for FIRE agents. For this, we use data on US households' expectations from the Michigan Survey of Consumers (MSC). We focus on how US households perceive the relationship between output and inflation, which plays a key role in macroeconomics. While respondents in the MSC are broadly representative for the US population, their thinking may be more strongly determined by intuitive reasoning than in the case of, say, professional forecasters. Thus, we are more likely to find patterns associated with intuitive thinking in the MSC rather than, say, the Survey of Professional Forecasters, and we may therefore tilt the evidence in favor of our approach. However, as argued by Coibion and Gorodnichenko (2015b) and Candia et al. (2021), expectations of price setters in firms appear to resemble more those of average households than professional forecasters. Moreover, on average, aggregate consumption decisions are heavily influenced by households rather than sophisticated professionals. We therefore see the MSC as an appropriate source for motivating evidence on the presence of intuitive thinking in expectation formation.

The MSC collects a wide range of responses, including households' one-year-ahead expectations regarding changes in general business conditions, and the price level. We use quarterly data for the entire available horizon of 1960:Q1 until 2022:Q4, giving us 252 observations. Our expectation measures are based on the answers to the following two questions from the MSC:[5]

- INF1: *By what percent do you expect prices to go up, on the average, during the next 12 months?*
- GDP1: *And how about a year from now, do you expect that in the country as a whole, business conditions will be better, or worse than they are at present, or just about the same?* (Answers: "Better", "Same", "Worse", "Don't Know / NA")

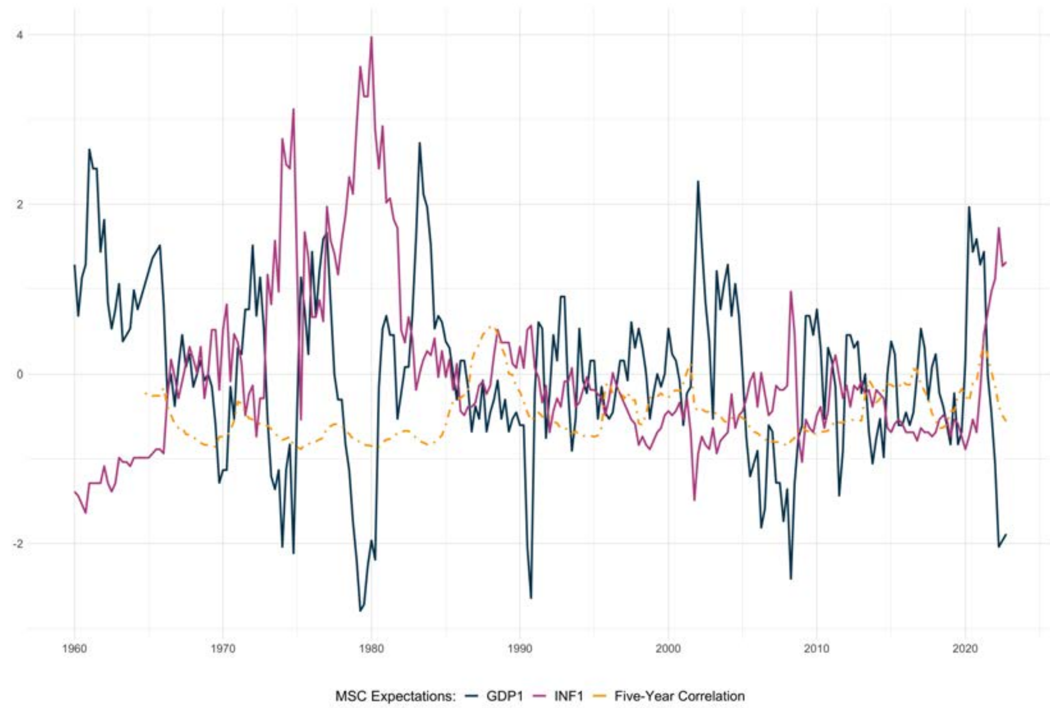
If, on average, consumers think about the business cycle as being driven primarily by demand factors, we should observe a positive contemporaneous correlation between the INF1 and GDP1 series. Even though many experts view modern booms and busts mostly demand-driven, the MSC respondents appear to perceive the opposite, namely supply-driven cycles. For a first impression, this can be seen in Figure 1 which plots the two raw series INF1 and GDP1, together with their correlation over a 5-year rolling window. The latter is almost always negative. It is also clearly visible that spikes of inflation expectations in the positive domain are accompanied by spikes of output expectations in the negative domain, and vice versa. This pattern has been well documented by, among others, Kamdar (2019), Candia et al. (2020), and Bhandari et al. (2023).

Of course, Figure 1 does not allow for any direct conclusion about whether individual expectations are in conflict with reality. After all, to the extent that fluctuations in output and inflation are driven by supply factors, we would expect the relationship between the two series to be negative. For instance, technology shocks have a positive effect on output and, at the same time, also lower the costs of production. What is more, individuals may anticipate future technology shocks (in the form of "news shocks", see Barsky and Sims 2012).

To get a better sense to what degree expectations are consistent with facts, we follow Enders et al. (2021) and use various measures of exogenous technology shocks to "purify" MSC expectations from variation attributed to those shocks. We use the following purification schemes: *(i)* the current value, and 8 lags of TFP using Fernald (2014)'s measures of Business Sector TFP and utilization-adjusted TFP, and IST (investment-specific technology) news shocks from Ben Zeev (2018); *(ii)* the previous list, agumented by 8 leads for TFP, and 8 lags of oil news

[5] More precisely, we use the column "Mean" from Table 32 of the MSC for INF1, and the column "Relative" from Table 26 for question GDP1. "Relative" is the difference between the proportion of respondents answering "Better" and the proportion of respondents answering "Worse" plus 100.

Figure 1: Inverse Relationship Between Output and Inflation Expectations in the MSC



NOTES: The data plotted correspond to the column "Relative" in Table 26 of MSC ("GDP1") and the column "Mean" in Table 32 of MSC ("INF1"). These two time series are standardised to a mean of zero and a variance of one, such that, for example, a value of 2 on the vertical axis implies a 2 standard deviations difference to the mean of the series. The correlation plotted is the five-year rolling-window Pearson correlation coefficient between the two standardised series.

shock measures from Känzig (2021)[6]; *(iii)* we add to the previous list the current value and 8 lags and 8 leads of real government consumption expenditures and gross Investment taken from the FRED database. Current values and lags of TFP, IST and oil shocks are obvious controls for supply shocks. We also include TFP *leads* since MSC respondents may actually be able to anticipate future TFP developments. We do not include leads for the IST and oil shocks, however, since the mentioned measures are already news shocks and hence of anticipatory nature. In the last scheme we add 8 lags and leads of government expenditures as a crude control for demand factors, which we will use to shed additional light on some results.[7] To obtain purified series for MSC expectations, we regress both *INF1* and *GDP1* separately on the measures in

[6]The oil news supply shock from Känzig (2021) is only available for the subsample of 1975:Q1 – 2022:Q4. When adding this measure, we lose 15 years of observations.

[7]We do not use measures of identified demand shocks since they tend to capture very little variation in output and inflation (Ramey 2016).

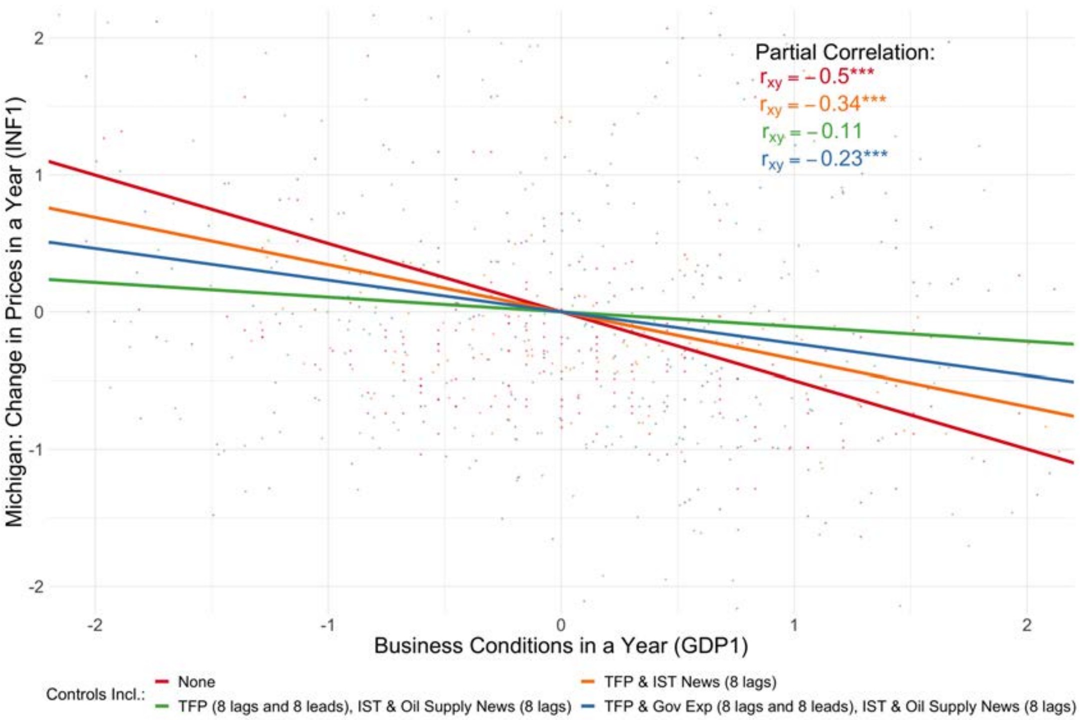
the above list to obtain residuals.[8] Arguably, TFP, IST, and oil shocks capture a large part of the variation in technology and other supply side factors. The remaining drivers of fluctuations in output and inflation are therefore likely to be associated with the demand side. Thus, we would expect the correlation of purified output and inflation expectations turning to positive. The data, however, tell a different story.

Figure 2 shows regression lines for regressions pertaining to the different purification schemes. It also includes a regression line for the case of no purification. In each case, both variables are standardized. To start with, the red line refers to the case of no purification. Due to standardization, the regression coefficient in this univariate regression is equal to the correlation coefficient and amounts to -0.5 (significant at the 1-percent level). The orange line represents the regression in which both expectation variables are regressed on 8 lags of TFP and IST news shocks. While the line becomes flatter, the correlation still amounts to -0.34 and remains significant at the 1-percent level. As a next step, we add 8 leads of TFP and 8 lags of oil news shocks to the purification. The correlation then amounts to -0.11 and is no longer significant. Recall, however, that the respective purification scheme only includes supply factors. To the degree that demand fluctuations play an important role for macroeconomic fluctuations, we would expect the correlation to be significantly positive rather than zero. When we add 8 lags and leads of government spending as a crude control for demand factors, the correlation becomes again more negative and is again statistically significant at the 1 percent level.

The empirical evidence so far suggests that agents' perception of the relationship between output and inflation is not (exclusively) driven by its factual nature but rather (also) by an intuitive model. This raises a question that is of particular importance in this paper: do such intuitive models help us to better understand macroeconomic fluctuations? Adapting this question to a standard macroeconomic framework leads us to ask whether *innovations* or *shocks* to intuitive models trigger impulse responses in key macroeconomic variables. As a next step in this direction, it is therefore of interest to consider a VAR estimation for the MSC series and to investigate whether traces of the same intuitive model that appeared in the results discussed above also appear in VAR innovations. For technical reasons, we do not directly estimate a VAR for the purified series used above. Rather we run VAR-X specifications with the two MSC expectations as the VAR variables and the purification variables as the exogenous

[8] We run regressions for many other purification schemes and obtain results very similar to the ones presented below.

Figure 2: Relationship between change in output and inflation in MSC Expectations



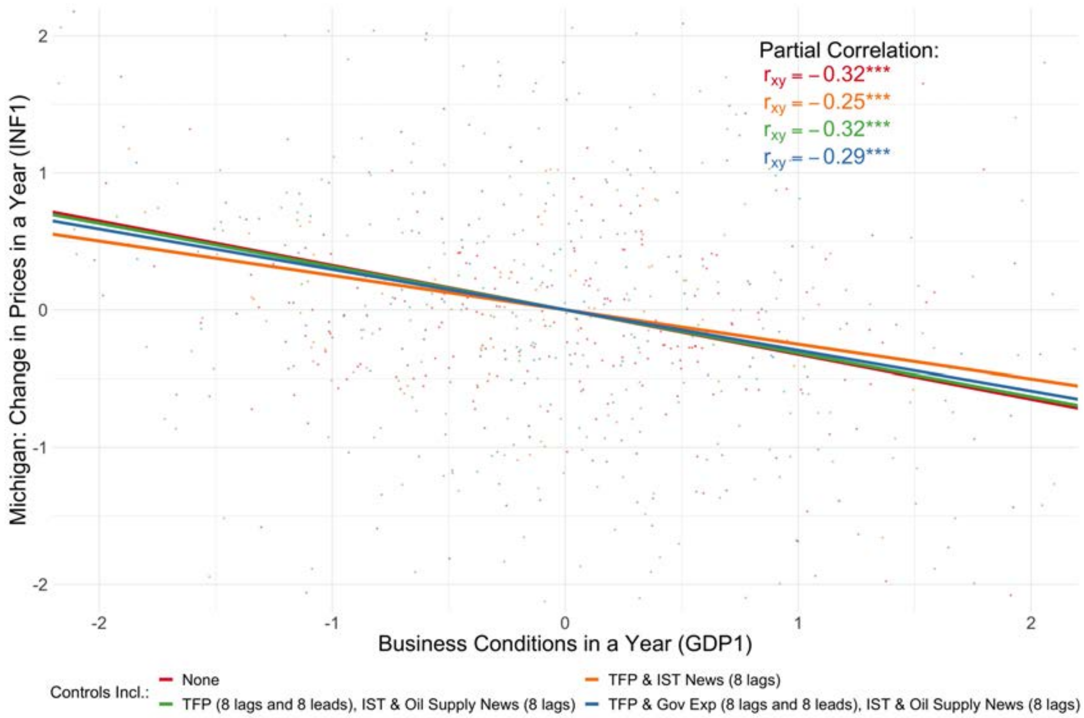
NOTES: The slopes of the plotted lines are estimated from regressions of our measure for INF1 ("mean") on GDP1 ("relative"). Both variables are purified by regressing them on the current realization, lags and leads of TFP, and the current realization and lags of oil news and IST news shocks, as discussed in the main text. The red line refers to the case with no purification. For each regression, both variables are standardized after purification. Significance levels (p-values): *** $p \leq 0.01$; ** $0.01 < p \leq 0.05$; * $0.05 < p \leq 0.1$.

X-controls.[9] Below, we report results where the VAR includes one autoregressive lag for MSC expectations. We use the VAR-X estimations to extract residuals and then run regressions for the residuals for inflation expectations on the residuals for output expectations. Figure 3 shows the results for the regressions for the VAR-X residuals, or innovations. All correlations are now significantly negative at the 1-percent level and the different purification schemes play almost no role for the value of the correlation.

Since the VAR-X innovations are correlated, it is standard practice to ask whether there is a meaningful scheme that identifies independent "structural" expectation shocks. Our model

[9]If we directly used the purified MSC series from the previous regressions, there would emerge a correlation between VAR innovations and contemporaneous or lead values of the supply side controls (such as TFP) since the (RHS) lags of the VAR variables have not been purified with literally the same contemporaneous or lead values as the (LHS) contemporaneous VAR variables. This correlation would defy the purpose of purification and lead to biases in the local projection estimates presented below.

Figure 3: Downward-Sloping NKPC in VAR(1) of MSC Forecast Errors



NOTES: The slopes of the plotted lines are based on estimated residuals from a VAR-X model for INF1 and GDP1, where the X-controls include the purification variables as explained in the main text and indicated in the legend. The red line refers to the case with no purification. Significance levels (p-values): *** $p \leq 0.01$; ** $0.01 < p \leq 0.05$; * $0.05 < p \leq 0.1$.

of intuitive thinking based on associative memory (see Section 4.4) suggests a particular linear identification scheme.[10] Denote the VAR innovations for output and inflation expectations by η_t^y and η_t^π, respectively. Denote the structural shocks that are to be identified by ε_t^y and ε_t^π, respectively. Identification then relies on finding a 2-by-2 matrix H such that $\eta_t = H\varepsilon_t$, where η_t and ε_t are vectors collecting the output and inflation expectation components, respectively. Writing this equation explicitly, we have

However, to the degree that a (linear) structural decomposition *is* meaningful, the model in Section 4.4 suggests a particular (linear) identification scheme. Denote the VAR innovations for output and inflation expectations by η_t^y and η_t^π, respectively. Denote the structural shocks that are to be identified by ε_t^y and ε_t^π, respectively. Identification then relies on finding a 2-by-2

[10]As will become clear in Section 4.4, with associative memory, is not always meaningful to decompose correlated innovations to intuitive thinking into independent components. The reason is that they may come “packaged together”, and looking “through” this package is not necessarily meaningful.

matrix H such that $\eta_t = H\varepsilon_t$, where the two vectors η_t and ε_t are containing an output- and an inflation-related component, respectively. Writing this equation explicitly, we have

$$\begin{aligned} \eta_t^y &= h_{yy}\varepsilon_t^y + h_{y\pi}\varepsilon_t^\pi \\ \eta_t^\pi &= h_{\pi y}\varepsilon_t^y + h_{\pi\pi}\varepsilon_t^\pi \end{aligned} \tag{1}$$

where the subscripts of the coefficients h_{ij} characterize their positions in H in an obvious way. Obviously, it is natural to assume that $h_{yy} > 0$ and $h_{\pi\pi} > 0$. Therefore, for the covariance between the two innovations in η_t to be negative, one or both of the off-diagonal elements $h_{y\pi}$ and $h_{\pi y}$ must be negative. The correlation is guaranteed to be negative if both off-diagonal elements are negative. Moreover, to the degree that a linear identification scheme is meaningful, the model in Section 4.4 also suggests that *both* elements should be negative. In a nutshell, this reflects an "all good/bad in one" heuristic. In the intuitive model, a high innovation to output expectation ("good") is associated with a low innovation to inflation expectation ("good"). The higher the innovation to output expectation, the lower the innovation to inflation. Hence, $h_{\pi y}$ should be negative. Similarly, a high innovation in inflation expectation ("bad") is associated with a low innovation in output expectation ("bad") and so the same logic applies. Our identification scheme now sets

$$h_{y\pi} = \tau h_{\pi y}, \tag{2}$$

with $\tau > 0$ exogenously fixed to a value not too far from 1. This means that, within associative memory, the cross-effect from being mentally triggered ("cued") by an observation about output to innovations for inflation expectations is not too different from the cross-effect where the two domains—output and inflation—are reversed. With this assumption, H and thus the series for the structural shocks ε_t is identified. It turns out that the value of H depends very little on τ for values between 0.75 and 1.[11]

Having identified ε_t, we estimate impulse responses by running local projections (Jordà 2005, Montiel Olea and Plagborg-Møller 2021).[12] We do so again for all purification schemes described above. The LHS variables are contemporaneous values and leads of GDP and inflation. Both are subject to the same purification schemes as MSC expectations. More precisely, we use purified versions of contemporaneous values and leads of cyclical real US GDP and leads

[11] It turns out that, for $\tau > 1$, there is no solution for H with h_{yy} and $h_{\pi\pi}$ both strictly positive and we hence disregard this range of τ.

[12] In the current context, we prefer local projections to more model-driven estimates since we intentionally want to "let the data speak", as far as possible.

of quarterly US CPI growth als LHS variables.[13] To explore the robustness of the estimations, we run two different specifications of local projections, given by

$$y_{t+h} = \beta_0^{(h)} + \beta_1^{(h)} \hat{\varepsilon}_t + \beta_2^{(h)} y_{t-1} + v_t^{(h)}, \tag{3}$$

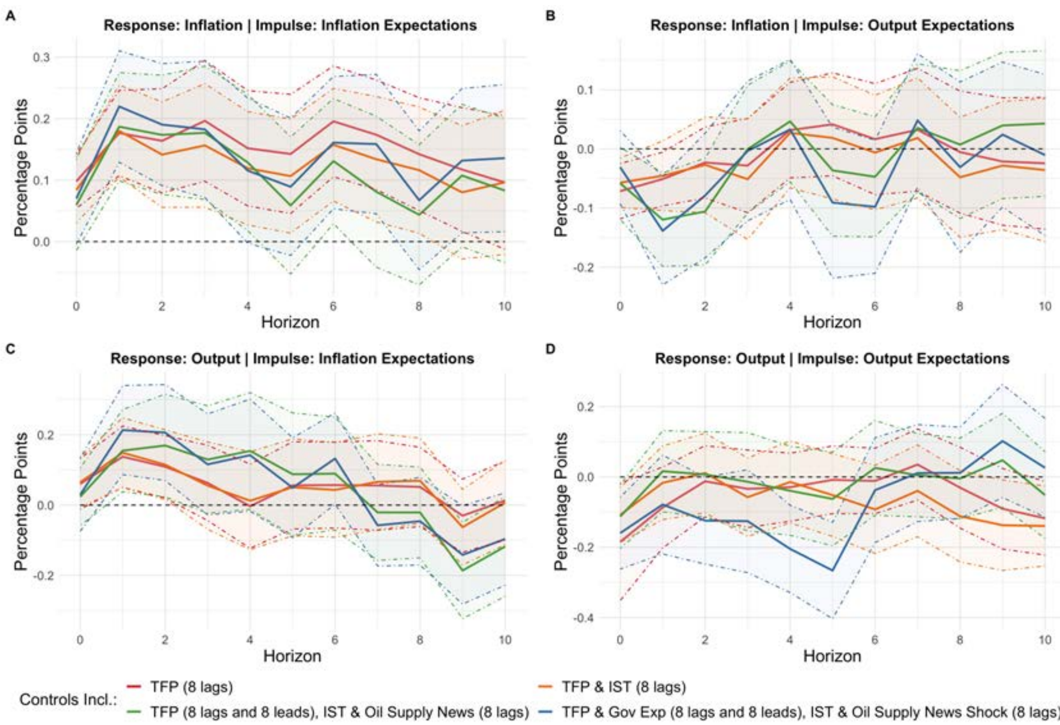
and

$$y_{t+h} - y_{t-1} = \beta_0^{(h)} + \sum_{k=0}^{K} \beta_{k+1}^{(h)} \hat{\varepsilon}_{t-k} + \beta_{K+1}^{(h)} (y_{t-1} - y_{t-2}) + v_t^{(h)} \tag{4}$$

where $h \geq 0$ indicates the lead horizon, y_{t+h} stands for the h-lead of a purified macroeconomic variable (output or inflation), and $\hat{\varepsilon}_t$ stands for the impulse variable, i.e. the estimated shocks to intuitive expectations. The first version derives from the analysis in Montiel Olea and Plagborg-Møller (2021), while the second follows the specifications used by Enders et al. (2021). In the latter case, we choose $K = 8$ lags for the expectation shocks.

[13] We use the series *GDPC1* and *CPIAUCSL* from the FRED database. The cyclical portion of the variation in real US GDP is estimated using Hodrick-Prescott filtering of the logarithm of *GDPC1* ($\lambda = 1600$). Alternative procedures, such as band-pass filtering at frequencies corresponding to 6 – 32 quarters (Baxter-King or Christiano-Fitzgerald) do not materially change the results.

Figure 4: Linear projection estimates of structural impulse responses for specification (3)



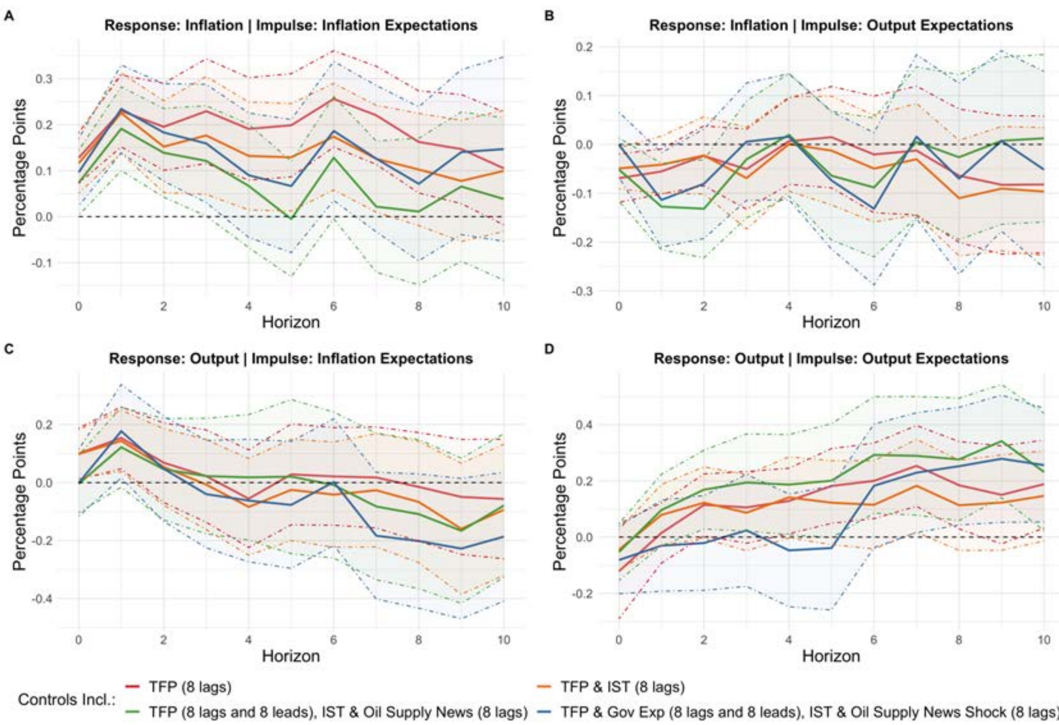
NOTES: The figure plots impulse responses for the linear projection specification (3). The confidence bands indicate 90% significance regions. The magnitude of the initial impulse is one standard deviation of the corresponding SVAR residual. As an example, a value of 0.1 on the vertical axis means that a one standard deviation shock in, say, (purified) inflation expectations is associated with a 10% standard deviation higher value of, say, (purified) inflation. The number of observations in the regression for the contemporaneous impact are 243, 200, 140, and 140, respectively.

Figure 4 plots the estimated impulse responses for the specification (3). What stands out is the pattern in Panel A: for all purification schemes, the estimate for the short-run impact of a shock to intuitive inflation expectations is significantly positive. Depending on the purification, the effects of a shock of one standard deviation of the expectation shock amounts to between 0.15 and 0.2 standard deviations of purified inflation. The patterns shown in the other panels are less clear-cut. Panel C shows a tendency for output first reacting positively to intuitive inflation expectation shocks. Output then reverts back to zero, with some specifications also showing more negative effects after 7 quarters. Panel B and D show the reaction of inflation and output to an intuitive output expectation shock. Panel B suggests that the reaction of inflation is first negative and then turns back to values close to zero. Panel D suggests that output is first negatively affected and then turns back to zero. This effect is somewhat counterintuitive; however, it is not very robust. Turning to Figure 5, we see that specification (4) yields mostly positive estimates for the effect of intuitive output expectation shocks on output. For some purifications, these are significant at a horizon between 6 and 10 quarters. For all other cases, the patterns in Figure 5 are very similar to Figure 4. In particular, the relatively strong effect of intuitive inflation expectations on inflation in Panel A are very similar across the two specifications.[14]

In sum, the estimated effects provide suggestive evidence that shocks to intuitive expectations may indeed affect actual macroeconomic outcomes. At this stage, however, it is somewhat difficult to make full sense of this observation; we therefore now proceed to a theoretical model of dual-system processing and a specific model of intuitive expectation formation based on associative memory.

[14]We also run many other specifications, using different structural VARs for the identification of structural shocks, different specifications for the linear projections, and different expectation measures from the MSC.

Figure 5: Linear projection estimates of structural impulse responses for specification (4)



NOTES: The figure plots impulse responses for the linear projection specification (4). The confidence bands indicate 90% significance regions. The magnitude of the initial impulse is one standard deviation of the corresponding SVAR residual. As an example, a value of 0.1 on the vertical axis means that a one standard deviation shock in, say, (purified) inflation expectations is associated with a 10% standard deviation higher value of, say, (purified) inflation. The number of observations in the regression for the contemporaneous impact are 235, 192, 132, and 132, respectively.

4 A Macroeconomy with Dual-System Thinkers

4.1 Starting point

In this section, we develop a general-equilibrium framework of a macroeconomy with dual-system thinkers. We aim for a framework that nests standard rational expectations as a special case. We limit the scope to (log)linearized representative agent models. A natural starting point is therefore the standard recursive Blanchard-Kahn equation (Blanchard and Kahn 1980). To simplify, we consider an economy with no predetermined endogenous variables, such as a capital stock. The Blanchard-Kahn equation then takes the form

$$x_t = AE_t x_{t+1} + Bu_t, \tag{5}$$

where x_t represents a vector of n endogenous non-predetermined variables, t denotes time, A is an $n \times n$ matrix and B is an $n \times 1$ matrix. In an application of our framework to a simple New Keynesian model in Section (5.1), x_t will include the two variables output and inflation. The term u_t denotes an exogenous stochastic state variable which, for now, we assume to be unidimensional to simplify the exposition. To fix ideas, think of it as total factor productivity (TFP). For tractability, we assume, that u_t follows the AR1 process

$$u_t = \rho u_{t-1} + \epsilon_t, \tag{6}$$

where $|\rho| < 1$, and ϵ_t represents a standard shock term that is *iid* with zero mean and finite variance.

In this paper we are concerned with expectations $E_t x_{t+1}$ in equation (5); in particular, we provide a model of how agents form such expectations when they have limited cognitive capacity. To keep a clear focus, we do not consider any other deviations from the standard model. We thus implicitly assume that the agent is perfectly aware of the primitive building blocks of the economy as well as the nature and parameterization of equilibrium relationships—as far as no expectation terms are involved. In a nutshell—in "as-if terms"—the agent is able to derive the Blanchard-Kahn form of equilibrium relationships in exactly the same way as a standard rational agent does.[15] It is only when it comes to the expectational recursion (5) that cognitive limitations materialize. This assumption is mainly driven by a methodological desire for focus. However, it can be justified by the fact that the generic form of macroeconomic relationships remain relatively stable over time, whereas expectation formation—or forecasting—crucially depends on an appropriate reading of signals, with nuances of their informational weight varying over time. As a consequence, the agent may have had better opportunities for learning about generic macroeconomic relationships than for assessing the informational content of signals.

In standard single-agent DSGE models with full-information rational expectations (FIRE),

[15] See Appendix A for a more detailed discussion. See Angeletos and Lian (2023) for a discussion of models of faulty intuitive reasoning about equilibrium relationships.

the agent forms model-consistent expectations. Thus, expectations $E_t x_{t+1}$ are determined recursively:

$$E_t x_{t+1} = A E_t x_{t+2} + B u_{t+1}. \tag{7}$$

Solving the recursion forward, and assuming that the respective sum converges, it is given by

$$\sum_{h=1}^{\infty} A^h B E_t u_{t+h} = \rho_u (I - \rho_u A)^{-1} B u_t \tag{8}$$

In this paper, we ask how the agent may form the corresponding expectations when FIRE exceed the agent's cognitive capacities. We pursue an approach that dominates in the psychological literature: dual system processing. A very interesting recent application in economics is provided by Ilut and Valchev (2023) who consider a dynamic consumption-savings choice under dual processing. Decades of research in psychology have established that when facing cognitively challenging tasks, humans combine two types of mental processes: intuitive thinking ("fast thinking" or "System 1"); and rigorous thinking ("slow thinking" or "System 2"; Kahneman 2011, Evans and Stanovich 2013). Intuitive thinking is the default; rigorous thinking is activated particularly when we become aware that our intuitive reasoning may lead to faulty outcomes (Ilut and Valchev 2023). However, even when rigorous thinking is active, the result may nevertheless be partly influenced by intuitive thinking. In the model of Ilut and Valchev (2023), this happens because intuitive thinking provides a "prior", upon which rigorous thinking may add an "update". Alternatively, a cognitive task may simply be too difficult to fully think through in a rigorous way and intuitive thinking is used for shortcuts, often without our awareness. Evidence by Bianchi et al. (2022) show that elements of intuitive thinking are even prevalent in professional forecasters, in the form of overweighing the predictive power of private information. Candia et al. (2020) show that subjective expectations held by households, and also by firms, are much less accurate than those by professional forecasters and show clear traces of intuitive thinking.

The stance that we take in this paper is that, in reality, forming expectations akin to (8) is cognitively challenging indeed, and possibly beyond capacity of real individuals. To fully appreciate this view, consider that the relevant domain of macroeconomic reality about which we want to learn something through the lens of a model features many more than a single exogenous state u_t. In reality, there are many candidate exogenous states and shocks, they may be hard to identify, and it may be hard to assess their impulse responses. While we do not want to model all the elements that contribute to the complexity of expectation formation, it is against this background that we view the assumption that expectation formation in macroeconomics is cognitively challenging. It is important to keep this background in mind even if equations like (5) or (8) do not look particularly complex to a trained economist.

4.2 The interplay of intuitive and rigorous thinking

The simplest model of the interplay between intuitive and rigorous thinking is one in which each system has a fixed influence on expectation formation; this is the approach we pursue here. Ilut and Valchev (2023) provide a fully microfounded model in which the relative influence of the two systems depends on history and familiarity of a task. In their case, the cognitively challenging task is to assess the optimal consumption level in the form of a policy function of an exogenous state. The latter is unidimensional and perfectly observable. In our case, we are interested in obtaining a parsimonious dual-system model that remains tractable in a general equilibrium environment. Moreover, we want to preserve scope for a microfoundation for the inner workings of intuitive thinking based on the concepts of similarity and context (Bordalo et al. 2020a, Bordalo et al. 2020a). We therefore see a model with fixed influence for each system well-suited to the aims of our analysis.

Assuming a fixed influence of each system, our specification of dual-system expectations, E_t^D, transforms equation (5) into

$$E_t^D x_{t+1} = \delta \left(A E_t^D x_{t+2} + B E_t u_{t+1} \right) + (1 - \delta)\, s_t, \tag{9}$$

with $0 \leq \delta \leq 1$. Here, the expression $A E_t^D x_{t+2} + B E_t u_{t+1}$ on the RHS is identical to the RHS of (7), except for the first expectations operator which now has a superscript D. This reflects the contribution of rigorous thinking to expectation formation. Importantly, (9) represents a recursive equation for $E_t^D x_{t+1}$, and so rigorous thinking remains recursive. However, only a share δ—the fixed share of rigorous thinking—finds its way into the recursion. Note that $E_t u_{t+1}$ represents a standard mathematical expectation that, given (6), is equal to $\rho_u u_t$. This is natural since, in the realm of rigorous thinking, expectations are formed in the standard rational way. A share $1 - \delta$ of expectation formation is contributed by intuitive thinking. Rather than thinking through the recursion, intuitive thinking directly comes with an answer for its part of expectation formation. The answer of intuitive thinking is "direct" in the sense that there is no analytical underpinning of how expectations of x_{t+1} depend on fundamental shocks. Rather, intuitive thinking "feels" an answer without engaging in any analytical reasoning. We refer to the term s_t that represents intuitive thinking in (9) as *sentiments*. Sentiments share their dimensionality with x_t, i.e. $s_t \in \mathbb{R}^n$. Reading (9) as a recursion for $E_t^D x_{t+1}$, sentiments can be understood as a perturbation of rigorous thinking. We make the following assumptions about sentiments.

ASSUMPTION 1 *(i) Sentiments s_t are independent of $u_{t'}$ for all $t, t' \geq 0$. (ii) $E_t^D s_{t+h} = s_t$.*

The first part of the assumption states that we limit our interest to sentiments that are not triggered by any fundamental event in the economy. For instance, we exclude that sentiments represent news about future total factor productivity (Barsky and Sims 2012) or news about oil shocks (Känzig 2021). Linking this to the empirical analysis in Section 3, the empirical counterpart of sentiments are thus not survey measures about macroeconomic variables, but rather residuals of these survey measures from regressions on all relevant macroeconomic fundamentals such as total factor productivity, etc.

The meaning of the second part of the assumption is that the agent's rigorous thinking is not aware of the perturbations of intuitive thinking. Sentiments "quietly slip in" into the calculus of rigorous thinking. As a result of this unawareness, rigorous thinking is not in a position to form rational forecasts for future sentiments akin to $E_t u_{t+1}$, $E_t u_{t+2}$, etc., which it correctly estimates as $\rho_u u_t$, $\rho_u^2 u_t$ etc. Rather, since rigorous thinking is unaware of the presence of intuitive thinking, the former is not able to foresee the future impact of sentiments. The logic of this is simple: if rigorous thinking were aware of the perturbations of intuitive thinking and their intertemporal dynamics, it could correct for it and we would be back to a FIRE world. Rather, according to a psychological view, individuals suffer from the illusion that they do their very best to eliminate any influence of intuitive thinking. However, unaware to them, their assessments are still influenced by recent conversations, by shopping experiences (prices), or news. The weight of these influences are measured by $(1 - \delta)$. This unawareness also explains why professional forecasters are found to overweigh private information (Bianchi et al. 2022).

In our model, we treat δ as a "deep" parameter, although with a more complete microfoundation, it would depend on the difficulty of a particular expectation formation task, on historical circumstances, opportunities for social learning, and education, among other factors. An economy with a comparateively high δ—i.e. a strong influence of rigorous thinking—would be one in which the set of exogenous states and shocks affecting macroeconomic variables are well known, the respective relationships are stable, and individuals are well-trained in rational problem solving. Conversely, an economy would be characterized by a comparatively lower

δ if exogenous states are harder to identify, macroeconomic relationships are less stable, and individuals have a lower propensity to rational analytical thinking.

In the following proposition, we provide the solution to (5) when the expectation operator E_t is replaced by E_t^D, which in turn is defined by (9).

PROPOSITION 1 *Assume that* $\sum_{h=0}^{\infty}(\delta A)^h BE_t u_{t+h}$ *converges and the inverse matrix* $(I-\delta A)^{-1}$ *exist. Then the solution of (5) under dual-system expectations is given by*

$$x_t = x_t^r(\delta) + (1-\delta)Ax_t^i(s_t,\delta) \equiv x_t^D(s_t,\delta) \tag{10}$$

where

$$x_t^r \equiv \sum_{h=0}^{\infty}(\delta A)^h BE_t u_{t+h} = (I-\delta\rho_u A)^{-1}Bu_t \tag{11}$$

is the standard (rational-expectations) forward solution for the stochastic difference equation

$$x_t = \delta AE_t x_{t+1} + Bu_t \tag{12}$$

that reflects the contribution of rigorous thinking; and

$$x_t^i \equiv \sum_{h=0}^{\infty}(\delta A)^h s_t = (I-\delta A)^{-1}s_t \tag{13}$$

is the standard forward solution of the deterministic difference equation

$$x_t = \delta Ax_{t+1} + s_t, \tag{14}$$

with $s_{t+h} \equiv s_t$, *reflecting the contribution of intuitive thinking.*

Proposition 1 provides the formulas for calculating dual-system expectations that can be used for deriving equilibria and impulse response functions for any DSGE economy with no predetermined endogenous state variables.[16] From an economic point of view, we are particularly interested in how dual system expectations modify the impulse responses to shocks in exogenous states. They are characterized by the following corollary.

Corollary 1 *Impulse responses of* x_{t+h} *to a shock* ϵ_t *at a horizon* $h \geq 1$ *are equal to the FIRE impulse responses times* δ^h. *Thus, compared to a FIRE economy, impulse responses are*

[16] Our model can be adapted to the case of predetermined exogenous states without losing tractability.

dampened by δ^h*.*

Recall that the shock ϵ_t appears in equation (6); as an example, think of a TFP shock. The result of Corollary 1 follows from the fact that only a fraction δ of the expectation formation process enters the recursion. As a result, impulse responses for ϵ_t at a horizon h are dampened by a factor δ^h. This dampening can also be understood as a form of "attenuation bias". Using an analogy from regression analysis, rigorous thinking attempts to solve an attribution problem by assessing how much of the variation in x_t is attributable to exogenous states at different horizons. However, rigorous thinking only governs a fraction δ of the expectation process. Intuitive thinking interferes into this attribution problem by adding sentiments, which are not attributed to any fundamental shocks. From the view of rational thinking, intuitive thinking therefore perturbs the solution of the attribution problem by adding noise that, in a FIRE world, would be irrelevant for expectation formation. This noise leads to an "attenuation bias" that is reflected in the dampening factors δ^h. As suggested by the local projection estimates in the previous section, and as we will pursue further in Section 5.1, information about sentiments is not irrelevant to a rational external observer who has data about sentiments. The reason is that sentiments may lead to self-fulfilling prophecies.

As a result of the discussed dampening, our model shares an important feature with the New Keynesian model with cognitive discounting of Gabaix (2020). He defines this concept as: "an innovation happening in k periods has a direct impact on agents' expectations that is shrunk by a factor $\bar{m}^k$ relative to the rational response, where $\bar{m} \in [0, 1]$ is a parameter capturing cognitive discounting." In our model, the parameter δ leads to exactly this effect. While in Gabaix' model, $E_t x_t$ experiences a shrinkage towards the steady state, our model posits a shrinkage towards the current realization of the sentiment s_t. Since sentiments vary over time (see Subsection 4.3), this generates a new source of variation in $E_t x_{t+1}$ – and hence x_t – that is not linked to the fundamentals of the economy (i.e. u_t). We therefore refer to the source of variation induced by sentiments as *autonomous*. As such, our sentiments are reminiscent to confidence shocks in Angeletos et al. (2018b).

A natural property of expectations formed in dual-processing mode would be that the influence of intuitive thinking does not decrease for terms associated with a more distant future. Under dual-system expectations as defined above, the ratios of the shares of the contributions of rigorous and intuitive thinking are, in fact, constant. To see this, let us expand (9), with terms arranged as follows:

$$
\begin{aligned}
E_t^D x_{t+1} = & \delta B \left(E_t u_{t+1}\right) && +(1-\delta)s_t + \\
& \delta \left(\delta A B E_t u_{t+2}\right) && +(1-\delta)\delta A s_t + \\
& \delta \left(\delta^2 A^2 B E_t u_{t+3}\right) && +(1-\delta)\delta^2 A^2 s_t + \\
& \ldots + && \\
& \delta \left(\delta^{h-1} A^{h-1} B E_t u_{t+h}\right) && +(1-\delta)\delta^{h-1} A^{h-1} s_t + \\
& \ldots && ,
\end{aligned}
\tag{15}
$$

We can interpret the first term in each line as the contribution of rigorous thinking, while the second term represents the contribution of intuitive thinking for a particular time horizon $h \geq 1$. These contributions add up to the overall dual-system expectation for x_{t+1}. The contributions of rigorous thinking take the form $A^h B E_t u_{t+h}$, multiplied by a weight δ^h; the contributions of intuitive thinking take the form $A^h s_t$, multiplied by a weight $(1-\delta)\delta^h$. For each horizon h, the ratio of the weights for intuitive over rigorous thinking is $(1-\delta)/\delta$, which is constant.

4.3 The inner workings of intuitive thinking

Associative Memory

After the discussion of the relative influence of intuitive and rigorous thinking on expectation formation, we now turn to the inner workings of intuitive thinking. So far, the sentiments s_t have just been an abstraction. We now provide a microfoundation based on an adaptation of the models of associative memory in Bordalo et al. (2020a) and Bordalo et al. (2023).[17] This microfoundation of intuitive thinking is a self-contained module. Readers who are not interested in a deeper understanding of the properties of s_t can proceed directly to the New Keynesian Economy with dual-systems expectations in Section 5. However, the associative-memory model of intuitive thinking provides crucial insights into the correlations between the different components in s_t. As a result, the model can explain why intuitive expectations may differ from statistical estimations based on historical data. Furthermore, it sheds light on why model-consistent expectations may be in conflict with the cognitive mechanics of human

[17] See also Enke et al. (2020) and Andre et al. (2022) for a discussion of associative memory. The latter study provides direct empirical evidence for the prevalence of associative memory in macroeconomic assessments.

memory.

In this subsection, we present a generic model of associative memory. In the next subsection, we link associative memory back to our evidence in Section 3. In particular, we discuss how associative memory can give rise to a negative correlation of output growth and inflation in survey expectation data. The model will also shed light on the link between VAR innovations and shocks in a structural VAR.

Models of associative memory are commonly presented as characterizing probability or expectation formation of a decision maker as a whole. In our case, associative memory is linked (only) to an agent's intuitive thinking—which reins with a weight of $(1 - \delta)$. To keep the overall perspective of our dual-system model, we use the shortcut IDM as a mnemonic for the intuitive decision maker in our model. Following Bordalo et al. (2020a) and Bordalo et al. (2023), we assume that the IDM holds in memory a mental database of variables associated with macroeconomic events. It may help to imagine this database as a large "mental spreadsheet". Each row in the spreadsheet corresponds to an experience e that the IDM has made on their own or learned through other sources. The (metaphoric) spreadsheet features two categories of columns. The first category corresponds to "hard" macroeconomic variables pertaining to equation (5). Following Bordalo et al. (2020a), we denote the second and "soft" type of variables as context variables. In a FIRE world, the latter variables would not contain any information relevant to forming expectations. However, from a psychological point of view, these variables may be salient. They may record any psychologically relevant features that the IDM associates with macroeconomic events. Most notably, they may record the *valence* of an event, i.e. the degree to which it feels "good", "bad", or "neutral" from an emotional point of view. For instance, the period of high inflation during the Volcker era may exhibit a value of "very bad" for a variable with label "emotional value associated with inflation". Other context variables may relate to timing (e.g. recent vs. more distant), or whether events were associated with vivid conversations with peers etc. Importantly, context variables may be functions of the "hard" macroeconomic variables (such as would be the case for valence), or they may be independent, e.g. the identity of friends with whom one was talking about an event.

In a nutshell, the IDM forms expectations based on associative memory according to the following procedure. The IDM makes current observations that bear a (psychological) relationship to different variables in equation (5) and the mental database. In the terminology of cognitive psychology, these observations are called "cues". The IDM uses the relationship between the cues and corresponding variables in the database to make an intuitive associative estimate about $E_t x_{t+1}$. This estimate then becomes the sentiment s_t introduced in the previous subsection. Associative memory works such that the IDM draws a sample from experiences ("rows") in the database. This sample is drawn probabilistically. The probability of any experience being drawn into the sample depends on two factors. First, the probability is proportional to a measure of *similarity* between the cue and an experience in the data base (or rather a subset of it).

Second, the probability is proportional to the similarity within a group of experiences that share certain properties. More specifically, in our model, the probability of an experience in the database being part of the mental sample decreases with the entropy of the respective group. Entropy is a measure of "disorder" of members in a group, in particular also of experiences as entries (or "data points", "rows") in a (mental) dataset. Entropy is low when a large fraction of entries within a group are very similar to each group, i.e. when they share common values among those variables that particularly matter.[18] Entropy has also been used as a descriptive model for the mechanics of associative memory in neuroscience (Pineda et al. 2021, Pineda and Morales 2023). In the same vein as in Bordalo et al. (2023), the sampling or retrieval probability of an experience is negatively related to the entropy of the respective group. Due to this mechanism, experiences from small groups with very low entropy get (possibly heavily) oversampled. Put simply, low entropy makes them salient and "stick" in memory. They are conducive to stereotypes and may therefore play an outsized role in shaping expectations or forecasts.

As an example, consider that an IDM is cued by unexpectedly high grocery or gasoline prices, and friends were talking about postponing the purchase of a new home, due to economic difficulties. These observations relate to measures of inflation and output in the mental database. Moreover, they both have a negative individual valence, and they also form jointly a "package" of negative overall valence ("bad"). Suppose that the mental database contains the variables inflation, output, the valences of inflation and output, and the overall valence of the experience. The cue then shows high similarity to all experiences associated with "bad economic circumstances". Everything else equal, experiences relating to "bad economic circumstances" may have a high probability of being sampled in a reaction to the cue. However, in associative memory, there is a second mechanism at work. There may or may not be many

[18] Bordalo et al. (2023) use a measure called self-similarity which, in our applications, would lead to similar conclusions. We use entropy since it is very tractable in the domain of our macroeconomic applications. Entropy is also used as a measure of information gain in the literature on rational inattention (Maćkowiak et al. 2023).

other experiences in the database relating to"bad economic circumstances"—or also just to "bad". They together form a group of "bad" experiences. To the degree that this class is very dissimilar (has a high entropy), the sampling probability of experiences in this group becomes lower. By contrast, assume now that the class of experiences with a valence of "very bad" (rather than just "bad") has a very low degree of dissimilarity (entropy). Then the sampling probability of an experience from the group of "very bad" experiences increases relative to the group of "bad experiences", although the experiences from the "very bad" group may be less similar to the current cue than "bad" experiences. In Subsection 4.4, we show in detail how a model along the lines of this example can explain the negative reduced-form Phillips curve that is prevalent in the Michigan Survey of Consumers, as discussed in the Section 3.

A Formal Model of Associative Memory

We now formalize the mechanisms outlined above. The IDM's mental database consists of a matrix ("spreadsheet") $\boldsymbol{E}$. The rows of this matrix are referred to as experiences and denoted by a vector e. The columns are referred to as variables. Each experience contains M_x macroeconomic variables that take on values from a set $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_{M_x}$. Furthermore, e contains M_c context variables taking on values from a set $\mathcal{C} = \mathcal{C}_1 \times \mathcal{C}_2 \times \ldots \times \mathcal{C}_{M_c}$. Overall, we have $e \in \mathcal{X} \times \mathcal{C} \equiv \mathcal{E}$. All sets of component values $\mathcal{X}_k$ and $\mathcal{C}_l$ are finite and either numerical or categorical. For variables where the underlying macroeconomic measure is a continuum (such as output or inflation), the corresponding set contains discretized values corresponding to value ranges.

We now turn to similarity between experiences in $\boldsymbol{E}$. The dimensions M_x or M_c are potentially quite large as the IDM may ultimately remember many details. However, a cue may have (strongly) reduced dimensionality compared to the columns of $\boldsymbol{E}$, or dimensions of $\mathcal{E}$ (e.g. the cue may just consist of grocery or gasoline prices and an associated valence). Hence, only a few columns in $\boldsymbol{E}$ (possibly made salient by the cue) may matter for perceived similarity to a cue. Denote the cue by $\varkappa$. The cue leads to the selection of a (small) subset of columns from $\boldsymbol{E}$ that forms the basis for perceived similarity. We denote the reduced matrix of selected columns by $\bar{\boldsymbol{E}}$, with typical row $\bar{e} \in \bar{\mathcal{E}}$, where the latter set results from a projection $\Pi_\varkappa$ onto the product of sets $\mathcal{X}_k$ and $\mathcal{C}_l$ that contain the possible values of the columns in $\bar{\boldsymbol{E}}$. Abusing the concept of projection, we also write $\bar{\boldsymbol{E}} = \Pi_\varkappa(\boldsymbol{E})$. The subscript $\varkappa$ indicates, that the column selection or projection may depend on the cue. We allow for the case that either all columns of hard macro variables or all columns of soft context variables completely disappear, but not both.

As a simple example, $\boldsymbol{E}$ may contain all macroeconomic variables, and their "soft associates", that an individual can potentially think of. By contrast, $\bar{\boldsymbol{E}}$ may only contain the valence of inflation as a single variable.

We assume that similarity as perceived by the IDM can be described as a function of a distance metric between two elements $\bar{e}, \bar{e}' \in \bar{\mathcal{E}}$. (We thus also assume that $\bar{\mathcal{E}}$ is a metric space.) Let $\Delta : \bar{\mathcal{E}} \times \bar{\mathcal{E}} \to \mathbb{R}_+$ be any distance metric on $\bar{\mathcal{E}}$. We then define similarity between a pair of experiences $\bar{e}$ and $\bar{e}'$ as a function S of their mutual distance $\Delta(\bar{e}, \bar{e}')$, i.e. $\mathsf{S} : \mathbb{R}_+ \to \mathbb{R}_+$ with the following properties: *(i)* $\mathsf{S}(0) = 1$ (i.e. the case $\bar{e} = \bar{e}'$ and hence $\Delta(\bar{e}, \bar{e}') = 0$), intuitively, similarity amounts to "100 percent" in this case; *(ii)* $\lim_{\Delta(\bar{e},\bar{e}') \to \infty} \mathsf{S}(\Delta(\bar{e}, \bar{e}')) = 0$; and *(iii)* $\Delta(\bar{e}, \bar{e}')$ (weakly) decreases with an increasing $\Delta(\bar{e}, \bar{e}')$.

Next, we formalize the idea that an experience e belongs to a certain group of similar experiences. Since this group takes on the form of a multiset, which is also referred to as a "bag", we denote the respective group a "similarity bag" of an experience $\bar{e} \in \bar{\boldsymbol{E}}$. It links the experiences $\bar{e}$ from the reduced experience data $\bar{\boldsymbol{E}}$ back to the full experience data $\boldsymbol{E}$ that includes everything the IDM can potentially think of. Let $\Pi_{\varkappa}^{-1}$ denote the inverse projection expanding a reduced experience $\bar{e}$ in the reduced dataset $\bar{\boldsymbol{E}}$ back to the full experience e in the full dataset $\boldsymbol{E}$. The similarity bag for a reduced experience $\bar{e}$ that appears as an entry in $\bar{\boldsymbol{E}}$ is then defined as follows.

Definition 1 (Similarity bag) *A similarity bag $\boldsymbol{S}_{\varkappa}(\bar{e})$ consists of all experiences $e \in \boldsymbol{E}$ for which $\Pi_{\varkappa}(e) = \bar{e}$.*

Thus, a similarity bag consists of those entries in $\boldsymbol{E}$ for which the columns that determine similarity take on a particular value $\bar{e}$. This particular value may appear in multiple rows, which is why the similarity bag is a data matrix or database, not a set in a set-theoretic sense. As an example, suppose that $\bar{e}$ contains only the valence of inflation. Hence, similarity is only perceived along this dimension. Suppose that $\bar{e}$ contains the value of "neutral". Then the similarity bag $\boldsymbol{S}_{\varkappa}(\bar{e})$ consists of all entries in $\boldsymbol{E}$ for which inflation valence is "neutral". The similarity bags formally capture the idea of a group of experiences, the (dis)similarity of which is measured by entropy. Denote by $p_{\boldsymbol{S}_{\varkappa}}(e)$, the relative frequency of e in $\boldsymbol{S}_{\varkappa}(\bar{e}))$. Denote by $\mathcal{S}_{\varkappa}(\bar{e}))$ the set of all possible values that the entries ("points") in $\boldsymbol{S}_{\varkappa}(\bar{e}))$ may take on. (Recall that a possible value from the (mathematical) set $\mathcal{S}_{\varkappa}(\bar{e}))$ may appear multiple times in the dataset $\boldsymbol{S}_{\varkappa}(\bar{e})$).) The entropy of $S_{\varkappa}(\bar{e})$ is then defined as

$$H\left(\boldsymbol{S}_{\varkappa}(\bar{e})\right)=-\sum_{e\in\mathcal{S}(\bar{e})} p_{\boldsymbol{S}_{\varkappa}}(e)\log p_{\boldsymbol{S}_{\varkappa}}(e). \tag{16}$$

Entropy takes on a value zero if all elements in the similarity bag take on the same value (since $\lim_{p\to 0} p\log p=0$), and it takes on its maximum value if all possible values are realized with equal probability (i.e. a uniform empirical frequency over possible values of e). The maximum value increases with the size of similarity bag. We only consider cases where $\boldsymbol{S}(\bar{e})$ is non-empty, such that entropy is always well-defined. Intuitively, when entropy of a similarity bag is zero, this bag lends itself for perfect stereotyping and is therefore very "memorable". A similarity bag with high entropy has no stereotype and therefore "pales" in memory.

Entropy refers to experiences in the full set $\boldsymbol{E}$, rather than the reduced set $\bar{\boldsymbol{E}}$ pertaining to the cue. The logic of this is the following. The IDM's associative thinking leads to a quick reaction of what is perceived similar to a cue. For this, only very few dimensions, or even a single one, may be used (e.g. the valence of inflation). This happens in the domain of $\bar{\boldsymbol{E}}$. However, the IDM may then have to make an (intuitive) assessment about a variable that was not part of $\bar{\boldsymbol{E}}$ but only of $\boldsymbol{E}$, e.g. inflation or business conditions (as is the case, e.g., for participants in the Michigan Survey of Consumers). For this, the IDM needs to expand the reduced experience set $\bar{\boldsymbol{E}}$ to the full experience set $\boldsymbol{E}$. So the IDM mentally draws a sample from the full set $\boldsymbol{E}$.

We are left with the question of what comes to the associative mind most easily, or rather most frequently. Before characterizing mental sampling, however, we need to discuss cues in somewhat more detail. Similar to a (reduced) experience $\bar{e}$, a cue $\varkappa$ consists of cues relating to hard macroeconomic variables $\varkappa^x$, and soft context variables $\varkappa^c$; thus $\varkappa=(\varkappa^x,\varkappa^c)\in\bar{\mathcal{X}}\times\bar{\mathcal{C}}$. Cues may not represent the same variables as those in $\bar{\boldsymbol{E}}$, but there is a one-to-one association between the cue elements and variables in $\bar{\boldsymbol{E}}$. Moreover, we assume that the possible values of a cue are the same as for $\bar{e}\in\bar{\mathcal{E}}$. As an example, any cues related to pricing, such as grocery or gazoline prices, may be subsumed in a cue with the label "price cue"; they are then associated with the inflation variable in $\boldsymbol{E}$. Moreover, a valence of the price cue of, say, "bad" is associated with valence of inflation of "bad" in $\bar{\boldsymbol{E}}$ etc. While we could formalize the corresponding associations more rigorously, the cost of more formalism does not justify the added clarity. It should become clear in the Phillips curve example in Section 4.4 below that the nature of the associations leads to little danger of confusion. Since a cue $\varkappa$ takes on the same values as experiences $\bar{e} \in \bar{\mathcal{E}}$, it is also straightforward to calculate the similarity between any $\varkappa$ and $\bar{e}$.

We are now in a position to state the sample or *retrieval* probability r for an experience e contained in the memory database $\boldsymbol{E}$ when the IDM associatively samples from memory upon a cue $\varkappa$. This is given by

$$r(e,\varkappa) = \frac{\mathsf{S}(\Pi_\varkappa(e),\varkappa)\; e^{-\alpha H(\boldsymbol{S}_\varkappa(\Pi_\varkappa(e)))} p_{\boldsymbol{E}}(e)}{\sum_{e\in\mathcal{E}} \left[\mathsf{S}(\Pi_\varkappa(e),\varkappa)\; e^{-\alpha H(\boldsymbol{S}_\varkappa(\Pi_\varkappa(e)))} p_{\boldsymbol{E}}(e)\right]}, \tag{17}$$

where $p_{\boldsymbol{E}}(e)$ refers to the relative frequency of of $e \in \boldsymbol{E}$. While this expression may look complicated, we show in the next subsection that it is comparatively tractable. The parameter α governs how strongly entropy of the similarity bag decreases the retrieval rate from that bag since it is not very "memorable" for the lack of a suitable stereotype. Everything else equal, the retrieval rate of an experience e increases with the similarity between the cue and the experience. However, it decreases with the entropy the associated similarity bag. Finally, the retrieval rate also increases with the empirical frequency $p_{\boldsymbol{E}}(e)$, the factors similarity and entropy may easily trump the empirical frequency. In sum, our model is consistent with important patterns discussed in Bordalo et al. (2023).[19]

To form expectations, we assume that the IDM draws a single experience e from the mental database $\mathcal{E}$ with sampling probability $r(e,\varkappa)$. It would be straightforward to allow for a larger sample. However, we find that important insights can be gained even in this simple and tractable setting. Denote the retrieved experience as $e^* \in \mathcal{E}$. When the IDM forms an expectation about a macro variable X_k, the result is simply the kth element of e^*, which we denote by e^*_k.

Associative Memory in an Intertemporal Setting

As a final step, we embed intuitive expectation formation into an intertemporal setting where a cue is drawn every period t. We make the link back to sentiments s_t from the previous subsection. In a dynamic setting, it is natural that if the IDM has formed an (intuitive) expectation about a certain variable recently, the previous expectation may still influence

[19] Many of the details of our model defer from the model in Bordalo et al. (2023). In particular, they do not use entropy as a measure of group similarity and there is no counterpart to the parameter α that governs the sensitivity to entropy, as most of their analysis is generic. This parameter plays an important role in the Phillips curve application in Section 4.4

the current one to some degree. Furthermore, there may also be influences across variables, e.g. yesterday's (intuitive) inflation expectation may affect today's output expectation. The simplest way to capture these intertemporal dependencies is to embed the model of associative memory into a vector autoregression (VAR):

$$s_t = \Phi s_{t-1} + \eta_t, \tag{18}$$

where $s_t, \eta_t \in \mathbb{R}^n$ and Φ is a coefficient matrix of dimension $n \times n$. Use the projection Π^* to denote the selection of macroeconomic variables from an experience e that the agent forms expectations about. With this notation, we can link the VAR innovations η_t to our model of associative memory according to

$$\eta_t = \Pi^*(e_t^*) \tag{19}$$

where we have now added a time index to e_t^* such that it refers to the retrieved sample based on the current cue $\varkappa_t$. Concerning (18), we assume that all eigenvalues of Φ lie strictly inside the unit circle. This means that earlier expectations eventually lose their influence on newly formed intuitive expectations.

It is noteworthy that our model of associative memory does not imply that the elements of the innovations η_t are mutually uncorrelated. To the contrary, the cue $\varkappa$ may have led to correlated retrievals from the mental database. Since this result is of interest with respect to several macroeconomic relationships, and with respect to surveys about macroeconomic variables, we state it in the form of a proposition.

PROPOSITION 2 *Suppose that the components of $\varkappa^x$ are mutually independent. As a result of associative memory, the components of $\eta_t = \Pi^*(e^*)$ may be correlated. Similarly, if two components in $\varkappa^x$ are mutually correlated, the mutual correlation of the corresponding components in η_t may have the opposite sign. Furthermore, the correlation of two components in η_t, s_t, and also $E^D x_{t+h}$, $h \geq 1$ may have the opposite sign of the factual empirical correlation in the data.*

To save space, we will not provide a general proof, but rather show a "proof by example" in the following subsection. The proposition shows that individuals' expectations may become correlated, or change the sign of a correlation in comparison to the factual empirical distribution thanks to associative memory. Compared to the FIRE case, this may lead to a type of (partially) self-fulfilling prophecies.

From an econometric point of view, it is of interest whether it is meaningful to transform the above VAR into a structural VAR with uncorrelated structural disturbances ε_t, such that $\eta_t = H\varepsilon_t$ for a suitable matrix H. We may interpret these structural shocks as *belief shocks* associated with associative memory. In light of our model, identifying these structural shocks is meaningful if the elements $\varkappa^x$ of the underlying cue are independent. In several applications, this may be a relevant case. In other applications, it may be less natural and the result may be an artificial and somewhat arbitrary orthogonalization of $\varkappa^x$ that may not have a clear interpretation.

4.4 Associative memory and a negative Phillips Curve

In this subsection, we show how a very simple application of the model of the previous subsection can explain how associative memory leads to a reduced-form Phillips curve, in agents' expectations, with a negative slope. This holds independent of the (empirical) distribution of supply and demand shocks.

We first discuss the working of the model in an informal way. The IDM holds a mental database of past output and inflation outcomes. We assume that this database is unbiased. We do so since we want to show that associative memory leads to distortions even if the underlying database exhibits no biases. The general model requires that values in the database are discretized. We consider a very simple case where both inflation and output can only take on the three values "low", "middle", and "high". Output and inflation represent the two only hard macroeconomic variables in the database. The database further contains three soft context variables: the valence, or emotional value, of output; the valence of inflation; and an overall valence of the experience. We assume that the individual valences of output and inflation take on the three values "good", "neutral", and "bad". The overall valence is expanded by the two values "very good" and "very bad", with the idea that two times "good" adds to "very good", while two times "bad" adds to "very bad". We provide a complete list of the respective relationships below.

The IDM obtains a price cue from, e.g., shopping or paying at the gasoline station, and a cue from the output domain, e.g. from talking to friends or reading news. The cue also comes with an overall valence of the situation, which is a function of the price and output cue. The IDM's perception of similarity is exclusively based on the overall valence. A cue with an overall valence of "very good" is perfectly similar to itself. It is somewhat similar to "good", and less but still somewhat similar to "neutral". Finally, it is not at all similar to "bad" and "very bad". A cue with an overall valence of "neutral" is somewhat similar to "good" or "bad", and it exhibits also some (lower) degree of similarity to "very good" and "very bad". Hence, everything else equal, based on a cue of "very good", the IDM would mainly retrieve a sample of "very good" or then also some "good" combinations of output and inflation. Triggered by a cue of "neutral", the IDM would retrieve mainly "neutral" combinations of output and inflation, but also some "good" and "bad" combinations, and even some "very good" and "very bad" ones.

As will become more clear in the formal discussion of the model, output and inflation combinations with an overall valence of "neutral" are substantially more diverse among each other than output and inflation combinations with a valence of "very good" or "very bad". Formally, the former have a higher entropy or lower self-similarity than the latter. Therefore, they do not come easily to mind and lack a stereotype. By contrast, "very good" or "very bad" combinations of output almost exclusively consist of stereotypes and are therefore highly memorable. Thus, they are over-represented in a mental sample retrieved from the memory database. Whenever the effect that "stereotypes stick" is sufficiently strong, the IDM's output and inflation expectations are negatively correlated since the sample will sufficiently frequently contain combinations of high output and low inflation, or low output and high inflation.

Due to mechanics of associative memory, there endogenously emerges an "all good/bad in one" heuristic. Groups of experiences where good things occur together have a low entropy (members of the group are very similar to each other). The same holds for experiences where bad things occur together. In multidimensional settings, there are obviously more outcomes where the valences of single dimensions are not aligned, such that the overall experience feels "neutral" (or also "unclear", "uncertain", "hard to classify"). As a result, high entropy or low within-group similarity emerges and retrieval rates from groups with mixed valences ("good" and "bad" combined) are low.

We now show how the result of a reduced-form Phillips curve in (intuitive) expectations can be obtained using the formalism of the previous subsection. Denote output (growth) by y, and inflation by π. The possible values of output are $\mathcal{X}_y = \{y^l, y^m, y^h\}$ and the possible values of inflation are $\mathcal{X}_\pi = \{\pi^l, \pi^m, \pi^h\}$. The valences of output and inflation both take on values of "good", "neutral", "bad". Formally, we represent this as $\mathcal{C}_y = \mathcal{C}_\pi = \{g, n, b\}$. Finally, the overall valence takes on values from "very good" to "very bad", i.e. $\mathcal{C}_o = \{vg, g, n, b, vb\}$. The memory database $\boldsymbol{E}$ contains the collection of variables, or columns, $(Y, \Pi, V_y, V_\pi, V_o)$, where V_y, V_π and V_o represent the valences of output, inflation, and the overall valence, respectively.

These variables take on values $e = (y, \pi, v_y, v_\pi, v_o) \in \mathcal{X}_y \times \mathcal{X}_\pi \times \mathcal{C}_y \times \mathcal{C}_\pi \times \mathcal{C}_o \equiv \mathcal{E}$.

We assume that the IDM perceives output levels of y^l, y^m, y^h as "bad", "neutral", and "good" respectively. The same holds for the inflation levels π^l, π^m, π^h (with orders reversed in comparison to output). The table below shows the relationship between outcomes of output, inflation, and overall valence. Note that two times "good" adds to "very good", and two times "bad" adds to "very bad", while "neutral" corresponds to a zero element in this addition scheme. We find this scheme for overall valence natural. However, the insights developed below

v_o	$\pi = \pi^l$	$\pi = \pi^m$	$\pi = \pi^h$
$y = y^l$	n	b	vb
$y = y^m$	g	n	b
$y = y^h$	vg	g	n

Table 1: Overall valence as a function of output and inflation outcomes

do not depend on the details of the specification. A valence of "neutral" captures outcomes that are hard to evaluate either because they have conflicting individual valences, or because they are "nothing special". The neutral category can thus also be characterized as "hard to evaluate", or "do not know what I should make of it", etc. What is essential for our results below is that uncertain valence is assigned to a larger set of diverse outcomes (with high entropy), while more distinctive valences as assigned to outcomes with homogeneous outcomes (with low entropy) that lend themselves to stereotypes.

The IDM perceives similarity only along the V_o dimension, i.e. overall valence, with values $v_o \in \{vg, g, n, b, vb\}$. The reduced memory database $\bar{\boldsymbol{E}}$ thus consists of a single column for V_0. A cue $\varkappa$ consists of an observation related to the output and price domain—the "hard" components—which are augmented by derived "soft" components of the valences of the output and inflation observation, and overall valence, respectively. The reduction of the observation and memory database is formally given by the projection $\Pi_\varkappa$ with $\Pi_\varkappa(\varkappa) = \varkappa_o$ and $\Pi_\varkappa(\boldsymbol{E}) = V_o$ (where we use the same variable names for cues than for experiences in $\boldsymbol{E}$). In this application, the projection does, in fact, not depend on $\varkappa$; similarity is always evaluated along V_o.

The distance metric for two entries in $\bar{\boldsymbol{E}}$ is given in Table 2. It results from assigning adjacent natural numbers to the values in $\mathcal{C}_o$ and using the absolute value of their difference as a metric. Similarity is shown in Table 3, where $\mathsf{S}(0) = 1 > \mathsf{S}(1) > \mathsf{S}(2) > 0 = \mathsf{S}(3) = \mathsf{S}(4)$. Thus, vg is similar to g and somewhat similar to n, but not similar to b and vb. We do not specify the values of $\mathsf{S}(1)$ and $\mathsf{S}(2)$ since our main result does not depend on their particular values.

	vg	g	n	b	vb
vg	0	1	2	3	4
g	1	0	1	2	3
n	2	1	0	1	2
b	3	2	1	0	1
vb	4	3	2	1	0

Table 2: Distance metric

	vg	g	n	b	vb
vg	1	S(1)	S(2)	0	0
g	S(1)	1	S(1)	S(2)	0
n	S(2)	S(1)	1	S(1)	S(2)
b	0	S(2)	S(1)	1	S(1)
vb	0	0	S(2)	S(1)	1

Table 3: Similarity values

For characterizing similarity bags, it is useful to define the sets of possible values for output and inflation in each bag. These are as follows:

$$\begin{aligned}
\mathcal{S}(vg) &= \{(y^h, \pi^l)\} \\
\mathcal{S}(g) &= \{(y^h, \pi^m), (y^m, \pi^l)\} \\
\mathcal{S}(n) &= \{(y^h, \pi^h), (y^m, \pi^m), (y^l, \pi^l)\} \\
\mathcal{S}(b) &= \{(y^m, \pi^h), (y^l, \pi^m)\} \\
\mathcal{S}(vb) &= \{(y^l, \pi^h)\}
\end{aligned} \tag{20}$$

As an example, consider now the similarity bag for vg. It is defined as the collection of all rows in $\boldsymbol{E}$ for which the pair (y, π) take on values in $\mathcal{S}(vg)$. In general, for any $v_0 \in \mathcal{C}_o$, the similarity bag $\boldsymbol{S}(v_o)$ is defined by

$$\boldsymbol{S}(v_o) = \{e \in \boldsymbol{E} : \Pi_{\varkappa}(e) \in \mathcal{S}(v_o)\} \tag{21}$$

The relative frequency of an outcome e in $\boldsymbol{S}(v_o)$ is fully determined by the outcomes for output and inflation (y, π) since the valence variables are deterministic functions of output and inflation. Denote the relative frequencies of a pair $(y, \pi) \in \mathcal{X}_y \times \mathcal{X}_\pi$ in the similarity bag by $p^{kl}_{\boldsymbol{S}(v_o)}$, with $k, l \in \{l, m, h\}$. With this, we can calculate the entropy of a similarity bag. As an example, consider the entropy values for $\boldsymbol{S}(vg)$ and $\boldsymbol{S}(n)$. In the first case, we have

$$\begin{aligned} H(\boldsymbol{S}(vg)) &= -\sum_{e \in \mathcal{S}(vg)} p^{jk}_{\boldsymbol{S}(vg)} \log p^{jk}_{\boldsymbol{S}(vg)} \\ &= -p^{hl}_{\boldsymbol{S}(vg)} \log p^{hl}_{\boldsymbol{S}(vg)} = -1 \times \log 1 = 1 \times \log 1 = 0. \end{aligned}$$

In the first line, the superscripts k and j refer to l, m, h. All elements in the similarity bag for "very good" are perfectly similar to each other, so the entropy is zero. By contrast, for the "neutral" case, we have

$$\begin{aligned} H(\boldsymbol{S}(n)) &= -\sum_{e \in \mathcal{S}(n)} p^{jk}_{\boldsymbol{S}(n)} \log p^{jk}_{\boldsymbol{S}(n)} \\ &= -p^{hh}_{\boldsymbol{S}(n)} \log p^{hh}_{\boldsymbol{S}(n)} - p^{mm}_{\boldsymbol{S}(n)} \log p^{mm}_{\boldsymbol{S}(n)} - p^{ll}_{\boldsymbol{S}(n)} \log p^{ll}_{\boldsymbol{S}(n)} > 0. \end{aligned}$$

For the case that the three value pairs in the neutral bag are uniformly distributed, we obtain $H(\boldsymbol{S}(n)) = 1.099$. For $\boldsymbol{S}(g)$ and $\boldsymbol{S}(b)$, entropy is also strictly positive; for a uniform distribution within those bags, we obtain an entropy value of 0.35.

We now have everything ready to state the retrieval rates. According to (17), the sum in the numerator for a retrieval rate is over all values $e \in \mathcal{E}$. In our case, these are all possible pairs of values for output and inflation (each taking on the values l, m, or h). In our application, we can simplify the sum by using the fact that overall valence V_o is a deterministic function of output and inflation according to Table 1. We thus can take sums over $\mathcal{C}_o$ rather than all value pairs. For this, we use the notation $p_{\boldsymbol{E}}(v_o)$ for the frequency of the value v_o in $\boldsymbol{E}$. The below expressions also make use of the fact that $\mathsf{S}(3) = \mathsf{S}(4) = 0$. With this, the retrieval rates are as follows:

$$r(e,vg) = \frac{\mathsf{S}(\Delta(e,vg))e^{-\alpha H(S(e))}p_E(e)}{p_E(vg) + \mathsf{S}(1)e^{-\alpha H(S(g))}p_E(g) + \mathsf{S}(2)e^{-\alpha H(S(n))}p_E(n)} \tag{22}$$

$$r(e,g) = \frac{\mathsf{S}(\Delta(e,g))e^{-\alpha H(S(e))}p_E(e)}{\mathsf{S}(1)e^{-\alpha H(S(vg))}p_E(vg) + e^{-\alpha H(S(g))}p_E(g) + \mathsf{S}(1)e^{-\alpha H(S(n))}p_E(n) + \mathsf{S}(2)e^{-\alpha H(S(b))}p_E(b)} \tag{23}$$

$$r(e,n) = \frac{\mathsf{S}(\Delta(e,n))e^{-\alpha H(S(e))}p_E(e)}{\mathsf{S}(2)e^{-\alpha H(S(vg))}p_{\boldsymbol{E}}(vg) + \mathsf{S}(1)e^{-\alpha H(S(g))}p_E(g) + e^{-\alpha H(S(n))}p_{\boldsymbol{E}}(n) + \mathsf{S}(1)e^{-\alpha H(S(b))}p_E(b) + \mathsf{S}(2)e^{-\alpha H(S(vb))}p_E(vb)} \tag{24}$$

$$r(e,b) = \frac{\mathsf{S}(\Delta(e,b))e^{-\alpha H(S(e))}p_E(e)}{\mathsf{S}(2)e^{-\alpha H(S(g))}p_E(g) + \mathsf{S}(1)e^{-\alpha H(S(n))}p_E(n) + e^{-\alpha H(S(b))}p_E(b) + \mathsf{S}(1)e^{-\alpha H(S(vb))}p_E(vb)} \tag{25}$$

$$r(e,vb) = \frac{\mathsf{S}(\Delta(e,vb))e^{-\alpha H(S(e))}p_E(e)}{\mathsf{S}(2)e^{-\alpha H(S(n))}p_E(n) + \mathsf{S}(1)e^{-\alpha H(S(b))}p_E(b) + p_E(vb)}. \tag{26}$$

Consider the retrieval rate for "very good", given that the cue is "very good", i.e. $r(vg, vg)$.

The numerator is then equal to $p_{\boldsymbol{E}}(vg)$, since similarity is one and entropy is zero. Whenever similarity decreases sufficiently quickly or α is sufficiently large, the denominator is also close $p_{\boldsymbol{E}}(vg)$. Thus, the retrieval rate becomes very close to one. Furthermore, we have $r(vg, vg) > r(g, vg) > r(n, vg)$ because of the larger distances of g and n to vg, which reduces similarity. Moreover, $r(b, vg) = r(vb, vg) = 0$ since similarity is zero in these cases. Clearly, the parameter α plays an important role for the relative magnitudes of the terms in the denominator, and for retrieval rates in general. It governs the sensitivity of "memorability" of a member of a similarity bag to its entropy (or dis-similarity). If a similarity bag is a rather "mixed bag", then it may still be memorable if α is rather low; but it will become less and less memorable with an increasing α. Somewhat more loosely speaking, the parameter α can also be interpreted as the "need for stereotyping" for making an experience memorable. More generally, a lower α is associated with higher cognitive skills.

Consider now the case where the cue is "neutral". Let us begin with $r(n, n)$. Whenever the relative frequencies in the similarity bag for neutral experiences come close to a uniform distribution, then $H(\boldsymbol{S}(e))$ in the numerator is relatively high for experiences e with $V_o = n$. Thus, the numerator and hence $r(n, n)$ become relatively low. Independent of the relative frequencies, let α continuously increase from zero towards infinity. When the sensitivity of associative memory to entropy increases, the expressions $e^{-\alpha H(\boldsymbol{S}(v_o))}$ continuously decrease towards zero for $v_0 = g, n, b$. The IDM's associative memory relies on the possibility for stereotyping, which relies on low entropy and a low sensitivity to entropy. When α increases, the case g, n, b become less and less memorable and retrievable. Eventually, only the perfectly "stereotypeable" cases vg and vb come to mind and win the retrieval race. The same arguments can be made with respect to $r(e, g)$ and $r(e, b)$.

In our model, expectations are formed based on a single retrieved sample, where the retrieval probability is $r(e, v_o)$. Denote the retrieved sample by e^*. The projection $\Pi^*(e^*)$ then selects the output and inflation component, which yield the IDM's expectations. With α sufficiently high, the selection consists almost exclusively of (y^h, π_l) or (y^l, π_h), with their relative frequency depending on relative frequencies of the two outcomes in $\boldsymbol{E}$.[20] Since y^h and π^h exceed their means, whereas y^l and π_l are lower than their means, respectively, we obtain a negative correlation between the two expectations. Denote by y_t^* and π_t^* the values retrieved when forming expectations in period t and use the notation $(y_t^*, \pi_t^*) = (\eta_t^y, \eta_t^\pi) \equiv \eta_t$. We then

[20] This result could be refined if we were to allow for the "emotional" weight of an experience in $\boldsymbol{E}$ to depend on its valence and, say, high inflation and low output have a particularly strong negative valence.

have the following result.

PROPOSITION 3 *There exists a critical threshold $\bar{\alpha}$ such that it $\alpha > \bar{\alpha}$, $Cov(\eta_t^y, \eta_t^\pi) < 0$.*

We refer to this negative correlation as *stagflationary expectations*. This result directly speaks to the reduced-form Phillips curve in VAR residuals shown in Figure 3. The associative memory model explains this pattern as a result of stereotyping. When an IDM is asked to make a forecast for inflation or output, what comes to mind most frequently is a combination of values where one variable exceeds its mean whereas the other falls below its mean, since these cases belong to the similarity bags with the lowest entropy and are therefore highly memorably. This is the case since these similarity bags consist of outcomes where the valences align to "good"/"good" or "bad"/"bad". It is as if the IDM follows an "all good/bad in one" heuristic. However, this heuristic derives endogenously from associative memory. Intuitively, the condition that α exceeds a certain threshold means that an item in the memory database comes to mind more frequently only when it is sufficiently "stereotypable". To the degree that stereotypes are prevalent in intuitive thinking, it is realistic that this condition holds in practice.

5 A New Keynesian Economy with Dual-System Expectations

In this section, we illustrate our framework from the previous section by applying it to the textbook version of the New Keynesian model in Galì (2015). In a first subsection, we describe the theoretical model and derive its predictions about impulse responses to fundamental macroeconomic shocks, as well as to autonomous innovations in intuitive thinking (sentiment innovations, for short). In the second subsection, we provide some quantitative results.

5.1 A Simple New Keynesian Model with Dual-System Expectations

Following Galì (2015), we consider an economy with a representative household whose objective function in period t is

$$E_t \sum_{h=0}^{\infty} \beta^h \left(\frac{C_{t+h}^{1-\sigma} - 1}{1-\sigma} - \frac{N_{t+h}^{1+\phi}}{1+\phi} \right) Z_{t+h} \tag{27}$$

where β is the discount factor, C_{t+h} denotes a CES index of consumption, σ the coefficient of relative risk aversion, N_{t+h} stands for hours worked, and ϕ is the inverse of the Frisch elasticity of labor supply. Z_{t+h} is a stochastic exogenous state variable that acts as generating fluctuations in the discount rate and is commonly used to model exogenous fluctuations in demand. Firms operate in monopolistic competition. Their production function is

$$Y_t = A_t N_t^{1-\alpha}, \tag{28}$$

where Y_t denotes output, A_t total factor productivity, and α determines how marginal productivity—and thus marginal costs—change with hours worked N_t.

To apply the DSE framework of Section 4, the model needs to be transformed into its linear difference (or "Blanchard-Kahn") form. For this, the model is log-linearized. We deviate from Galì (2015) in a minor way: we consider output deviations from the steady state rather than from the natural level. We do so since we find it more plausible that the intuitive thinking part of dual-system agents centers around deviations from steady state ("trend") rather than deviations from the natural level. The latter is cognitively more demanding concept, partly also because it is unobservable; as such, it falls more into the realm of rigorous thinking. Furthermore, it is also more straightforward to link deviations from steady state—rather than from the natural level—to empirical data (in the form of deviations from trend growth). As discussed at the beginning of Section 4.1 and in Appendix A, for the derivation of the Blanchard-Kahn form, we assume that dual-system agents share the same understanding of the primitive building blocks and their mutual relationships as standard FIRE agents, as long as expectations only appear in symbolic form and are not "solved" (see Appendix A for the respective terminology). Thus, as long as expectations are not "solved", we can use standard mathematical (conditional) expectation operators E_t and dual-system expectations E_t^D interchangeably. Since the latter nest the former for the case $\delta = 1$, we use the E_t^D for stating the key macroeconomic relationships, such as the IS and Phillips curve.

We follow the standard practice of using lower-case letters for variables in logs. In particular, y_t denotes the log deviation of output from its steady state value. We henceforth refer to y_t simply as output. Furthermore, π_t denotes inflation. In logs, the resulting IS equation is given by

$$y_t = E_t^D y_{t+1} - \frac{1}{\sigma}\left(i_t - E_t^D y_{t+1} - \rho_d\right) + \frac{1}{\sigma}(z_t - E_t^D z_{t+1}), \tag{29}$$

where ρ_d denotes the discount rate. Using derivation steps almost identical to Galì (2015), the

New Keynesian Phillips curve is given by

$$\pi_t = \beta E_t^D y_{t+1} + \kappa y_t - \lambda \frac{1+\phi}{1-\alpha} a_t, \tag{30}$$

where $\lambda \equiv \frac{(1-\theta)(1-\beta\theta)(1-\alpha)}{\theta(1-\alpha+\alpha\epsilon)}$ and $\kappa \equiv \lambda(\sigma + \frac{\varphi+\alpha}{1-\alpha})$. As a third equation, we have the Taylor rule

$$i_t = \rho_d + \phi_\pi \pi_t + \phi_y y_t + \nu_t \tag{31}$$

where the parameters $\phi_\pi \geq 0$ and $\phi_y \geq 0$ specify the reaction of the central bank to inflation and the deviation of output from steady state, respectively, and ν_t is an exogenous state variable that leads to exogenous fluctuations in the nominal interest rate. For the logs of the three exogenous state variables a_t (TFP), z_t (discount rate shifter) and ν_t (monetary policy), we follow the standard assumption that they each follow an AR1 process. Thus, we have

$$\begin{aligned} a_t &= \rho_a a_{t-1} + \varepsilon_t^a, \\ z_t &= \rho_z z_{t-1} + \varepsilon_t^z, \\ \nu_t &= \rho_\nu \nu_{t-1} + \varepsilon_t^\nu. \end{aligned}$$

It is straightforward, to use the IS equation, the New Keynesian Phillips curve and the Taylor rule to derive the linear difference form of the model, which—in symbolic representation—is given by

$$\begin{pmatrix} y_t \\ \pi_t \end{pmatrix} = A \begin{pmatrix} E_t^D y_{t+1} \\ E_t^D y_{t+1} \end{pmatrix} + B(z_t - \nu_t) - B E_t^D z_{t+1} + \Omega \lambda \frac{1+\phi}{1-\alpha} \begin{pmatrix} \phi_\pi \\ -\sigma - \phi_y \end{pmatrix} a_t \tag{32}$$

with

$$A = \Omega \begin{bmatrix} \sigma & 1 - \beta\phi_\pi \\ \sigma\kappa & \kappa + \beta(\sigma + \phi_y) \end{bmatrix}; \quad B = \Omega \begin{pmatrix} 1 \\ \kappa \end{pmatrix}; \quad \Omega = [\sigma + \kappa\phi_\pi + \phi_y]^{-1}$$

For dual-system expectations, according to Proposition 1, the solution of equation (32) consists of the rigorous-thinking part x_t^r plus $1-\delta$ times the intuitive thinking part x_t^i. According to the proposition, each of the two solutions can be derived as the solution from a separate (stochastic) difference equation. When deriving the difference equation for the rigorous-thinking part, we are confronted with the fact that the exogenous discount rate shifter z_t appears also appears as an argument of the dual-system expectation operator $E_t^D y_{t+1}$ (originating from the IS equation). For this to be well-defined, we adopt the convention that the agent has no sentiments about exogenous state variables. More precisely, the respective sentiments would take on a value of zero. Formally, this means that, for any exogenous state variable ξ_t, we assume

$$E_t^D \xi_{t+1} = \delta E_t \xi_{t+1} \tag{33}$$

Inserting dual-system expectations into (32), and using Proposition (1), the difference equation for the rigorous-thinking part is given by the solution to the stochastic difference equation

$$\begin{pmatrix} y_t \\ \pi_t \end{pmatrix} = \delta A \begin{pmatrix} E_t y_{t+1} \\ E_t \pi_{t+1} \end{pmatrix} + B(z_t - \nu_t) - \delta B E_t z_{t+1} + \Omega \lambda \frac{1+\phi}{1-\alpha} \begin{pmatrix} \phi_\pi \\ -\sigma - \phi_y \end{pmatrix} a_t. \tag{34}$$

Note that all expectation terms have now the usual meaning of a conditional mathematical expectation operator. The intuitive-thinking part is straightforward and is given by the solution to the difference equation

$$\begin{pmatrix} y_t \\ \pi_t \end{pmatrix} = \delta A \begin{pmatrix} y_{t+1} \\ \pi_{t+1} \end{pmatrix} + \begin{pmatrix} s_t^y \\ s_t^\pi \end{pmatrix}, \tag{35}$$

where (s_t^y, s_t^π) represent an inflation and output sentiment. Because of Assumption 1*(ii)*, we have to set $(s_{t+h}^y, s_{t+h}^\pi) \equiv (s_t^y, s_t^\pi)$. The reason is that the agent is unaware of sentiments perturbing expectation formation and—due to this unawareness—cannot make any prediction about how sentiments would change in the future. In the perception of the agent, sentiments are always zero as the agents see themselves as purely rigorous thinkers. The actual law of motion for sentiments is given by the VAR equation

$$\begin{pmatrix} s_t^y \\ s_t^\pi \end{pmatrix} = \Phi \begin{pmatrix} s_{t-1}^y \\ s_{t-1}^\pi \end{pmatrix} + \begin{pmatrix} \eta_t^y \\ \eta_t^\pi \end{pmatrix} \tag{36}$$

where the innovation terms are generally correlated, i.e., $cov(\eta_t^y, \eta_t^\pi) \neq 0$.

For the rigorous-thinking equation, it is straightforward to find the solution using the method of undetermined coefficients. The complete solution is obtained by adding the intuitive-thinking part. To state the solution, we make use of the following definitions:

$$\begin{aligned}
\Psi_{ya} &= \frac{\lambda(1+\varphi)(\phi_\pi - \delta\rho_a)}{(1-\alpha)[(\sigma(1-\delta\rho_a)+\phi_y)(1-\delta\beta\rho_a)+\kappa(\phi_\pi-\delta\rho_a)]},\\
\Psi_{y\nu} &= -\frac{(1-\delta\beta\rho_v)}{(\sigma(1-\delta\rho_v)+\phi_y)(1-\delta\beta\rho_v)+\kappa(\phi_\pi-\delta\rho_v)}\\
\Psi_{yz} &= \frac{(1-\delta\rho_z)(1-\delta\beta\rho_z)}{(\sigma(1-\delta\rho_z)+\phi_y)(1-\delta\beta\rho_z)+\kappa(\phi_\pi-\delta\rho_z)}\\
\Psi_{\pi a} &= -\frac{\lambda(1+\varphi)(\sigma(1-\delta\rho_a)+\phi_y)}{(1-\alpha)[(\sigma(1-\delta\rho_a)+\phi_y)(1-\delta\beta\rho_a)+\kappa(\phi_\pi-\delta\rho_a)]}\\
\Psi_{\pi\nu} &= -\frac{\kappa}{(\sigma(1-\delta\rho_v)+\phi_y)(1-\delta\beta\rho_v)+\kappa(\phi_\pi-\delta\rho_v)},\\
\Psi_{\pi z} &= \frac{\kappa(1-\delta\rho_z)}{(\sigma(1-\delta\rho_z)+\phi_y)(1-\delta\beta\rho_z)+\kappa(\phi_\pi-\delta\rho_z)},
\end{aligned} \tag{37}$$

Furthermore, we define $\Gamma = \left[(1-\delta\Omega\sigma)(1-\delta\Omega(\kappa+\beta(\sigma+\phi_y))) - \delta^2\Omega^2\sigma\kappa(1-\beta\phi_\pi)\right]^{-1}$. The solution for the New Keynesian model under dual-system expectations is then given by the following proposition.

PROPOSITION 4 *Suppose that*

$$\delta < \phi_\pi + \frac{1-\beta\delta}{\kappa}\phi_y + \frac{\sigma}{\kappa}\left[1-\delta(1+\beta(1-\delta))\right]. \tag{38}$$

Then both eigenvalues of δA are strictly inside the unit circle and there exists a unique solution to (34). This solution is given by

$$y_t^D = \Psi_{ya}a_t + \Psi_{yv}v_t + \Psi_{yz}z_t + (1-\delta)\Gamma\Omega\left[\sigma(1-\delta\Omega\beta(\sigma+\phi_y+\kappa\phi_\pi))s_t^y + (1-\beta\phi_\pi)s_t^\pi\right], \tag{39}$$

and

$$\pi_t^D = \Psi_{\pi a}a_t + \Psi_{\pi v}v_t + \Psi_{\pi z}z_t + (1-\delta)\Gamma\Omega\left[\sigma\kappa s_t^y + [\kappa+\beta(\sigma+\phi_y) - \delta\Omega\sigma\beta(\kappa\phi_\pi+\sigma+\phi_y)]s_t^\pi\right], \tag{40}$$

respectively. The evolution of sentiments s_t^y and s_t^π is given by (36).

Condition (38) is required for the uniqueness of the solution to the dual-system version of the New Keynesian model. For $\delta = 1$, this reduces to the standard condition $\kappa(\phi_\pi - 1) + (1-\beta)\phi_y > 0$.[21] It is well known that, in the standard case, the two Taylor coefficients ϕ_π and ϕ_y need to

[21] See equation (14) in Ch. 4 in Galì (2015).

be sufficiently large to fulfill this condition. In our case, the condition can still be fulfilled for $\phi_\pi = \phi_y = 0$, provided that δ is sufficiently small. In fact, one can show that the parameter space satisfying (38) is somewhat larger than the corresponding parameter space in Gabaix' (2020) model of cognitive discounting.[22]

Equations (39) and (40) in Proposition 4 highlight the two channels by which dual-system expectations change the fluctuation dynamics in a New Keynesian macroeconomy, in comparison to the standard case. First, dual-system expectations moderate the impulse responses to shocks in exogenous fundamentals, such as TFP, discount rate shifts, and monetary policy. The expressions for the coefficients in (37) show that δ always appears in combination with one of the autoregressive coefficients ρ_a, ρ_ν, or ρ_z. Thus, when δ is lower, it has the same effect as reducing the autoregressive coefficients of the exogenous fundamentals. In almost all cases, this leads to a dampening effect. The exception is the impulse response of output to a shock in the discount rate shifter. Since δ appears in the denominator of Ψ_{yz} twice, a lower δ increases the impulse response. The deeper reason for this is that, for any $h > 0$, the discount rate shifter appears twice in the IS equation, once as $E_t z_{t+h}$, and once as $E_t z_{t+h+1}$ (forward the RHS of (34) by h periods to see this). While multiplying the first with δ has a dampening effect, multiplying the second has an amplifying effect, which happens to dominate.

Second, with dual-system expectations there is a new source of fluctuations in the form of *sentiments and their innovations.* The lower δ, the more important is their influence. The resulting expressions are easiest to understand for the extreme case that $\delta = 0$, such that expectations are exclusively determined by sentiments. Assuming that all exogenous fundamentals states—i.e. TFP, discount factor, monetary policy—are zero, we then get

$$\begin{aligned} y_t^D &= \Omega \left[\sigma s_t^y + (1 - \beta\phi_\pi) s_t^\pi\right] \\ \pi_t^D &= \Omega \left[\sigma\kappa s_t^y + (\kappa + \beta(\sigma + \phi_y)) s_t^\pi\right] \end{aligned} \tag{41}$$

Inspecting the equations, it is immediately clear that both inflation and output increase with the output sentiment s_t^y. However, if we are interested in the effect of an exogenous innovation in output sentiment, we should take into account that, when sentiments are shaped by associative memory, an innovation in the output sentiment is negatively correlated with an innovation in the inflation sentiment (see Section 3). Thus, everything else equal, an increase

[22]More precisely, this holds if one sets $M = M^f$ (and equal to our δ) in Gabaix' (2020) Proposition 3. In his case, there would be a one instead of δ on the LHS if the inequality (38).

in the output sentiment is associated with a lower expected value of the inflation sentiment. Under the intuitive stagflationary model, the inflation-driving effect of the output sentiment is thus dampened by the negative inflation sentiment. The effect of the inflation sentiment on output depends on the sign of $(1-\beta\phi_\pi)$. A typical calibration for ϕ_π is 1.5, while β is assumed to be close to one for quarterly frequencies. This implies $(1-\beta\phi_\pi) < 0$. In this constellation, a negative inflation sentiment in the equation for output reinforces a positive output sentiment. Similar reasoning also applies to the case of $\delta > 0$, provided it is not too large.

One notable difference between the impulse responses to a shock in the exogenous fundamentals (TFP, discount rate, monetary policy) and sentiments is that the former crucially depend on the autoregressive coefficient. In the case of sentiments, this effect is completely absent. This follows from the unconscious nature of intuitive thinking, which makes the agent unaware of sentiments perturbing expectation formation. In the agent's own view, sentiments are identically zero. Hence, the agent is not in a position to anticipate that in the future, the impact of the current sentiment shock will be weaker and eventually vanish.

One question that may arise is whether the variation in output and inflation that we attribute to sentiments and their innovations could also be understood as resulting from discount rate shocks. Inspection of (37) shows that—in the standard case with $\delta = 1$—a positive discount rate shock leads to an increase in output and in inflation. In principle, if we consider an isolated positive shock in s_t^π, this has the same qualitative effect (see next subsection for a discussion of the quantitative effects). However, an important motivation for our sentiments is that they allow us to capture intuitive models and thus a negative association between output and inflation. For this, the output and inflation sentiment should be considered as a "package". This may raise the question whether a similar effect could be achieved when combining discount rate shocks with TFP shocks. However, a TFP shock is clearly associated with the idea of representing something "real", while the output sentiment belongs to the realm of intuitive thinking and is not, in the same sense, "real". Thus, if we were to use the menu of shocks available in the literature, we would have to combine a preference shock with a pure noise shock about TFP (Barsky and Sims 2012; Lorenzoni 2009), which is also not "real". This would allow for capturing the same beliefs about output and inflation as we do with sentiments. However, it is not a priori clear why discount rate shocks should be (positively) correlated with noise shocks about TFP. By contrast, a negative correlation between output and inflation sentiments is supported by the model of associative memory in Section 4.4.

5.2 Quantitative Results

We now provide quantitative results for the DSE version of the New Keynesian model. We start with impulse response functions for the conventional fundamental shocks (monetary policy, discount rate and technology). The parametrization is as follows: for δ, we use a baseline value of 0.85, guided by the cognitive discounting parameter from Gabaix (2020). We also add the two polar cases $\delta = 0$ and the standard case of $\delta = 1$. All other parameters are identical to those of the baseline model in Chapter 3 of Galì (2015).[23] The results are shown in Figure 6. As to be expected after the theoretical discussion in Subsection 5.1, the main pattern is that impulse responses are dampened under dual-system expectations relative to the standard case of $\delta = 1$. The exception is the discount rate shock, which has already been discussed in the previous subsection. Overall, the main impression is a striking similarity to the standard case of $\delta = 1$, at least for moderate deviations of δ from one.

We next turn to a quantitative analysis of the impact of sentiment shocks on output and inflation. Setting all other shocks to zero (i.e. $\nu_t = a_t = z_t = 0$), it follows directly from Proposition 4 that output and inflation are given by

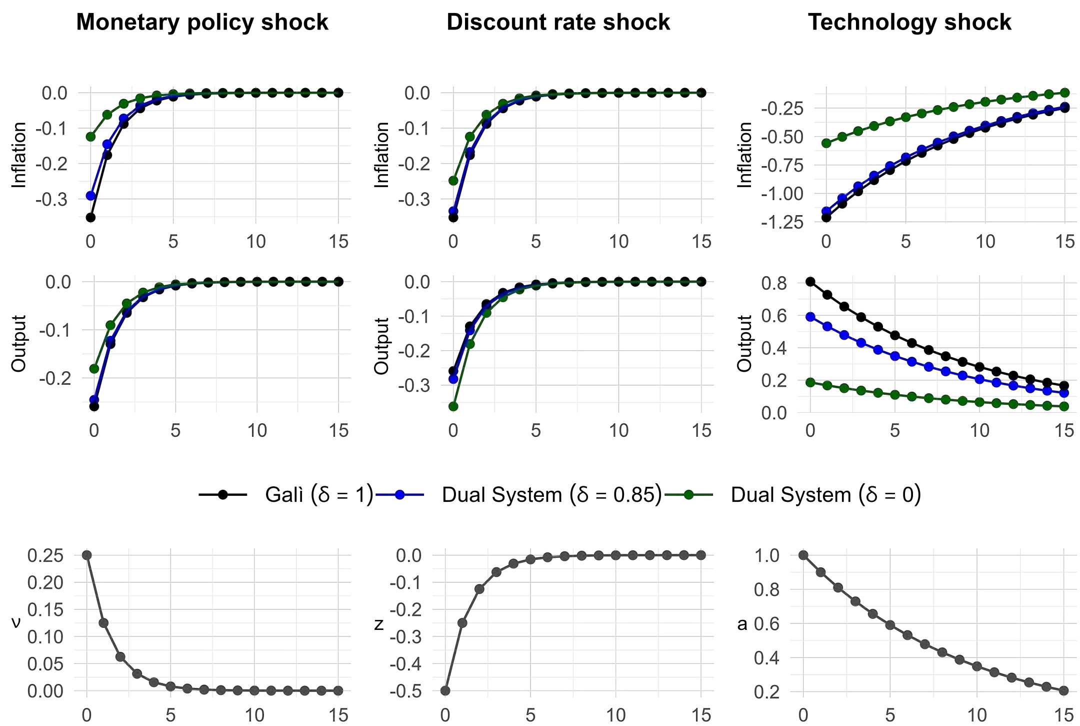
$$\begin{aligned} y_t^D &= (1-\delta)\Gamma\Omega\left[\sigma(1-\delta\Omega\beta(\sigma+\phi_y+\kappa\phi_\pi))s_t^y + (1-\beta\phi_\pi)s_t^\pi\right] \\ \pi_t^D &= (1-\delta)\Gamma\Omega\left[\sigma\kappa s_t^y + [\kappa+\beta(\sigma+\phi_y) - \delta\Omega\sigma\beta(\kappa\phi_\pi+\sigma+\phi_y)]s_t^\pi\right] \end{aligned} \tag{42}$$

where the sentiments s_t^y and s_t^π follow the VAR process specified in (36).

We now derive impulse responses for exogenous changes in sentiments. It follows from the analysis of associative memory in Section 4.4 that it is not always meaningful to decompose the VAR innovations into independent shocks. When intuitive expectations are formed by associative memory, expectation of output and inflation may generically be “packaged together”. This would manifest itself in correlated innovation terms in η_t in (36). It may then not always be meaningful to unpack these into independent components ε_t. Still, not least in the interest of allowing a comparison to typical results in the literature, we find it of interest to follow the identification strategy lined out at the end of Section 3, (see the discussion around equations (1) and (2)). The identifying assumption (2) with $\tau > 0$ is arguably the best way to capture the

[23] For simplicity, $\sigma = 1$, i.e. log-utility is assumed. The labor elasticity $\varphi = 5$ implies a Frisch elasticity of 0.2, the discount factor $\beta = 0.99$ implies a steady state real annualized return on financial assets of about 4% and the elasticity of substitution $\epsilon = 9$ implies a steady state markup of 12.5%. Furthermore, $\theta = 0.75$, the degree of price stickiness, implies an average price duration of 4 quarters and $\alpha = 0.25$ is the standard capital share. For the monetary policy rule, $\phi_\pi = 1.5$ and $\phi_y = 0.125$ are the values consistent with the original Taylor rule.

Figure 6: Monetary policy, discount rate and technology shocks



NOTES: The sub-figures in the upper panel show impulse responses for annualized inflation and output triggered by shocks to monetary policy, the discount rate, and technology for three different levels of δ. The lower panel shows the time path of the exogenous fundamental states (monetary policy ν_t, discount rate z_t, technology a_t) in response to a one-time exogenous shock to the respective state. The calibration of the shock sizes follows Galì (2015).

"all good/bad in one" nature of intuitive expectations within a linear framework. We therefore use it here for a calibration of impulse responses for sentiment shocks in our dual-system version of a simple NK model. As far as sentiments are concerned, our calibration of parameters is informed by the empirical analysis in Section 3.

A remaining challenge for the calibration of impulse responses for structural shocks ε_t (where $\eta_t = H\varepsilon_t$) is to determine the size of the shock in such a way that the magnitudes of the impulse responses have a natural meaning and refer to a "typical" shock size. By identification, the structural shocks ε_t have a standard deviation of 1. A shock sized "one standard deviation" and the resulting impulse responses have thus no natural quantitative interpretation. Our procedure to determine the size of a shock such that the resulting impulse responses can be naturally interpreted is by targeting two estimated impulse response coefficients from the local projection estimates in Section 3. Denote $\beta_{ij}^{(h)}$ as the local projection coefficient for the reaction of variable i to a shock in expectations about variable j at a horizon $h \geq 0$, where i and j refer to the variables output and inflation. The coefficient $\beta_{ij}^{(h)}$ indicates the size of the impulse response of the impacted variable i at horizon h due to an expectation shock of one standard deviation concerning variable j. From the perspective of model-based calibrations, rather than empirical estimations, the same impulse response is derived from equation (42), using (36) and (1). Denote the respective model-based impulse response by $\lambda_{ij}^{(h)}$. We can now choose the size of the shock ε_j in such a way that

$$\beta_{ij}^{(h)} = \lambda_{ij}^{(h)} \bar{\varepsilon}_j.$$

to target the model-based impulse response of variable i to expectations about j at horizon h to its empirical estimate. This yields $\bar{\varepsilon} = \beta_{ij}^{(h)} / \lambda_{ij}^{(h)}$. Note that the numerical value of $\lambda_{ij}^{(h)}$ is determined by the parameter calibrations discussed above, and empirical estimates for Φ and H.[24] To understand the logic of this target procedure, consider the case that $\beta_{ij}^{(h)}$ is large while $\lambda_{ij}^{(h)}$ is small. The targeting then means increasing the size of the shock in proportion to the inverse of $\lambda_{ij}^{(h)}$, to obtain an impulse response of exactly $\beta ij^{(h)}$.

Since our estimations are noisy, and since there is no habit formation in our model that would slow an initial response, we do not implement the targeting in a literal sense. Rather, we take a value of 0.15 as broadly reflecting the responses of inflation to intuitive inflation expectations within the first year. Furthermore, we take a value of 0.1 to broadly reflect the responses of output to output inflation expectations within the first year, based on Panel D in Figure 5. We then set $\lambda_{\pi\pi}^{(0)} \bar{\varepsilon}_j = 0.15$ and $\lambda_{yy}^{(0)} \bar{\varepsilon}_j = 0.1$. Since our "small" NK model is very stylized in any case and merely serves as an illustration for applying our general model, we view this "wholesale" approach to the calibration as justified. The values of $\lambda_{\pi\pi}^{(0)}$ and $\lambda_{yy}^{(0)}$ are determined using the calibrated parameter values discussed above.

The four calibrated impulse responses are shown in the upper panel of Figure 7. The left column shows the results for an inflation sentiment shock, and the right column for an output sentiment shock. The blue curves show calibrations for $\delta = 0.85$, the green curves refer to the extreme case of $\delta = 0$. In the latter case, the size of the shock is not targeted to the

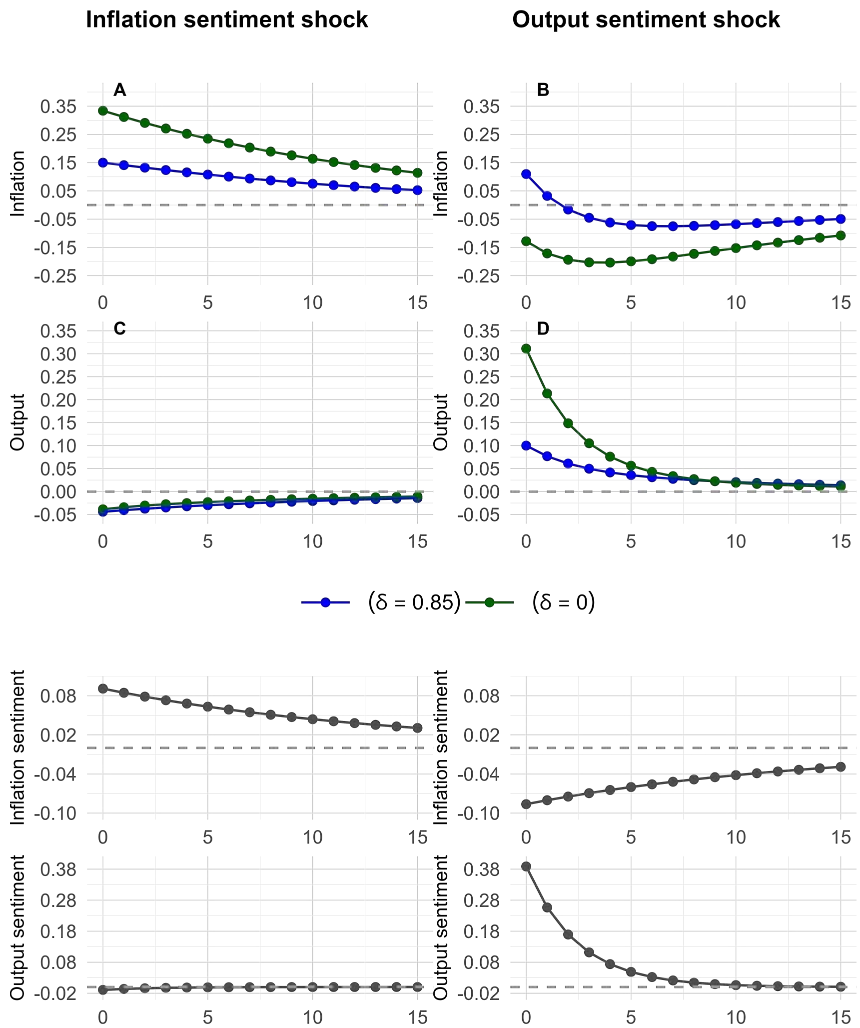
[24] Specifically, we take the purification with 8 lags and leads of TFP, and 8 lags for IST and oil news shocks. We get

$$\hat{\Phi} = \begin{pmatrix} 0.66 & -0.06 \\ 0.00 & 0.93 \end{pmatrix}.$$

and

$$\hat{H} = \begin{pmatrix} 0.99 & -0.10 \\ -0.22 & 0.98 \end{pmatrix}. \tag{43}$$

Figure 7: Impulse responses to sentiment shocks



NOTES: The sub-figures in the upper panel show impulse responses for annualized inflation and output, triggered by independent structural shocks to inflation and output expectations as discussed in the main text. The curves for $\delta = 0.85$ are targeted to the empirical impulse response by a commensurate choice of the shock size. The shock size for $\delta = 0$ is the same as for $\delta = 0.85$. The lower panel shows the impact of the structural shocks on the sentiments s_t.

empirical impulse response but kept at the same size as in the case of $\delta = 0.85$. The lower panel in Figure 7 shows the impact of the structural expectation shocks ε_t on sentiments s_t. The impulse responses for a positive inflation sentiment shock closely resemble those for a negative supply shock in a standard model.

The impulse responses for a positive output sentiment shock resemble those for a positive supply shock, with the added nuance that inflation first increases, at least in the case of

$\delta = 0.85$. The figure also illustrates that impulse responses tend to react quite strongly to δ. Comparing the impulse responses in Figure 7 to those in Figure 6, it is striking that none of the latter closely resemble the sentiment shocks.

Since our NK model is very simple, we would not expect that impulse responses, calculated for standard parameter values, would closely correspond to the empirical estimates in Figures 4 or 5. The overlap is highest for the effects of sentiment shocks on inflation (Panel A). The model-based impulse responses also capture the negative effect of an output sentiment shock on inflation (Panel B). The empirical curves resembles more closely to the model-based curve for a low value of δ rather than for $\delta = 0.85$, due to the fact that, for a low δ, the curve first decreases and returns towards zero. Both the empirical estimates in Figure 5 and the model-based curve for the effect of an output sentiment shock on output show a positive reaction (Panel D). However, in the model-based case, the impact is highest initially, whereas the empirical curves gradually increase. In the case of the output reaction to an inflation expectation shock, the empirical estimates are positive, though not significant, while the model-based effect is negative, but very small. The lower-left panel in the lower part of the figure shows that this is the case since the output sentiment does almost not react to a shock in the inflation sentiment. This is due to our estimate of the identification matrix H. Recall that for the standard rational-expectation case of $\delta = 1$, all model-based curves would coincide with the horizontal axis as there would be no effect.

6 Conclusion

Motivated by empirical evidence about individual expectations, we incorporate dual-system processing into an otherwise standard dynamic macroeconomic framework. Our framework features autonomous innovations in expectations resulting from intuitive thinking, which we understand as resulting from associative memory. As a concrete application, we implement our framework in the standard New Keynesian (NK) model.

While in the current paper the cognitive perturbations in expectations are assumed to be completely autonomous, i.e., independent of the state of the economy, it is plausible that in practice they can be connected with current economic events. For instance, it would be natural to consider that perturbations in expectations can be triggered by hyped media reporting (such as when a technical recession is portrayed as a slump) or by the use of decisive language by central banks (like a pledge to "do whatever it takes"). To what extent would such connections amplify the impact of the very shocks that underlie the hyped reporting or the use of decisive language? What would be the implications for the conduct of monetary policy? Our framework is a flexible tool within which these and related questions can be addressed. At this point, we leave a detailed exploration of these aspects for future research.

References

Andre, Peter et al. (2022). "Subjective Models of the Macroeconomy: Evidence From Experts and Representative Samples". *Review of Economic Studies* 89, pp. 2958–2991.

Andre, Peter et al. (2023). "Narratives about the Macroeconomy". CESifo Working Paper No. 10535.

Angeletos, George-Marios, Fabrice Collard, and Harris Dellas (2018a). "Quantifying Confidence". *Econometrica* 86.5, pp. 1689–1726.

— (2018b). "Quantifying Confidence". *Econometrica* 86.5, pp. 1689–1726.

Angeletos, George-Marios and Chen Lian (2023). "Dampening General Equilibrium: Incomplete Information and Bounded Rationality". In: *Handbook of Economic Expectations*. Elsevier, pp. 613–645.

Barsky, Robert B. and Eric R. Sims (2012). "Information, Animal Spirits, and the Meaning of Innovations in Consumer Confidence". *American Economic Review* 102.4, pp. 1343–77.

Ben Zeev, Nadav (2018). "What can we learn about news shocks from the late 1990s and early 2000s boom-bust period?" *Journal of Economic Dynamics and Control* 87, pp. 94–105. ISSN: 0165-1889. DOI: https://doi.org/10.1016/j.jedc.2017.12.003. URL: https://www.sciencedirect.com/science/article/pii/S0165188917302518.

Bhandari, Anmol, Jaroslav Borovička, and Paul Ho (2023). "Survey data and subjective beliefs in business cycle models". Federal Reserve Bank of Richmond Working Paper.

Bianchi, Francesco, Sydney C. Ludvigson, and Sai Ma (2022). "Belief Distortions and Macroeconomic Fluctuations". *American Economic Review* 112.7, pp. 2269–2315.

Blanchard, Olivier J. and Charles M. Kahn (1980). "The Solution of Linear Difference Models under Rational Expectations". *Econometrica* 48.5, pp. 1305–1311.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2020a). "Memory, Attention, and Choice". *Quarterly Journal of Economics* 135.3, pp. 1399–1442.

Bordalo, Pedro et al. (2020b). "Overreaction in Macroeconomic Expectations". *American Economic Review* 110.9, pp. 2748–2782.

Bordalo, Pedro et al. (2023). "Memory and Probability". *Quarterly Journal of Economics* 138.1, pp. 265–311.

Candia, Bernardo, Olivier Coibion, and Yuriy Gorodnichenko (2020). "Communication and the Beliefs of Economic Agents". NBER Working Paper 27800.

— (2021). "The Inflation Expectations of U.S. Firms: Evidence from a new survey". NBER Working Paper 28836.

Coibion, Olivier and Yuriy Gorodnichenko (2012). "What Can Survey Forecasts Tell Us about Information Rigidities?" *Journal of Political Economy* 120.1, pp. 116–159.

— (2015a). "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts". *American Economic Review* 105.8, pp. 2644–2678.

— (2015b). "Is the Phillips Curve Alive and Well after All? Inflation Expectations and the Missing Disinflation". *American Economic Journal: Macroeconomics* 7.1, pp. 197–232.

Dräger, Lena, Michael J. Lamla, and Damjan Pfajfar (2016). "Are survey expectations theory-consistent? The role of central bank communication and news". *European Economic Review* 85, pp. 84–111.

Eliaz, Kfir and Ran Spiegler (2020). "A Model of Competing Narratives". *American Economic Review* 110.12, pp. 3786–3816.

Enders, Zeno, Michael Kleemann, and Gernot J. Müller (2021). "Growth Expectations, Undue Optimism, and Short-Run Fluctuations". *The Review of Economics and Statistics* 103.5, pp. 905–921.

Enke, Benjamin, Frederik Schwerter, and Florian Zimmermann (2020). "Associative Memory and Belief Formation". NBER Working Paper 26664.

Evans, Jonathan St. B. T. and Keith E. Stanovich (2013). "Dual-Process Theories of Higher Cognition: Advancing the Debate". *Perspectives on Psychological Science* 8.3. PMID: 26172965, pp. 223–241. DOI: 10.1177/1745691612460685. eprint: https://doi.org/10.1177/1745691612460685. URL: https://doi.org/10.1177/1745691612460685.

Fernald, John (2014). "A quarterly, utilization-adjusted series on total factor productivity". In: Federal Reserve Bank of San Francisco.

Flynn, Joel P. and Karthik Sastry (2022). "The Macroeconomics of Narratives". SSRN Working Paper.

Gabaix, Xavier (2020). "A Behavioral New Keynesian Model". *American Economic Review* 110.8, pp. 2271–2327.

Galì, Jordi (2015). *Monetary Policy, Inflation, and the Business Cycle*. Princeton University Press.

Garcìa-Schmidt, Mariana and Michael Woodford (2019). "Are Low Interest Rates Deflationary? A Paradox of Perfect-Foresight Analysis". *American Economic Review* 109.1, pp. 86–120.

Hommes, Cars, Domenico Massaro, and Matthias Weber (2019). "Monetary Policy under Behavioral Expectations: Theory and Experiment". *European Economic Review* 118, pp. 193 –212.

Ilut, Cosmin and Rosen Valchev (2023). "Economic Agents as Imperfect Problem Solvers". *Quarterly Journal of Economics* 138.1, 313–362.

Jordà, Òscar (2005). "Estimation and Inference of Impulse Responses by Local Projections". *The American Economic Review* 95.1, pp. 161–182.

Kahneman, Daniel (2003). "Maps of Bounded Rationality: Psychology for Behavioral Economics". *American Economic Review* 93.5, pp. 1449–1475.

— (2011). *Thinking, Fast and Slow.* Random House.

Kamdar, Rupal (2019). "The inattentive consumer: Sentiment and expectations". Working Paper.

Känzig, Diego R (2021). "The macroeconomic effects of oil supply news: Evidence from OPEC announcements". *American Economic Review* 111.4, pp. 1092–1125.

Lorenzoni, Guido (2009). "A Theory of Demand Shocks". *American Economic Review* 99.5, pp. 2050–2084.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt (2023). "Rational Inattention: A Review". *Journal of Economic Literature* 61.1, pp. 226–73. DOI: 10.1257/jel.20211524. URL: https://www.aeaweb.org/articles?id=10.1257/jel.20211524.

Montiel Olea, José Luis and Mikkel Plagborg-Møller (2021). "Local Projection Inference Is Simpler and More Robust Than You Think". *Econometrica* 89.4, pp. 1789–1823.

Pineda, Luis A., Gibrán Fuentes, and Rafael Morales (Mar. 2021). "An entropic associative memory". *Scientific Reports* 11.1, p. 6948. ISSN: 2045-2322. DOI: 10.1038/s41598-021-86270-7. URL: https://doi.org/10.1038/s41598-021-86270-7.

Pineda, Luis A. and Rafael Morales (June 2023). "Imagery in the entropic associative memory". *Scientific Reports* 13.1, p. 9553. ISSN: 2045-2322. DOI: 10.1038/s41598-023-36761-6. URL: https://doi.org/10.1038/s41598-023-36761-6.

Ramey, V.A. (2016). "Chapter 2 - Macroeconomic Shocks and Their Propagation". In: ed. by John B. Taylor and Harald Uhlig. Vol. 2. Handbook of Macroeconomics. Elsevier, pp. 71–162.

Shiller, Robert J. (2017). "Narrative Economics". *American Economic Review* 107.4, pp. 967–1004.

Shiller, Robert J. (2019). *Narrative Economics. How Stories Go Viral and Drive Major Economic Events*. Princeton University Press.

Appendix A: Symbolic vs. Quantitative Representation of Relationships with Expectations

In this Appendix, we discuss how a dual-system processing agent would arrive the Blachard-Kahn equation (5), or (34) in the case of the NK model, from first principles. One possible reaction to this is to view these equations as reasonably good approximations of how individuals reason about the economy (or, rather, have learned to reason about it). As a matter of fact, this may not reflect the reasoning of any single individual, but rather how a "statistical aggregate" of individuals happens to reason about the economy collectively. While we are sympathetic to this view, we still find it intellectually rewarding to consider more rigorously what assumptions are needed such that (5) or (34) would emerge in the reasoning of an agent who starts from first principles. For this, we introduce the concept of a symbolic representation of a macroeconomic relationship involving expectations, as opposed to its quantitative representation.

We define a symbolic representation of an economic relationship involving expectations as one in which expectation terms have a purely symbolic meaning, i.e. no operation is allowed that would "solve" any expression that appears as an argument to the expectation operator. In particular, is not permissible to express their symbols in terms of other entities that are not by themselves symbols for expectations. In symbolic representation, we denote time t expectation terms about x_{t+h}, $h > 0$, by "$E_t x_{t+h}$". It is permissible to perform time shifts to symbolic expectations. In particular, we assume that a shift of x_t by h periods into the future leads to "$E_t \pi_{t+h}$", and shifting "$E_t \pi_{t+1}$" by h periods leads to "$E_t \pi_{t+h+1}$", $h \geq 1$. We finally take expectation terms of white noise as an exception and assume that it is permissible to set "$E_t \varepsilon_t$" $= 0$, where ε_t represents white noise. We refer to a representation in which expectations terms can be solved in terms of other entities as a quantitative representation.

Imagine the description of an economy as represented by computer code organized in a set of programs. In a symbolic representation, whenever an expectation term appears, this is understood as a reference to a subroutine that remains unspecified. This subroutine is not part of the symbolic representation, it is only its name that has meaning. It derives from the fact that once the routine is added to the collection, the overall program will run and produce a quantitative representation of the economy. The point is that the subroutine could contain very different instructions, depending on how expectations are formed. The instructions may correspond to FIRE, to dual-system expectations, or any other type of expectations. The symbolic representation expresses, and only expresses, relationships between entities, some of them in the form of symbolic expectations, some as deterministic entities. It is silent on how expectations would obtain quantitative meaning.

Note that if we have y_t = "$E_t y_{t+1}$" and "$E_t y_{t+1}$" = "$E_t z_{t+1}$", it is permissible to form the equation y_t = "$E_t z_{t+1}$". In this case, we did not replace a symbolic expectation by a non-symbolic one. Consider next the equation x_t = "$E_t x_{t+1}$" $+ u_t$ with $u_t = \rho u_{t+1} + \varepsilon_t$ (where the latter equation replicates (6)). It follows from our definition above that it is admissible to derive "$E_t x_{t+1}$" = "$E_t x_{t+2}$" + "$E_t u_{t+1}$". However, it is *not* admissible to derive "$E_t x_{t+1}$" = "$E_t x_{t+2}$"$+E_t u_{t+1}$ with $E_t u_{t+1} = \rho u_t$ and form the recursive solution $x_t = \sum_{h=0}^{\infty} \rho^h u_t = \frac{1}{1-\rho} u_t$. This is not permissible since the symbolic expectation "$E_t u_{t+1}$" must not be equated to the standard mathematical expectation expression $E_t u_{t+1}$. In the image used above, this would mean adding a routine with code for the execution of a standard mathematical expectation, which is not a meaningful operation within a symbolic representation. It would only be allowed once we leave the realm of the symbolic representation and assume that an agent fills the routine with instructions to calculate standard mathematical operations.

Consider now the primitive building blocks of a dynamic macroeconomy. It includes households, firms, their objective functions and constraints, and possibly other actors such as the central bank or government. The behavioral equations for the different actors generally include expectation terms whenever agents are not myopic. For any standard dynamic economic model for which, after (log)linearization, equation (5) is obtained in linear form as the core equilibrium characterization, an equation exhibiting the same form can also be derived with replacing standard mathematical expectation terms by symbolic ones. This follows from the fact that expectations still appear in (5), thus have not yet been "solved".[25] We can simply keep all expressions of the form $E_t x_{t+h}$ in the equations, and we can also relabel them to "$E_t x_{t+h}$" and perform the same operations. As a result, a FIRE agent, i.e. an agent with $\delta = 1$ and $s_t \equiv 0$, and a DSE agent who start with the same primitive description of an economy, will agree on its symbolic representation. It's only when it comes to the quantitative representation, that the FIRE agent calls a "FIRE routine" for solving expectations of the form "$E_t x_{t+h}$", while the dual-system agent calls a "dual system routine". As a result, they will arrive at different quantitative values for x_t. This is why we refer to any solution of the core equations for the economy as belonging to its quantitative representation.

[25] In some versions of dynamic macro models, the standard form of (5) may have been obtained by solving some trivial stochastic difference equations and substituting the results. However, for linearized models, it is always possible to arrive at an equilibrium equation in the form of (5) while keeping the respective equations "unsolved".

That FIRE and dual-system agents agree on the symbolic representation of an economy is not a realistic assumption. If an agent has difficulties to form expectations about future macroeconomic variables, it is plausible that the same agent would also face difficulties in deriving equilibrium relationships from first principles. This provides an interesting research agenda, to which seminal contributions have been made by Garcìa-Schmidt and Woodford (2019) and Angeletos and Lian (2023). We do not pursue this here, for two main reasons. First, the assumption that dual-system agents share the symbolic representation of the economy with FIRE agents, increases tractability. We are interested in what insights we can obtain while keeping this tractability. Second, agents may actually not derive an equation like (**??**) from primitives but may have learned to reason about economic relationships "akin to" this equation.