NBER WORKING PAPER SERIES

FORWARD GUIDANCE WITHOUT COMMON KNOWLEDGE

George-Marios Angeletos
Chen Lian

Working Paper 22785
http://www.nber.org/papers/w22785

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2016, Revised February 2018

We have no financial interests or research support to disclose. We are grateful to the editor, John Leahy, and three anonymous referees for extensive feedback, and to Marco Del Negro, Alisdair McKay, Alexandre Kohlhas, Kristoffer Nimark, Paul Pichler, Franck Portier, and Mirko Wiederholt for conference discussions. For helpful comments, we also thank Adrien Auclert, Harris Dellas, Xavier Gabaix, Alessandro Pavan, Karthik Sastry, and seminar participants at the University of Chicago, Duke University, Northwestern University, Princeton University, the 2016 and 2017 NBER Summer Institutes, the 2016 ESEM in Edinburgh, the 2017 NBER EFG and ME meetings, the 2017 SED meeting, the 2018 AEA meeting, the NYU Stern IUB workshop, the ECB NSM workshop, the 2017 Hydra Workshop in Sardinia, the 20th Central Bank Macroeconomic Modeling Workshop at the Bank of France, the 2017 SNB/CEP Workshop at Study Center Gerzensee, the 2017 Workshop on Sentiments and Beliefs at the Bank of Chile. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Forward Guidance without Common Knowledge
George-Marios Angeletos and Chen Lian
NBER Working Paper No. 22785
October 2016, Revised February 2018
JEL No. C72,D82,E03,E32,E43,E52,E58

ABSTRACT

How does the economy respond to news about future policies or future fundamentals? Standard practice assumes that agents have common knowledge of such news and face no uncertainty about how others will respond. Relaxing this assumption attenuates the general-equilibrium effects of news and rationalizes a form of myopia at the aggregate level. We establish these insights within a class of games which nests, but is not limited to, the New Keynesian model. Our results help resolve the forward-guidance puzzle, offer a rationale for the front-loading of fiscal stimuli, and illustrate more broadly the fragility of predictions that rest on long series of forward-looking feedback loops.

George-Marios Angeletos
Department of Economics, E52-530
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
angelet@mit.edu

Chen Lian
Department of Economics
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
lianchen@mit.edu

1 Introduction

How does aggregate demand respond to forward guidance, that is, to news about future monetary policy? How does aggregate supply respond to news about future productivity? How do asset prices respond to news about future earnings?

Standard practice in macroeconomics and finance is to address such questions by assuming rational expectations along with common knowledge of the relevant news and of the structure of the economy. This imposes that the agents have a perfect and common understanding of the effects of the news on current and future economic outcomes. For instance, in the context of forward guidance, it is standard to assume not only that everybody is aware of and attentive to the policy announcement, but also that nobody doubts the awareness, the attentiveness, and the responsiveness of other agents. As a result, the agents have no doubts about how inflation and income—the product of the behavior of others—will adjust.

In this paper, we accommodate such doubts by removing common knowledge of the relevant news and introducing higher-order uncertainty, that is, by allowing the agents to be uncertain about the beliefs and the responses of others. This friction can be the by-product of either dispersed private information, as in Lucas (1972) and Morris and Shin (2002), or rational inattention and costly contemplation, as in Sims (2003) and Tirole (2015). It can thus be interpreted interchangeably as a form of coordination failure that is consistent with equilibrium uniqueness, and as a form of bounded rationality that is consistent with the rational-expectations equilibrium concept.

Our contribution is to show that this friction attenuates general-equilibrium effects, such as those associated with the Keynesian multiplier and the inflation-spending feedback in the New Keynesian model, and causes the economy to respond to news about the future *as if* the agents were myopic. These insights shed new light on topical policy questions such as the power of forward guidance and the optimal timing of fiscal stimuli. More generally, they highlight the fragility of predictions that rest on long series of forward-looking, general-equilibrium feedback loops.

The New Keynesian Model without Common Knowledge. In Sections 3 and 4, we revisit the New Keynesian model, relax its common-knowledge assumptions, and show how its building blocks, the modern versions of the IS and Philips curves, can be embedded into a class of games that we call dynamic beauty contests. This serves three goals. First, it clarifies the three general-equilibrium (GE) mechanisms that operate within the New Keynesian model: the spending-income multiplier running within the demand block, the strategic complementarity in price-setting running within the supply block, and the inflation-spending feedback running across the two blocks. Second, it unearths the role of higher-order beliefs that lie underneath these mechanisms. Third, it motivates the more abstract analysis of Section 5.

GE Attenuation and the Horizon Effect in Dynamic Beauty Contests. Section 5 studies a class of games in which optimal decisions today depend positively on expectations of the future decisions of others. It nests the building blocks of the New Keynesian model, but is not limited to them.[1]

[1]Note the crucial difference from the *static* beauty contests studied in Morris and Shin (2002), Woodford (2003*a*), Angeletos and Pavan (2007), and Bergemann and Morris (2013): in these works, behavior is *not* forward-looking.

Within this framework, we ask the following question: How does the aggregate outcome today (at $t = 0$) respond to the news about the fundamental in a future date (at $t = T \geq 1$)? In our leading application (forward guidance), this question refers to the response of aggregate spending and employment to the news about the policy rate T periods later.

Clearly, the answer to this question depends on how precise and credible the news is, because these factors control how much the forecasts of the policy instrument, or other fundamentals, vary in the first place. But this is not what we are after in this paper. Instead, we study the response of the aggregate outcome *relative* to that of the aforementioned forecasts.

Denote this relative response by ϕ_T. Next, let ϕ_T^* be the value that obtains in the frictionless, common-knowledge benchmark. This can be decomposed into two components:

$$\phi_T^* = \phi_T^{PE} + \phi_T^{GE},$$

where ϕ_T^{PE} captures the direct or partial-equilibrium (PE) effect, namely the response of the typical agent holding constant the response of others, and ϕ_T^{GE} captures the additional, general-equilibrium (GE) effect.[2] Under quite general assumptions on the information structure, we establish the following three properties about the value of ϕ_T that obtains away from the common-knowledge benchmark.

1. The lack of common knowledge dulls the GE effect. As a result, ϕ_T is bounded between ϕ_T^{PE} and ϕ_T^*, and is closer to ϕ_T^{PE} when there is more higher-order uncertainty. We refer to this property as "GE attenuation."

2. This attenuation is stronger the longer the horizon of the news or, equivalently, the longer the series of forward-looking, general-equilibrium feedback loops. As a result, the ratio ϕ_T/ϕ_T^* decreases with T. We refer to this property as the "horizon effect."

3. Under a mild condition, the attenuation grows without bound, and ϕ_T becomes vanishingly small relative to ϕ_T^*, as $T \to \infty$. This is our "limit result."

Let us explain the reasoning behind these findings. To begin with, note that the GE effect is driven by beliefs of the future behavior of others. In equilibrium, these beliefs can be expressed as functions of the hierarchy of beliefs of the future fundamental (e.g., the future policy rate). Understanding the GE effect is therefore akin to understanding the hierarchy of beliefs.

In the common-knowledge benchmark, the role of higher-order beliefs is concealed due to the property that higher-order beliefs co-move perfectly with first-order beliefs. But this property is degenerate: away from that benchmark, higher-order beliefs tend to move less than lower-order beliefs, because they are more anchored to the common prior. By the same token, the expectations of future endogenous outcomes adjust relatively little to the news and the GE effect is attenuated. This proves our first result.

[2] In our framework, the two components have the same sign, meaning that the GE effect reinforces the PE effect. Furthermore, PE_T can decrease with T, because of the discounting embedded in individual preferences. Yet, GE_T can increase with T, and may even explode as $T \to \infty$, because of powerful general-equilibrium feedback loops.

Our second result, the horizon effect, follows from the combination of the aforementioned anchoring of higher-order beliefs with a property that is embedded in the class of forward-looking games studied in this paper. In this class, longer horizons raise the relative importance of higher-order beliefs in equilibrium outcomes, thus also reinforcing the attenuation effect. This is because longer horizons involve more iterations of the forward-looking, Euler-like equations of the model, which in turn map to beliefs of higher order. In a nutshell, iterating on dynamic GE feedback loops is akin to ascending the hierarchy of beliefs.

Finally, our limit result follows from combining the above insight with the fact that infinite-order beliefs are pegged at the common prior and are therefore unresponsive to the news, even if the level of inattentiveness, or the idiosyncratic noise in the observation and interpretation of the news, is arbitrarily small. This result can be seen as a sharp illustration of the horizon effect. But it also contains the following lesson: predictions that hinge on long series of GE feedback loops are particularly fragile to relaxations of the complete-information, rational-expectations benchmark.[3]

We complement these results with another, which recasts the informational friction as a form of myopia: accommodating higher-order uncertainty is akin to having the representative agent discount the future more heavily at the aggregate level than what is rational at the individual level. Furthermore, the as-if discounting is larger the stronger the underlying GE effect.[4]

Policy Lessons. Suppose that the economy is in a liquidity trap and the zero lower bound (ZLB) on the nominal interest rate binds between today, $t = 0$, and some future date, $t = T - 1$ for some known $T \geq 2$. Because of this constraint, the monetary authority is unable to stimulate the economy by reducing its current policy rate. It can nevertheless try to achieve the same goal by committing to low rates at $t = T$ (or later). The baseline New Keynesian model predicts that the effectiveness of this kind of forward guidance increases with T and explodes as $T \to \infty$. What is more, the effect is quantitatively huge even for modest T.

These predictions constitute the so-called forward guidance puzzle (Del Negro, Giannoni and Patterson, 2015; McKay, Nakamura and Steinsson, 2016*b*). They are at odds not only with the available evidence (Campbell et al., 2012), but also with the logic that news regarding the distant future should be heavily discounted. This logic is based on PE reasoning. The puzzling predictions are driven by GE effects.

As noted earlier, there are three GE mechanisms at work: the modern version of the Keynesian multiplier, which runs inside the demand block; the dynamic strategic complementarity in the price-setting decisions of the firms, which runs inside the supply block; and the inflation-spending feedback, which runs across the two blocks. There mechanisms act as multipliers of the PE effects of monetary policy. Removing common knowledge arrests these multipliers, thus also bringing the predictions of the model closer to the aforementioned PE logic.

To illustrate, suppose that each agent worries that any other agent is unaware of, or inattentive to, the policy news with a 25 percent probability. Under a textbook parameterization (Galí, 2008), this kind of

[3] In the New Keynesian model, examples of such predictions include not only those regarding the power of forward guidance, but also those regarding the indeterminacy of interest-rate pegs and the so-called neo-Fisherian effects.

[4] Apart from offering a sharp representation of how the absence of common knowledge modifies the forward-looking behavior of the economy, this result also rationalizes the kind of myopia assumed in Gabaix (2016).

doubt reduces the power of forward guidance by about 90 percent at the 5-year horizon. Importantly, this attenuation is *relative* to the movements in *expectations* of interest rates. Our theory therefore helps explain why forward guidance may have a modest effect on expectations of inflation and income, and on actual activity, *relative* to its effect on the yield curve.

Turning to fiscal policy, the standard New Keynesian model predicts that, in the presence of a binding ZLB constraint, a fiscal stimulus of a given size is more effective when it is back-loaded, i.e. when it is announced now but implemented later on. This prediction hinges on the same GE feedback loops as those that govern the power of forward guidance. A variant of the aforementioned results therefore offers a rationale for the front-loading of fiscal stimuli: such front-loading improves coordination in the sense of reducing the bite of higher-order uncertainty.

Remark. Our work invites the analyst to study the role of higher-order beliefs. It does not, however, require that the agents themselves engage in higher-order reasoning. Instead, as in any rational-expectations context, it suffices that their expectations of the relevant economic outcomes, such as inflation and income, are consistent with actual behavior. From this perspective, our analysis is a revision of the predictions that the analyst can make, once she liberates the rational-expectations hypothesis from the auxiliary assumption that the agents face no doubts about the information, or the attentiveness, of others.

Layout. Section 2 discusses the related literature. Section 3 introduces our version of the New Keynesian model. Section 4 nests the demand and supply blocks of that model within the class of dynamic beauty contests that we consider. Section 5, which is self-contained, studies this class of games and develops the key theoretical results. Sections 6 and 7 work out the implications for, respectively, monetary and fiscal policy. Section 8 concludes the main text. The Appendix contains the proofs and additional results.

2 Related Literature

On the theoretical side, our paper builds heavily on the macroeconomic literature on incomplete information and beauty contests.[5] Some of this literature, including Morris and Shin (2002) and Woodford (2003*a*), focuses on *static* beauty contests, namely settings in which the agents only have to forecast the concurrent actions of others. By contrast, Allen, Morris and Shin (2006), Bacchetta and van Wincoop (2006), Morris and Shin (2006), and Nimark (2008, 2017) study *dynamic* beauty contests, namely settings in which the agents must forecast the future actions of others. Our paper shares with the latter set of works the emphasis on forward-looking expectations. But whereas these works focus on the effects of learning, our paper focuses on the horizon of the events that the agents have to forecast.

On the applied side, our paper adds to the recent literature on the forward-guidance puzzle. Del Negro, Giannoni and Patterson (2015) and McKay, Nakamura and Steinsson (2016*b*) seek to resolve the puzzle by introducing short horizons and liquidity constraints, and Andrade et al. (2015) do so by letting forward guidance confound good news about the future policy instrument with bad news about other fundamen-

[5] See Angeletos and Lian (2016*c*) for a survey.

tals. We make an orthogonal point: we show how higher-order uncertainty anchors income and inflation expectations, thus attenuating the GE effects of forward guidance.

Related forms of GE attenuation can be found in Farhi and Werning (2017) and Gabaix (2016). Farhi and Werning (2017) obtain such attenuation by replacing Rational Expectations Equilibrium (REE) with Level-k Thinking, an approach that follows the lead of Garcıa-Schmidt and Woodford (2015). Gabaix (2016) achieves the same goal by assuming that beliefs are biased in the direction of underestimating the persistence of the underlying shocks and of their effects on the economy. We refer the interested reader to Angeletos and Lian (2016*a*) for a detailed exposition of how these approaches connect to and differ from ours.[6]

Closely related is the earlier work of Wiederholt (2015). That paper notes that higher-order uncertainty can dull the inflation-spending feedback of the New Keynesian model, thus also reducing the power of forward guidance. However, that paper does not consider how higher-order uncertainty can attenuate the GE effects that run inside each block of the model. Most importantly, it does not consider how the impact of higher-order uncertainty on equilibrium outcomes and the resulting weakening of forward guidance vary with the horizon of the events that the agents need to forecast, which is the core contribution of our paper.

Finally, Chung, Herbst and Kiley (2015) and Kiley (2016) argue that some of the paradoxical predictions of the New Keynesian model are resolved once the nominal rigidity is attributed to sticky information as in Mankiw and Reis (2002). But this is largely because these works abstract entirely from the price stickiness seen at the micro data and from the forward-looking aspect in the firms' price-setting decisions. These works also rule out information frictions among the consumers. They therefore do not share our insights regarding the anchoring of forward-looking expectations and the attenuation of the associated GE effects.

3 Framework

In this section, we introduce the framework used for the applied purposes of the paper. This is the same as the textbook New Keynesian model (Woodford, 2003*b*; Galí, 2008), except that we allow the agents to face uncertainty about the information, the beliefs, and the behavior of others.

Consumers. There is a measure-one continuum of ex-ante identical consumers in the economy, indexed by $i \in \mathcal{I}_c = [0, 1]$. Preferences are given by

$$\mathcal{U}_{i,0} = \sum_{t=0}^{+\infty} \beta^t U(c_{i,t}, n_{i,t}), \tag{1}$$

where $c_{i,t}$ and $n_{i,t}$ denote the consumer's consumption and labor supply at period t, $\beta \in (0, 1)$ is the discount

[6] Let us briefly note the key differences. Because our approach is compatible with the REE concept, it bypasses a challenge faced by the aforementioned works, namely what the agents do when they experience again and again a reality that is inconsistent with their beliefs. Our approach is also easier to reconcile with the evidence in Coibion and Gorodnichenko (2012, 2015) and Vellekoop and Wiederholt (2017), which indicate that the dynamics of expectations, as measured in surveys, is broadly consistent with models that maintain the REE concept but allow for informational frictions and slow learning. Finally, while the form of GE attenuation captured by our approach is robust to settings in which GE effects offset PE effects (think of settings featuring strategic substitutability), this is not the case for Level-k Thinking.

factor, and U is the per-period utility function. The latter is specified as

$$U(c,n) = \tfrac{1}{1-1/\sigma} c^{1-1/\sigma} - \tfrac{1}{1+\epsilon} n^{1+\epsilon},$$

where $\sigma > 0$ is the elasticity of intertemporal substitution and $\epsilon > 0$ is the inverse of the Frisch elasticity. The budget constraint in period t is given, in real terms, by the following:

$$c_{i,t} + s_{i,t} = \frac{R_{t-1}}{\pi_t} s_{i,t-1} + w_{i,t} n_{i,t} + e_{i,t}, \tag{2}$$

where $s_{i,t}$ is the consumer's saving in period t, R_{t-1} is the nominal gross interest rate between $t-1$ and t, $\pi_t \equiv p_t/p_{t-1}$ is the gross inflation rate, p_t is the aggregate price level at t, and $w_{i,t}$ and $e_{i,t}$ are the real wage and the real dividend received by the consumer. The wage and the dividend are allowed to be consumer-specific for reasons that will be explained shortly. We finally denote aggregate consumption by $c_t = \int_{\mathcal{I}_c} c_{i,t} di$, aggregate labor supply by $n_t = \int_{\mathcal{I}_c} n_{i,t} di$, and so on.

Firms. There is a measure-one continuum of ex-ante identical firms, indexed by $j \in \mathcal{I}_f = (1,2]$. Each of these firms is a monopolist that produces a differentiated intermediate-good variety. The output of firm j is denoted by y_t^j, its nominal price is denoted by p_t^j, and its real profit is denoted by e_t^j. The technology is assumed to be linear in labor and productivity is fixed to one, so that

$$y_t^j = l_t^j, \tag{3}$$

where l_t^j is the labor input. These intermediate goods are used by a competitive sector as inputs in the production of the final good. The technology is CES with elasticity $\varsigma > 1$. Aggregate output is thus given by

$$y_t = \left(\int_{\mathcal{I}_f} \left(y_t^j \right)^{\frac{\varsigma-1}{\varsigma}} dj \right)^{\frac{\varsigma}{\varsigma-1}}, \tag{4}$$

The corresponding price index—that is, the nominal price level—is denoted by p_t.

Sticky Prices. Nominal rigidity takes the familiar, Calvo-like, form: in each period, a randomly selected fraction $\theta \in (0,1]$ of the firms must keep prices unchanged, while the rest can reset them.

Monetary Policy. For the time being, we do not need to specify how monetary policy is conducted. To fix ideas, however, it is useful to think of a situation where the agents know what the policy will be in the short run (say, because the ZLB is binding) but face uncertainty about the policy in the more distant future. In this context, the question of interest is how outcomes today vary with the expectations of the future policy.

Rational Expectations and Common Knowledge. Throughout, we impose Rational Expectations Equilibrium (REE). We nevertheless depart from standard practice by removing common knowledge of the state of Nature, accommodating higher-order uncertainty, and preventing the agents from reaching a consensus on the future path of the economy.[7]

[7] By definition, the state of Nature contains the entire profile of the information sets in the population. Furthermore, in any given

In abstract games, such as those studied in Morris and Shin (2002, 2003) or in Section 5 of our paper, lack of common knowledge can be *directly* imposed by endowing each agent with exogenous private information (or different Harsanyi types). In macro and finance applications, however, there is the complication that market signals, such as prices, aggregate information. To preserve the absence of common knowledge, we either have to make sure there is "noise" in the available market signals (Grossman and Stiglitz, 1980), or assume that the observation of all the relevant variables, including the available market signals, is contaminated by idiosyncratic noise due to rational inattention (Sims, 2003).

We follow the first approach in our New Keynesian application in order to make clear that our policy lessons are robust to allowing each consumer to observe her own income, each firm to observe her own demand and supply conditions, and everybody to have common knowledge of the *current* monetary policy and the *current* price level. We nevertheless like the second approach, too, because it bypasses the need for the auxiliary shocks described below and because it allows the re-interpretation of the assumed friction as "costly contemplation" (Tirole, 2015). Section 5 therefore allows for a flexible interpretation of the friction under consideration and Appendix B sketches in more detail how our insights can be recast under the lenses of rational inattention.

The Auxiliary Shocks. Take any period t and let w_t, e_t, and μ_t denote, respectively, the average real wage, the average firm profit, and the average markup in the economy. The real wage and the dividend received by consumer i at t are given by, respectively, $w_{i,t} = w_t\xi_{i,t}$ and $e_{i,t} = e_t\zeta_{i,t}$, where $\xi_{i,t}$ and $\zeta_{i,t}$ are i.i.d. across i and t, independent of one another, and independent of any other random variable in the economy. On the other hand, the real wage paid by firm j is $w_t^j = w_t u_t^j$, and the markup charged by it is $\mu_t^j = \mu_t \nu_t^j$, where u_t^j and ν_t^j are i.i.d. across both j and t, independent of one another, and independent of any other random variable in the economy. One can interpret u_t^j and ν_t^j as idiosyncratic shocks to a firm's marginal cost and her optimal markup, and $\xi_{i,t}$ and $\zeta_{i,t}$ as idiosyncratic shocks to a consumer's labor and financial income.[8] Finally, the aggregate markup shock, μ_t, is also i.i.d. over time and independent of any other variable.

As anticipated, the *sole* modeling role of all these shocks is to "noise up" the information that each agent can extract from the available market signals. We can thus accommodate the desired friction while allowing, in every t, each consumer i to have private knowledge of $(w_{i,t}, e_{i,t})$, each monopolist j to have private knowledge of (w_t^j, μ_t^j), and everybody to have common knowledge of the current interest rate, R_t, of the current prices, $(p_t^j)_{j\in[0,1]}$, and therefore also of the price level.

equilibrium, information sets pin down beliefs of all the endogenous outcomes. It follows that the state of Nature can encode arbitrary news about the future, and that the agents can lack a common belief about the future outcomes only if they lack common knowledge of the state of Nature.

[8] Although we treat these shocks as exogenous, it is not hard to fill in the missing micro-foundations. For instance, firm-specific markup shocks can be micro-founded by allowing for good-specific shocks to the elasticity of the demand faced by each monopolist. Similarly, consumer- and firm-specific wage shocks can be justified by introducing idiosyncratic taste and productivity shocks, letting labor markets be segmented, and allocating the agents in such a way that there is market-specific random variation in the equilibrium wage, translating to idiosyncratic variation in the wage received by a consumer or the wage paid by a firm.

4 The New Keynesian Model as a Pair of Beauty Contests

In this section, we develop our beauty-contest representation of the New Keynesian model. Apart from motivating the class of games studied in Section 5, this representation reveals how the GE effects of that model are related to higher-order beliefs.

To keep the analysis tractable, we work with the log-linearization of the model around a steady state in which inflation is zero and the nominal interest rate equals the natural rate (i.e. $\beta R = 1$). With abuse of notation, we henceforth reinterpret all the variables as the log-deviations from their steady-state counterparts and concentrate on the joint determination of π_t and c_t (or, equivalently, of π_t and y_t).

Proposition 1 (Beauty Contests) *In any equilibrium, and regardless of the level of higher-order uncertainty, the following properties are true. First, aggregate spending satisfies*

$$c_t = -\sigma \left\{ \sum_{k=1}^{+\infty} \beta^{k-1} \bar{E}_t^c[r_{t+k}] \right\} + \tfrac{1-\beta}{\beta} \left\{ \sum_{k=1}^{+\infty} \beta^k \bar{E}_t^c \left[c_{t+k}\right] \right\}, \tag{5}$$

where $\bar{E}_t^c[\cdot]$ denotes the average expectation of the consumers and $r_t \equiv R_{t-1} - \pi_t$ denotes the real interest rate between $t-1$ and t. Second, inflation satisfies

$$\pi_t = \varkappa \left\{ mc_t + \sum_{k=1}^{+\infty} (\beta\theta)^k \, \bar{E}_t^f \left[mc_{t+k}\right] \right\} + \tfrac{1-\theta}{\theta} \left\{ \sum_{k=1}^{+\infty} (\beta\theta)^k \, \bar{E}_t^f \left[\pi_{t+k}\right] \right\} + \varkappa \, \mu_t, \tag{6}$$

where $\bar{E}_t^f[\cdot]$ denotes the average expectation of the firms, mc_t denotes the real marginal cost in period t, and $\varkappa \equiv (1-\theta)\,(1-\beta\theta)\,/\theta$. Finally, $mc_t = (\epsilon + 1/\sigma)y_t$ and $y_t = c_t$.

Condition (5) is a GE variant of the textbook version of the Permanent Income Hypothesis. It follows from log-linearizing the individual consumption functions, aggregating them, and using market clearing in the product, labor, and debt markets. It gives aggregate consumption in any given period as an increasing function of the consumers' expectations of future aggregate income, which in equilibrium coincides with aggregate consumption, and as a decreasing function of their expectations of future real interest rates. Condition (6), on the other hand, follows from the optimal price-resetting behavior of the firms. It gives the rate of inflation as a function of the current real marginal cost and of the firms' expectations of the future marginal costs and the future inflation.

In the rest of this section, we first show that conditions (5) and (6) reduce to, respectively, the representative consumer's Euler condition and the New Keynesian Philips Curve (NKPC) once higher-order uncertainty is assumed away. This clarifies the sense in which our result "opens up" these familiar equations away from the common-knowledge benchmark. We then expand on the three GE mechanisms that are embedded in the model and on their relation to higher-order beliefs.

Relation to the Euler Condition and the NKPC. To begin with, rearrange condition (5) as follows:

$$c_t = -\sigma \bar{E}^c_t[r_{t+1}] + (1-\beta)\bar{E}^c_t[c_{t+1}] + \beta\left\{ -\sigma \sum_{k=1}^{+\infty} \beta^{k-1} \bar{E}^c_t[r_{t+1+k}] + \tfrac{1-\beta}{\beta} \sum_{k=1}^{+\infty} \beta^k \bar{E}^c_t\left[c_{t+1+k}\right] \right\}. \tag{7}$$

Next, compute c_{t+1} by applying condition (5) in period $t+1$, and take the period-t average expectation of it:

$$\bar{E}^c_t[c_{t+1}] = -\sigma \sum_{k=1}^{+\infty} \beta^{k-1} \bar{E}^c_t\left[\bar{E}^c_{t+1}[r_{t+1+k}]\right] + \tfrac{1-\beta}{\beta} \sum_{k=1}^{+\infty} \beta^k \bar{E}^c_t\left[\bar{E}^c_{t+1}[c_{t+1+k}]\right]. \tag{8}$$

Clearly, the term inside the big brackets in condition (7) is the same as the one in the right-hand side of condition (8), except for the difference in the expectation operators. Can we use this observation to recover the Euler equation of a representative consumer?

In the common-knowledge benchmark, the answer is yes. Because, in any period, all agents are assumed to share the same information and the same belief, we have $\bar{E}^c_t[\cdot] = E_t[\cdot]$, where $E_t[\cdot]$ denotes the rational expectation conditional on the common information set, i.e. the information of the representative agent. Using this fact and applying the Law of Iterated Expectations for the representative agent, we infer that, for every $k \geq 1$,

$$\bar{E}^c_t\left[\bar{E}^c_{t+1}\left[r_{t+1+k}\right]\right] = E_t\left[E_{t+1}\left[r_{t+1+k}\right]\right] = E_t\left[r_{t+1+k}\right] = \bar{E}^c_t\left[r_{t+1+k}\right],$$

and similarly $\bar{E}^c_t\left[\bar{E}^c_{t+1}\left[c_{t+1+k}\right]\right] = \bar{E}^c_t\left[c_{t+1+k}\right]$. That is, second-order beliefs coincide with first-order beliefs. It follows that $\bar{E}^c_t[c_{t+1}]$ coincides with the term inside the brackets in condition (7). Therefore, this condition reduces to the familiar, representative-agent Euler condition:

$$c_t = -\sigma\, E_t[r_{t+1}] + E_t[c_{t+1}]. \tag{9}$$

A similar argument establishes that, in the common-knowledge benchmark, condition (6) reduces to

$$\pi_t = \varkappa\, mc_t + \beta\, E_t\left[\pi_{t+1}\right] + \varkappa\, \mu_t, \tag{10}$$

which is the familiar NKPC.[9]

Away from the common-knowledge benchmark, the above argument breaks because the Law of Iterated Expectations no longer holds for the average beliefs. In particular, as it will become clear in the next section, the second-order beliefs of either the exogenous fundamentals or the endogenous outcomes do not coincide with the corresponding first-order beliefs. In the present context, this means that

$$\bar{E}^c_t\left[\bar{E}^c_{t+1}[r_{t+1+k}]\right] \neq \bar{E}^c_t\left[r_{t+1+k}\right] \qquad \text{and} \qquad \bar{E}^c_t\left[\bar{E}^c_{t+1}[c_{t+1+k}]\right] \neq \bar{E}^c_t\left[c_{t+1+k}\right].$$

[9] Using the facts that $mc_t = (\epsilon + 1/\sigma) y_t$ and $c_t = y_t$, the NKPC can be restated as $\pi_t = \kappa y_t + \beta E_t[\pi_{t+1}] + u_t$, where $\kappa \equiv (\epsilon + 1/\sigma)\varkappa$ is the slope of the NKPC with respect to output and $u_t \equiv \varkappa \mu_t$ is the so-called cost-push shock.

It follows that the term in the bracket in condition (7) does not coincide with $\bar{E}_t^c[c_{t+1}]$ and therefore condition (5) does not reduce to the familiar Euler condition. The same logic explains why condition (6) cannot be reduced to the familiar NKPC.

To recap, the key lesson so far is that the familiar representation of the New Keynesian model hinges on the assumption that the agents share the same beliefs, not only about the exogenous impulses, such as news about current or future monetary policy, but also about the future path of aggregate spending and inflation. When this assumption is imposed, one can understand the equilibrium of the economy as the solution to a single-agent decision problem: think of a fictitious agent, who controls the vector (c_t, π_t) and whose optimal behavior is described by the dynamic system (9)-(10). When, instead, this assumption is relaxed, the equilibrium of the economy is better understood as the solution to a complex, multi-layer game among a large number of differentially informed players.[10]

The Economy as a Pair of Dynamic Beauty Contests. Consider first the behavior of the consumers and, momentarily, treat the process for the real interest rate as exogenous. Condition (5) then defines a game in which the players are the consumers, the actions are the consumers' spending levels at different periods, and the payoff-relevant fundamental is the path of real interest rates. This game is a dynamic beauty contest in the sense that each consumer has an incentive to spend more now when she expects the other consumers to spend more in the future. This is because higher aggregate consumption translates to higher income, which in turn justifies more consumption today. Condition (5) therefore isolates the GE mechanism that runs inside the demand block of the model that is known as the Keynesian multiplier.

Consider next the behavior of the firms and, momentarily, treat the process of mc_t as exogenous. Condition (6) then defines a beauty-contest game in which the players are the firms, the actions are the prices, and the relevant fundamental is the average real marginal cost. A dynamic strategic complementary is present in this game because the optimal reset price of a firm depends on her expectation of the future path of the nominal price level, which in turn depends on the decisions of other firms. Condition (6) therefore encapsulates the GE mechanism that runs inside the supply block.

Finally, consider the interaction of the demand and the supply blocks, that is, the interaction of conditions (5) and (6). This interaction can be understood as a meta-game between the consumers and the firms. This meta-game features dynamic strategic complementarity in the following sense: the consumers find it optimal to spend more now when they expect the firms to raise their prices in the future, because high inflation means low real returns to saving; and, symmetrically, the firms find it optimal to raise their prices now when they expect the consumers to spend more in the future, because high aggregate output means high marginal costs. This GE mechanism is known as the inflation-spending feedback.

PE vs GE. Let us momentarily treat the process for the real rate as exogenous and concentrate on the consumption beauty contest defined by condition (5).

Next, fix a $t \geq 0$ and a $k \geq 2$, and ask the following question: how does c_t varies with the period-t beliefs of r_{t+k}, holding constant the beliefs of $r_{t+k'}$ for all $k' \neq k$? This formalizes the question of how much

[10] These points explain the sense in which higher-order uncertainty can be interpreted as frictional coordination.

aggregate spending moves in response to news about future real rates.

From condition (5), we readily see that, if we also hold constant $\bar{E}_t[c_{t+k'}]$ for all $k' \geq 1$, a one-unit reduction in $\bar{E}_t[r_{t+k}]$ raises c_t by $\beta^{k-1}\sigma$. This can be interpreted as a PE effect because it measures the response of the typical consumer to the interest rate faced by himself, holding constant his beliefs about the future behavior of all other consumers, thus also holding constant his expectations of future income.[11] By the same token, the GE effect is tied to how much the typical consumer expects the other consumers to adjust their spending in the future.

From GE to Higher-Order Beliefs. Understanding the adjustment in the expectations of aggregate spending and the associated GE effect is formally equivalent to understanding the hierarchy of beliefs.

For instance, consider $\bar{E}^c_t[c_{t+1}]$, the average expectation of the next-period aggregate spending. From condition (5), c_{t+1} can be expressed as a function of the period-$(t+1)$ *first-order* beliefs of the future real rates and the future aggregate spending. It follows that $\bar{E}^c_t[c_{t+1}]$ is a function of the period-t *second-order* beliefs of the future rates and the future aggregate spending, a fact already shown in condition (8). Using the same logic again and again, we can express the second-order beliefs of future spending—the last term in condition (8)—as functions of third- and higher-order beliefs of the future real rates. We conclude that understanding how the rational expectations of future aggregate spending and income adjust to news about future real rates is formally the same as understanding how the first- and higher-order beliefs of the future rates adjust to the news.

A similar point applies if we focus on the beauty contest defined by condition (6), or if we consider the meta-game between the demand and the supply blocks of the economy: understanding the GE adjustment to any news is equivalent to understanding the associated movement in the relevant higher-order beliefs. This explains where we are heading next: by recasting the GE effects of news about the future in terms of higher-order beliefs, we explain why these effects lose steam once one departs from the familiar but unrealistic common-knowledge benchmark.

5 Dynamic Beauty Contests

This section contains the broader theoretical contribution of our paper. In this section, we side-step the micro-foundations and establish our main insights within a more abstract class of dynamic beauty contests. This class nests not only the two blocks of the New Keynesian model, but also other applications, such as incomplete-information asset-pricing models in the tradition of Singleton (1987). Moreover, we remain agnostic about the origins of higher-order uncertainty. This helps clarify that our lessons do not hinge on whether such uncertainty is the product of dispersed private information or of cognitive constraints along

[11] Strictly speaking, this is the PE effect plus the "within-period" GE effect, namely, the feedback effect between aggregate spending and the *contemporaneous* level of income. In the version of the model we have considered, this within-period GE effect is not attenuated because each consumer knows perfectly her current income and each firm knows perfectly her current marginal cost. Had we relaxed these assumptions, we could have attenuated this effect as well. In any event, this effect is of little interest since it vanishes as the length of the time interval shrinks to zero, which justifies the interpretation of $\beta^{T-1}\sigma$ as the PE effect.

the line of Sims (2003) and Tirole (2015).[12]

5.1 Set up

Time is discrete, indexed by $t \in \{0, 1, ...\}$, and there is a continuum of players, indexed by $i \in [0, 1]$. In each period t, each agent i chooses an action $a_{i,t} \in \mathbb{R}$. We denote the corresponding average action by a_t and let $\Theta_t \in \mathbb{R}$ denote an exogenous fundamental that becomes commonly known in period t.

We specify the best response of player i in period t as follows:

$$a_{i,t} = \Theta_t + \gamma E_{i,t}\left[a_{i,t+1}\right] + \alpha E_{i,t}\left[a_{t+1}\right], \tag{11}$$

where $\alpha, \gamma > 0$ are fixed parameters. Note that a player's optimal action in any given period depends on her expectation of both her own and the aggregate action in the next period. The former effect is parameterized by γ, the latter by α.

By iterating condition (11) forward, aggregating across i, and letting $\bar{E}_t\left[\cdot\right]$ denote the average expectation in period t, we reach the following representation of the beauty contest under consideration:

$$a_t = \Theta_t + \gamma \left\{ \sum_{k=1}^{+\infty} \gamma^{k-1} \bar{E}_t\left[\Theta_{t+k}\right] \right\} + \alpha \left\{ \sum_{k=1}^{+\infty} \gamma^{k-1} \bar{E}_t\left[a_{t+k}\right] \right\} \qquad \forall t \geq 0. \tag{12}$$

This is the key equation we work with in this section. It relates the aggregate outcome in any given period to the concurrent fundamental, the average forecasts of the future fundamentals, and the average forecasts of the future aggregate outcomes. The first two terms capture the direct or PE effect of the current and future fundamentals; the GE effect is captured by the last term, which regards the actions of others. Under this interpretation, α parameterizes the importance of the GE effect.

When all agents share the same information, so that $\bar{E}_t[\cdot] = E_t[\cdot]$ is the rational expectation of a representative agent, we can use the Law of Iterated Expectations to restate condition (12) as follows:

$$a_t = \Theta_t + \delta E_t[a_{t+1}] \qquad \forall t \geq 0, \tag{13}$$

where $\delta \equiv \gamma + \alpha$. The above can be thought of as an aggregate-level Euler condition and δ as the effective discount factor that governs how much current outcomes depend on expectations of future outcomes. This dependence reflects the combination of PE and GE effects, as indicated by the fact that δ is given by the sum of γ and α. Yet, by looking at equation (13) alone, it is impossible to tell these effects apart. This explains why it is important to "open up" this equation to (12), just as we did with the representative agent's Euler condition and the NKPC in the previous section: the distinction between PE and GE effects becomes crucial

[12] By sidestepping the micro-foundations, we not only speak to a larger class of environments, but also abstract from the kind of auxiliary shocks that were necessary before in order to limit the endogenous aggregation of information through markets. This rules out the possibility that agents confuse one kind of fundamental for another (say, the news about future monetary policy with news about the idiosyncratic components of wage and markups), isolates the role of higher-order uncertainty about the fundamental of interest, and clarifies why our contribution is orthogonal to that of Lucas (1972). It also rules out the amplification mechanism studied in Chahrour and Gaballo (2017), which rests on such confusion.

once the common-knowledge assumption is dropped.

Interpretation. Condition (12) directly nests the supply block of our New Keynesian model: just interpret Θ_t as the real marginal cost scaled by $\varkappa$, interpret a_t as inflation, and set $\gamma = \beta\theta$ and $\alpha = \beta(1-\theta)$. Similarly, to nest the demand block, interpret Θ_t as the real interest rate scaled by $-\sigma$, interpret a_t as the aggregate level of spending, and set $\gamma = \beta$ and $\alpha = 1 - \beta$.[13] Moving to a different application, consider the class of incomplete-information asset-pricing models studied in Singleton (1987), Allen, Morris and Shin (2006), Bacchetta and van Wincoop (2006), and Nimark (2017). These models feature a single asset ("stock market") and overlapping generations of differentially informed traders. The equilibrium asset price is shown to satisfy the following condition:

$$p_t = d_t - s_t + \beta \bar{E}_t[p_{t+1}],$$

where p_t, d_t, and s_t denote the period-t price, dividend, and supply, respectively, and $\bar{E}_t[\cdot]$ is the average expectation of the period-t traders. The above can be nested in condition (12) by letting $\Theta_t = d_t - s_t$, $a_t = p_t$, $\alpha = \beta$ and $\gamma = 0$.[14] This indicates, not only the broader applicability of the insights we develop in the rest of this section, but also that the aggregate outcome a_t can be a market-clearing price in a Walrasian setting rather than the average action of a set of players.

The Question of Interest. How does aggregate spending respond to news about future real interest rates? How does inflation respond to news about future real marginal costs? How do asset prices respond to news about future dividends? In the rest of this section, we seek to answer this kind of questions without taking a specific stand on how precise or credible the available news is, or how exactly it has to be modeled. We achieve this by focusing on how the aggregate outcome covaries with the concurrent forecasts of future fundamentals, and by abstracting from the exact source and magnitude of the variation in these forecasts.

More specifically, we fix an arbitrary $T \geq 2$ and isolate the role of news about the period-T fundamental, Θ_T, by shutting down the uncertainty about the fundamentals in any other period. Without any loss, we then let $\Theta_t = 0$ for all $t \neq T$ and treat Θ_T as the *only* random fundamental.[15] We also anchor the beliefs of the aggregate outcome "at infinity" by imposing that $\lim_{k\to\infty} \gamma^k E_{i,t}[a_{t+k}] = 0$ with probability one. This is akin to ruling out infinite-horizon bubbles and can be justified either by letting the game end at *any* finite period $T' > T$, or by imposing that $\gamma < 1$ and that a_t is bounded. We finally state the question of interest as

[13] A minor qualification is needed here. In our version of the New Keynesian model, the consumers do not know the *real* interest rate between today and tomorrow because they have to forecast tomorrow's inflation. It follows that we cannot simply let $\Theta_t = -\sigma r_{t+1}$. That said, we can nest the demand block in condition (12) *as it is* if we assume that prices are completely rigid ($\theta = 1$), fix inflation at zero, and let $\Theta_t = -\sigma R_t$. For the more general case in which inflation is variable, we can either modify the model so that r_{t+1} is observed, or nest the existing version of the demand block in the variant of (12) that replaces Θ_t with $\bar{E}_t[\Theta_t]$. The results developed in the sequel are robust to such modifications.

[14] The restriction $\gamma = 0$ means that the aggregate outcome today depends only on the average expectation of the aggregate outcome tomorrow, as opposed to the entire path of the aggregate outcome in the future. Relative to the more general case we study here, this restriction decreases dramatically the dimensionality of the higher-order beliefs that the aforementioned works had to deal with. The most essential difference, however, between these works and ours is that they focus on the effects of learning whereas we focus on the effects of different horizons.

[15] Setting $\Theta_t = 0$ for all $t \neq T$ is equivalent to isolating the variation in hierarchy of beliefs about the period-T fundamental that is orthogonal to variations in belief hierarchy about the fundamentals in other periods. Our result can thus be read as an "orthogonalization" of the effects of different horizons. Without any loss, we also normalize the unconditional mean of Θ_T to be zero.

follows: how does a_0, the current outcome, covary with $\bar{E}_0[\Theta_T]$, the current average forecast of the future fundamental?

5.2 Complete vs. Incomplete Information

In what follows, we study how the answer to the aforementioned question varies as we move from the frictionless benchmark typically studied in applied work to the more realistic scenario in which the agents are uncertain about one another's beliefs and actions. To fix language, and to rule out trivial cases in which the informational friction relates only to variables that are immaterial for our purposes, we introduce the following definitions.

Definition 1 *We say that information is complete, or that the agents have common knowledge of the available news, if* $E_{i,t}[\Theta_T] = E_{j,t}[\Theta_T]$ *with probability one for all* (i,j,t) *such that* $i \neq j$ *and* $t \leq T-1$. *(And when the converse is true, we say that information is incomplete.)*

Definition 2 *We say that the agents are able to reach a consensus about the future trajectory of the economy if* $E_{i,t}[a_\tau] = E_{j,t}[a_\tau]$ *with probability one for all* (i,j,t,τ) *such that* $i \neq j$ *and* $t < \tau \leq T$.

Note that the first notion regards the beliefs the agents hold about the *exogenous* fundamental, whereas the second notion regards the beliefs the agents form about the *endogenous* outcomes. While conceptually distinct, the two notions are tied together under the rational-expectations hypothesis.

Proposition 2 *Along any rational-expectations equilibrium, lack of consensus about the future outcomes is possible only when information is incomplete.*

This clarifies the modeling role, and our preferred interpretation, of the introduced friction. What is at stake here is not how much the agents know about the fundamentals, but rather the ability to coordinate their beliefs and their responses to the exogenous impulses (the news). Accordingly, the comparisons we develop in the sequel between complete- and incomplete-information economies hold true even if the agents in the incomplete-information economy are *better* informed than their counterparts in the complete-information economy, in the sense of facing less first-order uncertainty about Θ_T.[16]

The Frictionless Benchmark. As a reference point for our subsequent analysis, we first answer the question of interest in the absence of higher-order uncertainty, that is, under complete information.

Lemma 1 *Suppose that information is complete. For all states of Nature,*

$$a_0 = \phi_T^* \cdot E_0[\Theta_T], \tag{14}$$

[16]In line with this point, the obtained lessons are also robust to replacing Θ_t in condition (11) with $\theta_{i,t} = \Theta_t + v_{i,t}$, where $v_{i,t}$ is i.i.d. across agents, and letting each agent i know at $t=0$ the *entire* path of $\theta_{i,t}$; that is, we could have eliminated the uncertainty the typical agent faces about the fundamentals that matter for her *own* decisions (e.g., the uncertainty a consumer faces about her own interest rates), and yet preserve her uncertainty about the beliefs and the decisions of *others* (e.g., the consumer's uncertainty about aggregate spending and inflation).

where $\phi^*_T = \delta^T \equiv (\gamma+\alpha)^T$.

The scalar ϕ^*_T measures how much the aggregate outcome covaries with the average forecast of the fundamental T periods later. This scalar is easily computed by iterating on condition (13), which is the aggregate-level Euler condition of the complete-information economy. Indeed, since $\Theta_t = 0$ for all $t \neq T$, we have $a_T = \Theta_T$ and $a_t = \phi^*_{T-t} E_t[\Theta_T]$ for every $t \leq T-1$, where $\{\phi^*_{T-t}\}_{t=0}^{T-1}$ solves the following recursion:

$$\phi^*_{T-t} = \delta \phi^*_{T-t-1} \quad \forall t \leq T-1, \tag{15}$$

with terminal condition $\phi^*_0 = 1$. It follows that $\phi^*_T = \delta^T$, where, recall, $\delta \equiv \gamma + \alpha$ and where γ and α parameterize, respectively, PE and GE effects.

The exact interpretation of ϕ^*_T and its magnitude depend, of course, on the application under consideration. Consider, for example, the demand block of the New Keynesian model, in isolation of the supply block. In this context, ϕ^*_T measures the response of aggregate spending in period 0 to the concurrent expectation of the real interest rate in period T, holding constant the expectations of the real interest rate in all other periods. In the textbook version of the New Keynesian model, this object is the same regardless of T.[17] In the variant studied by McKay, Nakamura and Steinsson (2016*a*), on the other hand, the corresponding object is decaying with T, due to liquidity constraints.

What is of interest to us, however, is not how ϕ^*_T varies as we move from one complete-information application to another, but rather how it compares to its incomplete-information counterpart. We address this question in the rest of the section. But lets us first note three facts about ϕ^*_T. First, ϕ^*_T is invariant to the precision of the representative agent's information that is available at $t = 0$. Second, ϕ^*_T is the same regardless of whether the representative agent expects to receive additional information between $t = 0$ and $t = T$ or not. Finally, the decomposition between PE and GE effects is irrelevant; all that matters is the combined effect, herein parameterized by $\delta \equiv \gamma + \alpha$.

The Frictional Case. We henceforth focus on the scenario in which the agents face higher-order uncertainty and are therefore unable to reach a perfect consensus about the response of the economy to any given news about Θ_T. This raises the delicate question of whether and how the level of higher-order uncertainty evolves over time. For the time being, we bypass this complication by making the following simplifying assumption.

Assumption 1 *There is no learning between $t = 0$ and $t = T$, at which point Θ_T becomes commonly known.*

As noted a moment ago, learning is inconsequential in the complete-information benchmark. This is no more true when information is incomplete, for reasons discussed later. Nevertheless, Assumption 1 is a useful starting point for three reasons. First, it affords an otherwise general specification of the information structure and yields the sharpest version of our results. Second, it separates our results from prior works that study the implications of learning. Third, it can be relaxed without compromising our main lessons.

[17]This is readily seen by iterating the Euler condition of the representative consumer. It also follows from our earlier observation that the demand block is nested with $\gamma = \beta$ and $\alpha = 1 - \beta$, which means that $\delta = 1$ and hence that ϕ^*_T does not vary with T.

Let us fill in the remaining details. Notwithstanding Assumption 1, we accommodate an otherwise arbitrary information structure. We first let a random variable s encapsulate the realization of the fundamental Θ_T along with *any* other aggregate shock that influences the information and the beliefs of the agents; we refer to s as the underlying state of Nature. We next represent the information of agent i by a random variable ω_i, which is itself correlated with s. Along with the fact that there is no learning, this means that, for all $t \leq T-1$, $E_{i,t}[\cdot] = E_{i,0}[\cdot] = \mathbb{E}[\cdot|\omega_i]$, where $\mathbb{E}[\cdot|\omega_i]$ is the rational expectation conditional on ω_i. We refer to s as the aggregate state of Nature and to ω_i as the signal of agent i. Conditional on s, this signal is an i.i.d. draw from a fixed distribution, whose p.d.f. is given by $\phi(\omega|s)$. This signal therefore encodes the information that agent i has, not only about the fundamental Θ_T, but also about the entire state of Nature and thereby about the beliefs of other agents. In short, ω_i is the Harsanyi type of agent i, that is, an object that governs the agent's expectations of all the endogenous outcomes along any given equilibrium. We finally assume that a law of large number applies in the sense that $\phi(\omega|s)$ is also the cross-sectional distribution of the signals in the population when the aggregate state is s.

Because the state s can be an arbitrary random variable, the specification we have introduced above allows for rich first- and higher-order uncertainty and nests a variety of examples that can be found in the literature. For instance, we may let the aggregate state be $s = (\Theta_T, u)$, where $\Theta_T \sim N(0, \sigma_\theta^2)$ and $u \sim N(0, \sigma_u^2)$ are independent of one another, and specify the signal as $\omega_i = (z, x_i)$, where $z = \Theta_T + u$, $x_i = \Theta_T + v_i$, and $v_i \sim N(0, \sigma_v^2)$ is independent of s and i.i.d. across i. This case nests the information structure assumed in Morris and Shin (2002): each agent receives two signals about Θ_T, a private one given by x_i, and a public one given by z. Alternatively, we may modify the aforementioned case by letting $\omega_i = (z_i, x_i)$, with x_i as before and $z_i = z + \eta_i$, $\eta_i \sim N(0, \sigma_\eta^2)$. In this case, there is a public signal z, whose observation is, however, contaminated by idiosyncratic noise, perhaps due to rational inattention a la Sims (2003). As yet another example, we could let the aggregate state be $s = (\Theta_T, \sigma)$, where $\Theta_T \sim N(0, \sigma_\theta^2)$ and $\sigma \sim U[\underline{\sigma}, \bar{\sigma}]$, and specify the signal as $\omega_i = (x_i, \sigma_i)$, where $x_i = \Theta_T + v_i$, $v_i \sim N(0, \sigma_i^2)$, $\sigma_i \sim U[\sigma - \Delta, \sigma + \Delta]$, and $\Delta, \underline{\sigma}, \bar{\sigma}$ being known scalars such that $0 < \Delta < \underline{\sigma} < \bar{\sigma}$. In this case, some agents are better informed than others, and each agent is uncertain about how informed or uninformed the other agents are.

By adopting this level of generality, we seek, not only to clarify the robustness of our insights, but also to bypass the need for taking a specific stand on what the available signals are and how the higher-order uncertainty is generated. This also explains why most of our results in this section are formulated in terms of how outcomes depend on the belief hierarchy, as opposed to how they depend on a specific set of signals. A concrete example, however, will also be considered.

5.3 Attenuation and the Horizon Effect

An instrumental step towards addressing the question of interest is to understand the role of higher-order beliefs. Since Θ_t is fixed at zero for all $t \neq T$ and Θ_T becomes commonly known at $t = T$, we have $a_t = 0$

for all $t \geq T+1$, $a_T = \Theta_T$, and, for all $t \leq T-1$,

$$a_t = \gamma^{T-t} \bar{E}_t [\Theta_T] + \alpha \sum_{k=1}^{T-t} \gamma^{k-1} \bar{E}_t [a_{t+k}] . \tag{16}$$

By Assumption 1 (no learning), $\bar{E}_t[\cdot] = \bar{E}_0[\cdot]$ for all $t \leq T-1$. Using this fact, and iterating the above condition backwards, we reach the following lemma, which represents the period-0 outcome as a linear function of the concurrent average first- and higher-order expectations of Θ_T.

Lemma 2 *For every (α, γ, T), there exist positive scalars $\{\chi_{h,T}\}_{h=1}^{T}$ such that, regardless of the information structure,*

$$a_0 = \sum_{h=1}^{T} \left\{ \chi_{h,T} \cdot \bar{E}_0^h [\Theta_T] \right\} , \tag{17}$$

where $\bar{E}_0^h[\cdot]$ is defined recursively by $\bar{E}_0^1[\cdot] = \bar{E}_0[\cdot]$ and $\bar{E}_0^h[\cdot] = \bar{E}_0[\bar{E}_0^{h-1}[\cdot]]$ for every $h \geq 2$.

The weights $\chi_{h,T}$ can be constructed recursively, as functions of α, γ, h, and T alone; see Appendix A for details. Here, we focus on the interpretation and the implications of Lemma 2.

Applied to the demand block of the New Keynesian model, this lemma means that today's aggregate spending is determined by the hierarchy of beliefs about future real interest rates. Applied to the supply block, it means that today's inflation is determined by the hierarchy of beliefs about future real marginal costs. In both the textbook version of the New Keynesian model and in the richer DSGE versions used for quantitative policy evaluation, this kind of higher-order beliefs is "swept under the carpet" because the complete-information assumption lets higher-order beliefs collapse to first-order beliefs.

To see this point in the abstract context under consideration, note that when higher-order beliefs coincide with first-order beliefs, condition (17) reduces to $a_0 = \left(\sum_{h=1}^{T} \chi_{h,T} \right) \cdot \bar{E}_0 [\Theta_T]$. Along with Lemma 1, this also means that the complete-information outcome satisfies the following restriction:

$$\phi_T^* = \sum_{h=1}^{T} \chi_{h,T}. \tag{18}$$

By the same token, the weights $\{\chi_{h,T}\}$ do not matter per se: all that matters for equilibrium behavior is their sum, ϕ_T^*.

When, instead, information is incomplete, higher-order beliefs no longer coincide with first-order beliefs. To understand how a_0 covaries with $\bar{E}_0[\Theta_T]$, we must therefore understand two things: first, how higher-order beliefs covary with first-order beliefs; and second, how the beliefs of different order load into a_0, that is, what the structure of the weights $\{\chi_{h,T}\}$ is. We complete these tasks in the sequel—but first we state the ultimate lesson.

Theorem 1 *Suppose that information is incomplete. There exists a positive scalar ϕ_T, which depends on both T and the information structure, such that the following properties hold:*

(i) For all states of Nature,

$$a_0 = \phi_T \cdot \bar{E}_0[\Theta_T] + \epsilon, \tag{19}$$

where ϵ is either identically zero or is random but orthogonal to $\bar{E}_0[\Theta_T]$.

(ii) The ratio ϕ_T/ϕ_T^* is strictly less than 1.

(iii) The ratio ϕ_T/ϕ_T^* is strictly decreasing in T.

Part (i) identifies ϕ_T as the incomplete-information counterpart of ϕ_T^*. Part (ii) establishes that the absence of common knowledge reduces the extent to which the aggregate outcome covaries with the average forecast of the fundamental, regardless of how precise the latter is. We refer to this finding as the "attenuation effect." Part (iii) establishes that this kind of attenuation increases with T: the longer the horizon, the larger the reduction in the responsiveness of the aggregate outcome to the forecasts of the fundamental relative to the frictionless benchmark. We refer to this finding as the "horizon effect."

Let us now sketch the proof of the result. Part (i) is trivial: it follows directly from projecting a_0 on $\bar{E}_0[\Theta_T]$ and letting ϕ_T be the coefficient of this projection and ϵ the residual. The latter captures any variation in higher-order beliefs that is orthogonal to the variation in first-order beliefs, such as the kind of "sentiment shocks" studied in Angeletos and La'O (2013).

To prove parts (ii) and (iii), note first that Lemma 2 implies that the following condition holds regardless of the information structure:

$$\phi_T = \sum_{h=1}^{T} \chi_{h,T}\beta_h, \tag{20}$$

where $\{\chi_{h,T}\}$ are the same scalars as those appearing in Lemma 2 and

$$\beta_h \equiv \frac{Cov\left(\bar{E}_0^h[\Theta_T], \bar{E}_0^1[\Theta_T]\right)}{Var\left(\bar{E}_0^1[\Theta_T]\right)} \tag{21}$$

is the coefficient of the projection of $\bar{E}_0^h\left[\Theta_T\right]$ on $\bar{E}_0^1\left[\Theta_T\right]$, for any $h \geq 1$. When information is complete, $\bar{E}_0^h\left[\Theta_T\right]$ coincides with $\bar{E}_0^1\left[\Theta_T\right]$ for all h and all realizations of uncertainty, which in turn means that β^h is identically 1 for all h. Away from this fragile benchmark, the following is true:

Proposition 3 *With incomplete information, β_h is bounded in $(0,1)$ for all $h \geq 2$ and is strictly decreasing in h.*[18]

In words, higher-order beliefs co-move less with first-order beliefs than lower-order beliefs. This result is easy to establish for the specific information structure studied in Morris and Shin (2002), but requires more work for the more general structure allowed here. It has a similar flavor as the result found in Samet (1998), which essentially states that higher-order beliefs are anchored to the common prior, but is distinct from it, because Samet's result regards only the asymptotic properties of higher-order beliefs as $h \to \infty$ and contains

[18]Obviously, for $h = 1$, we have $\beta_1 = 1$.

no information on whether the comovement between first- and h-order forecasts varies monotonically with h.

Let us now go back to the proof of parts (ii) and (iii) of Theorem 1. Part (ii) now follows directly from the fact that $\beta_h \in (0,1)$, along with conditions (18) and (20). In a nutshell, ϕ_T is less than ϕ_T^* simply because higher-order beliefs move less than one-to-one with first-order beliefs when, and only when, information is incomplete. Importantly, this is true regardless of how much the first-order beliefs themselves co-move with the fundamental. The following is therefore also true, underscoring once again that the introduced friction has to do *only* with the accommodation of higher-order uncertainty and frictional coordination, not with the first-order uncertainty about Θ_T.

Corollary 1 Fix (α, γ, T). *An incomplete-information economy features a lower response to expectations of future fundamentals (i.e., a lower ϕ_T) than a complete-information economy, even if the agents have more precise information about Θ_T in the former than in the latter.*

What remains to prove is part (iii), namely, the property that ϕ_T decreases with T relative to ϕ_T^*. This follows from combining our result that the comovement between first- and higher-order beliefs decreases with the belief order h (Proposition 3), with another result, which we establish next and which sheds light on how the horizon T affects the relative importance of higher-order beliefs in the period-0 outcome.

Theorem 2 Fix (α, γ). *For any (h,T) such that $1 \leq h \leq T$, let $s_{h,T}$ be the total weight on beliefs of order up to, and including, h; that is, $s_{h,T} \equiv \sum_{r=1}^{h} \chi_{r,T}$, where the χs are the same coefficients as those appearing in condition (17). The ratio $s_{h,T}/s_{T,T}$, which measures the relative contribution of the first h orders of beliefs to the aggregate outcome, strictly decreases with the horizon T and converges to 0 as $T \to \infty$.*

This result is crucial (which is why it qualifies, at least in our eyes, to be called a "theorem"). Longer horizons increase the number of loops from future aggregate actions to current actions. But when one increases the number of loops, one is effectively walking down the hierarchy of beliefs: forecasting outcomes further and further into the future maps to forecasting the forecasts of others at higher and higher orders. This in turn explains why longer horizons increase the relative importance of higher-order beliefs. Part (iii) of Theorem 1 then follows from combining this finding with the fact that high-order beliefs themselves covary less with first-order beliefs than do lower-order beliefs.

To sum up: the attenuation effect ($\phi_T/\phi_T^* < 1$) follows merely from the fact that incomplete information dampens the comovement of higher-order beliefs with first-order beliefs ($\beta_h < 1$ for all h). The horizon effect (ϕ_T/ϕ_T^* decreases with T), on the other hand, follows from the fact that this dampening increases with the order of belief (β_h decreases with h), together with the fact that longer horizons raise the relative importance of higher-order beliefs (Theorem 2).

5.4 The Limit as $T \to \infty$.

We complement the preceding results by establishing that, as long as higher-order beliefs are sufficiently anchored in the sense made precise below, ϕ_T becomes *vanishingly* small relative to its complete-information counterpart as the horizon gets larger and larger.

Proposition 4 (Limit) *If* $\lim_{h\to\infty} Var\left(\bar{E}^h[\Theta_T]\right) = 0$, *then*

$$\lim_{T\to\infty} \frac{\phi_T}{\phi_T^*} = 0.$$

To understand this result, note first that, by Theorem 1, the ratio ϕ_T/ϕ_T^* is strictly decreasing in T and bounded in $(0,1)$. It follows that this ratio necessarily converges to a constant $\varphi \in [0,1)$ as $T \to \infty$. In Appendix A, we show that $\varphi = \lim_{h\to\infty} \beta_h$, where β_h is defined as before. This fact is, essentially, a corollary of Theorem 2: in the limit as $T \to \infty$, only the infinite-order beliefs matter. Establishing that ϕ_T/ϕ_T^* converges to zero as $T \to \infty$ is therefore equivalent to establishing that β_h converges to zero as $h \to \infty$. A sufficient (in fact, also a necessary) condition for this to be the case is that $\lim_{h\to\infty} Var\left(\bar{E}^h[\Theta_T]\right) = 0$.

This condition need not hold for every information structure, but can be said to be generic in the following sense. Take, as a reference point, the familiar case in which Θ_T is Normal and each agent observes a noisy private signal and a noisy public signal, as in Morris and Shin (2002). In this case, $\lim_{h\to\infty} \beta^h$ is strictly positive and is pinned down by the precision of the public signal. But now let us perturb this economy by allowing each agent's observation of the public signal to be contaminated by an idiosyncratic noise, perhaps due to rational inattention or some other cognitive limitation. Then, $\lim_{h\to\infty} \beta^h$ is necessarily 0, even if the aforementioned noise is *arbitrarily* small.[19] In short, it takes a tiny perturbation to break the common-knowledge nature of a public signal and to guarantee the limit result in Proposition 4 holds.

As noted in the Introduction, this result illustrates how predictions that depend on long series of forward-looking, general-equilibrium, feedback loops can be particularly fragile to relaxations of common knowledge. It also builds a bridge between our results and the uniqueness result in global games (Morris and Shin, 1998, 2003): the common thread is the disproportionate effect of the (very) high orders of belief and the resulting discontinuity in small perturbations of common knowledge. Finally, it is worth noting that the result holds true even if ϕ_T^* itself explodes to infinity as $T \to \infty$. This case is relevant in the context of forward guidance, which we study in Section 6.

5.5 The Friction as Extra Discounting at the Aggregate level

The preceding analysis established our main results under a general specification of the information structure. We now use a tractable example to recast the friction under consideration as a certain form of myopia.

[19] This case is formalized by letting agent i observe $z_i = z + \eta_i$, where $z = \Theta_T + u$ is the underlying public signal and η_i is the contaminating idiosyncratic noise. Our statement is that $\lim_{h\to\infty} \beta_h = 0$ as long as $Var(\eta_i)$ is not exactly zero.

This sheds further light on how the absence of common knowledge interacts with the underlying GE mechanisms and how it modifies forward-looking behavior at the aggregate level.

For this example, we let $\Theta_T = z + \eta$, where $z \sim N(0, \sigma_z^2)$ and $\eta \sim N(0, \sigma_\eta^2)$ are independent of one another. We interpret z as the component of Θ_T that is realized at $t = 0$ and η as a residual that is realized at $t = T$. Finally, we let the signal of agent i received at $t = 0$ be $\omega_i = z + v_i$, where $v_i \sim N(0, \sigma_v^2)$ is i.i.d across i and orthogonal to both z and η. We can then think of z as news about Θ_T and v_i as an idiosyncratic noise in the observation, or interpretation, of this news by agent i.[20]

Regardless of its interpretation, the key property of this example is that it imposes an exponential structure in the belief hierarchy. By this we mean that, for every $h \geq 2$,

$$\bar{E}_0^h[\Theta_T] = \lambda \bar{E}_0^{h-1}[\Theta_T],$$

where $\lambda \equiv \sigma_v^{-2} / \left(\sigma_v^{-2} + \sigma_z^{-2}\right) \in (0,1)$. This scalar therefore controls the speed with which the comovement between first and h-order beliefs decay with h. A lower λ captures a lower degree of common knowledge; the frictionless, complete-information benchmark is nested in the limit as $\lambda \to 1$.

The following result can then be shown.

Proposition 5 (Discounting) *The elasticity ϕ_T in the incomplete-information economy described above is the same as the elasticity ϕ_T^* in a representative-agent economy in which, for all $t \leq T-2$, the Euler condition (13) holds with $\delta \equiv \gamma + \alpha$ replaced by $\delta' \equiv \gamma + \alpha\lambda$.*

This result illustrates that, under appropriate conditions, the object of interest can be calculated with the same ease as in the complete-information benchmark: all one has to do is to hold the PE effect constant and to discount the GE effect by the factor $\lambda \in (0,1)$. It also offers a sharp formalization of the idea that removing common knowledge causes the economy to act as if the representative agent is myopic and discounts the future more heavily than in the frictionless benchmark (the effective δ is reduced). The extent of this kind of myopia is inversely related to λ, the degree of common knowledge.

The logic behind this result is simple. In the example we consider, all the average higher-order forecasts are linear transformations of the average first-order forecast. This guarantees that, for all $t \in \{1, ..., T-1\}$, the aggregate outcome can be expressed as a multiple of the average first-order forecast: $a_t = \phi_{T-t}\bar{E}_0[\Theta_T]$, for some known scalar ϕ_{T-t}. It follows that $\bar{E}_{t-1}[a_t] = \phi_{T-t}\bar{E}_0^2[\Theta_T]$. And since $\bar{E}_0^2[\Theta_T] = \lambda\bar{E}_0[\Theta_T]$, we conclude that $\bar{E}_{t-1}[a_t] = \lambda a_t$. It is therefore *as if* the average agent systematically underestimates the variation in the future aggregate outcome, which in turn explains the result in Proposition 5. Clearly, the exact result relies on the particular information structure assumed above. The logic, however, applies more generally. As already mentioned, the key feature of the example we consider is that λ controls the speed with which β_h decays with h. The kind of myopia, or discounting, reported above is merely a manifestation of the more general property that β_h decays with h. Theorem 1 and Proposition 5 are therefore mirror images of each

[20] The interpretation of this noise as the product of rational inattention is discussed at Appendix B. For another complementary interpretation, see the discussion of Assumption 3 in Section 6.

other: saying that there is more attenuation at longer horizons is the same as saying that the agents are, effectively, myopic.

Finally, note that, for any $\lambda < 1$, the gap between δ and δ' increases with α, which proves the following.

Corollary 2 *The documented GE attenuation and the associated as-if discounting are larger, the stronger the underlying GE effect.*

This result offers a sharp illustration of how the friction accommodated in this paper interacts with GE mechanisms. It also suggests that the applied lessons delivered in Sections 6 and 7 are likely to be reinforced if one extends the analysis to more realistic versions of the New Keynesian model that strengthen the relevant GE mechanisms by introducing short horizons, borrowing constraints and hand-to-mouth consumers.[21]

5.6 Allowing for Learning

We now return to the role played by Assumption 1. Relaxing this assumption does not appear to invalidate the insights we have developed, but complicates the analysis and precludes us from establishing Theorem 1 for arbitrary information structures. This is due to the increased complexity of the kind of higher-order beliefs that emerge once information is changing between 0 and T.

To appreciate what we mean by this, note that the absence of Assumption 1 guarantees that the following properties hold for any $h \in \{2, ..., T\}$ and any $\{t_2, t_3, ...t_h\}$ such that $0 < t_2 < ... < t_h < T$:

$$\bar{E}_0[\bar{E}_{t_2}[\cdots[\bar{E}_{t_h}[\Theta_T]\cdots]] = \bar{E}_0^h[\cdot].$$

That is, Assumption 1 helps collapse the "cross-period" higher-order beliefs to the "within-period" higher-order beliefs. This is the key step for obtaining the convenient representation of a_0 in Lemma 2 and, thereby, for applying Proposition 3. Without Assumption 1, we must instead express a_0 as a function of all the aforementioned kind of "cross-period" higher-order beliefs, which greatly complicates the analysis. Note, in particular, that there are $T-1$ types of second-order beliefs (namely, $\bar{E}_0\left[\bar{E}_t\left[\cdot\right]\right]$ for all t such that $1 \leq t \leq T-1$), plus $(T-1)\times(T-2)/2$ types of third-order beliefs, plus $(T-1)\times(T-2)\times(T-3)/6$ types of fourth-order beliefs, and so on. What is more, the correlation structure between first- and higher-order beliefs is more intricate, reflecting the anticipation of learning. Intuitively, the first-order belief $\bar{E}_0\left[\Theta_{10}\right]$ can be *less* correlated with the second-order belief $\bar{E}_0\left[\bar{E}_1\left[\Theta_{10}\right]\right]$ than with the third-order belief $\bar{E}_0\left[\bar{E}_8\left[\bar{E}_9[\Theta_{10}]\right]\right]$ if there is little or no learning between periods 0 and 1 but a lot of learning between periods 1 and 8. While this possibility does not invalidate our lessons,[22] it precludes us from extending Theorem 1 to arbitrary forms of learning.

[21]Recall that the demand block of the New Keynesian model is nested with $\gamma = \beta$ and $\alpha = 1-\beta$. The results of Galí, López-Salido and Vallés (2007), Del Negro, Giannoni and Patterson (2015), Werning (2015), Kaplan, Moll and Violante (2016) and McKay, Nakamura and Steinsson (2016*b*,*a*) indicate that the aforementioned frictions map to a lower γ (weaker PE effect) but also a higher α (stronger GE effect). Corollary 2 then implies that these frictions are likely to reinforce our attenuation and horizon effects.

[22]This possibility is indeed allowed in the cases studied in Appendix B, where the theorem continues to apply.

It is possible, however, to overcome this caveat for two leading forms of learning studied in the literature. In the one, we let the agents become gradually aware of Θ_T, as in Mankiw and Reis (2002) and Wiederholt (2015). In the other, we allow the agents to receive a new private signal about Θ_T in each period prior to T, as in Woodford (2003*a*), Nimark (2008), and Mackowiak and Wiederholt (2009). In these two cases, we can not only recover Theorem 1, but also obtain a variant of Proposition 5. See Appendix B for details.

Also note that Theorem 2, which establishes that longer horizons raise the relative importance of higher-order beliefs, follows directly from the best-response structure and is therefore independent of the information structure. This is suggestive of why our horizon effect may extend beyond the aforementioned cases. In line with this idea, we are able to show that the result of Proposition 4, namely the property that ϕ_T becomes vanishingly small relative to ϕ_T^* as $T \to \infty$, extends to *arbitrary* forms of learning as long as the higher-order uncertainty remains bounded way from zero, in a sense made precise in Appendix B.

On the basis of these findings, we conclude that our lessons regarding the interaction of horizons and lack of common knowledge are robust to the introduction of learning. Two additional lessons, however, emerge once we take into account learning.

The first is that, holding the period-0 information constant, ϕ_T is closer to ϕ_T^* in the scenario in which subsequent learning is allowed relative to the scenario in which such learning is ruled out. Intuitively, the anticipation of more (less) information in the future eases (exacerbates) the friction in the present. Note that this effect hinges on the agents being forward-looking and is therefore absent in *static* beauty contests, such as those studied in Morris and Shin (2002), Woodford (2003*a*), and Angeletos and La'O (2010).

The second lesson is that, as time passes and agents accumulate more information, higher-order beliefs converge to first-order beliefs, causing the anchoring of the expectations and the attenuation of GE effects to decay with the lag of time since the news has arrived.[23] Although this prediction is not a core theme of our paper, it is worth noting that it is consistent with the available evidence on the impulse responses of expectations to identified shocks (Coibion and Gorodnichenko, 2012; Vellekoop and Wiederholt, 2017). By contrast, this evidence seems incompatible with the theories developed in Gabaix (2016) and Farhi and Werning (2017): by design, these theories attenuate the response of expectations to shocks but do not allow this attenuation to decay with the lag since the shock has hit the economy.

6 Revisiting Forward Guidance

In this section, we return to the New Keynesian model and study the implications of our insights for the effectiveness of forward guidance.[24]

[23] Formally, consider the setting with learning studied in Appendix B, pick any $\tau \in \{1, ..., T-2\}$, and look at the response of a_t at $t = \tau$ rather than at $t = 0$. It is easy to check that this response increases with the precision of the information that arrives between $t = 0$ and $t = \tau$. A variant of this property is at the core of Woodford (2003*a*), Bacchetta and van Wincoop (2006), Nimark (2008, 2017), and Angeletos and La'O (2010): these papers show how, following persistent shocks to the underlying fundamentals, there is initially a large "wedge" between first- and higher-order beliefs, but this wedge decays as time passes and learning occurs.

[24] Campbell et al. (2012) distinguish two types of forward guidance, "Delphic" and "Odyssean," depending on whether the central bank communication is interpreted partly as a signal of the underlying state of the economy or exclusively as a commitment about future policy. Throughout our analysis, we are concerned with the second kind of forward guidance.

To begin with, suppose either that the monetary authority commits on implementing a specific path for the *real* interest rate, or that prices are infinitely rigid ($\theta = 1$) so that inflation is identically zero and the real interest rate coincides with the nominal one. In either case, we can characterize the response of the economy to news about future rates by studying the demand block alone. The results of the previous section then directly imply that the absence of common knowledge reduces this response, and the more so the further into the future the change in the real rate is. In this sense, the effectiveness of forward guidance is reduced.

The obvious caveat is that this argument treats the path of the *real* interest rate as the exogenous impulse or, equivalently, that it shuts down inflation. While this is useful for understanding how monetary policy controls aggregate demand holding constant inflation expectations, it is inappropriate for understanding how forward guidance matters during a liquidity trap. In that context, it is essential to treat the real interest rate as endogenous and to capture the feedback loop between aggregate spending and inflation.[25]

As noted before, this feedback loop is a GE mechanism that runs *between* the two blocks of the model. This means that the analysis of Section 5, which effectively dealt with each block in isolation, is not directly applicable. To address this complication, let us assume momentarily that information is complete among the firms, even though it is incomplete among the consumers. In this special case, the standard NKPC remains valid, implying that inflation can be expressed as the present value of the future real marginal costs and, thereby, of future aggregate spending. Using this fact to substitute away inflation from condition (5), and letting $y_t = c_t$, we arrive at the following representation of the equilibrium.

Lemma 3 *Suppose the firms have complete information. The equilibrium level of aggregate output (also, spending) satisfies the following condition at every t:*

$$y_t = -\sigma R_t - \sigma \sum_{k=1}^{\infty} \beta^k \bar{E}_t^c[R_{t+k}] + \sum_{k=1}^{\infty} (1 - \beta + k\sigma\kappa)\, \beta^{k-1} \bar{E}_t^c[y_{t+k}], \tag{22}$$

where $\kappa \equiv (\epsilon + 1/\sigma)\varkappa = (\epsilon + 1/\sigma)(1-\theta)(1-\beta\theta)/\theta$.

This can, once again, be understood as a beauty contest among the consumers. But unlike the one seen earlier in condition (5), the one obtained here subsumes the feedback loop between inflation and spending. This in turn explains why the expectations of income show up with different weights from those in condition (5), as well as why these weights depend on κ, which is the slope of the NKPC with respect to aggregate output.

Although the beauty contest obtained in Lemma 3 is not directly nested in the framework of Section 5, all the lessons continue to hold. Proposition 3, which regards the belief hierarchy, remains intact, while Lemma 2 and Theorem 2, which hinge on the dynamic structure, are extended in Appendix C. It follows that Theorem 1 and all the other results of Section 5 can be directly applied to the present context.

[25] Accordingly, for the rest of the analysis, we let $\theta < 1$ (equivalently, $\kappa > 0$).

A Concrete Example. To make things more concrete, and to accommodate the possibility that the firms, too, have incomplete information, we henceforth impose the following assumptions.

Assumption 2 (Monetary Policy) *There exists a known $T \geq 2$ such that:*

(i) At any $t < T$, the nominal interest rate is pegged at zero.[26]

(ii) At any $t > T$, monetary policy replicates flexible-price outcomes.

(iii) The period-T nominal rate is such that

$$R_T = z + \eta, \tag{23}$$

where z and η are random variables, independent of one another and of any other shock in the economy, with $z \sim N(0, \sigma_z^2)$ and $\eta \sim N(0, \sigma_\eta^2)$. The former is realized at $t = 0$; the latter is realized at $t = T$ and is unpredictable prior to that point.

Assumption 3 (Information) *(i) At $t = 0$, each agent i, be it a consumer or a firm, observes a private signal of z, given by*

$$\omega_i = z + v_i,$$

where v_i is an idiosyncratic noise, Normally distributed, with mean zero and variance $\sigma_c^2 \geq 0$ or $\sigma_f^2 \geq 0$, depending on whether the agent is, respectively, a consumer or a firm.

(ii) No other exogenous information arrives till $t = T$, at which point R_T becomes common knowledge.

(iii) The volatilities of the markup and idiosyncratic shocks are arbitrarily large relative to that of z.

Assumption 2 specifies the monetary policy and identifies T as the horizon of forward guidance. Part (i) is motivated by the idea that the ZLB constraint is binding during a liquidity trap. Part (ii), on the other hand, permits us to concentrate on expectations of R_T, the nominal interest rate that the monetary authority will implement in the first period after the ZLB has ceased to bind.[27] Finally, part (iii) splits the randomness of R_T into two orthogonal components, along the lines of the example considered in the end of the last section. The first component, z, can be interpreted as the anticipated component of R_T, or as the news about future monetary policy; such news could be the product of a policy announcement. The second component, η, captures the residual uncertainty, or the unanticipated component, which is revealed at $t = T$.

Assumption 3 turns to the information structure. Part (i) lets each agent's observation of z be contaminated by an idiosyncratic noise, thereby introducing higher-order uncertainty. Part (ii) shuts down any *exogenous* learning. Finally, part (iii) shuts down the *endogenous* learning that obtains through the observation of the realized market outcomes (wages, prices, etc) in every period.

Part (i) of Assumption 3 may appear to be at odds with the fact that central bank communications enjoy ample coverage in the financial press; one may instead be tempted to treat forward guidance as a public

[26]Since R_t is the log-deviation from steady state gross nominal rate, this means $R_t = -\rho$, where ρ is the discount rate defined by $\beta = e^{-\rho}$.

[27]Needless to say, we let the central bank keep the interest rate low for just one period after the ZLB has ceased to bind, as opposed to many periods, only to simplify the exposition.

signal. But it is one thing to say that something is "public news" in the real world and it is a different thing to assume that something is a "public signal" in the theory. Doing the latter requires, not only that every agent herself is aware of and attentive to the news, but also that she is fully confident that every other agent is also aware of and attentive to the news, and so on. Clearly, that kind of common knowledge does not have an obvious counterpart in the real world. What is more, even if one insists on modeling the central bank communications as a perfectly public signal, we can recast the required idiosyncratic noise as idiosyncratic variation in the *interpretation* of such communications.[28]

To sum up, part (i) of Assumption 3 captures the friction we are interested in. Parts (ii) and (iii), on the other hand, are mostly for convenience: they facilitate a sharp characterization of the belief hierarchy.[29] In particular, similarly to the example studied in Section 5, the higher-order beliefs of, respectively, the consumers and the firms satisfy, for all $h \geq 2$,

$$\bar{E}_0^{c,h}[R_T] = \lambda_c^{h-1} \cdot \bar{E}_0^c[R_T] \qquad \text{and} \qquad \bar{E}_0^{f,h}[R_T] = \lambda_f^{h-1} \cdot \bar{E}_0^f[R_T],$$

where

$$\lambda_c \equiv \frac{\sigma_c^{-2}}{\sigma_c^{-2} + \sigma_z^{-2}} \in (0,1] \qquad \text{and} \qquad \lambda_f \equiv \frac{\sigma_f^{-2}}{\sigma_f^{-2} + \sigma_z^{-2}} \in (0,1].$$

Note that λ_c and λ_f control how much higher-order beliefs co-move with first-order beliefs; they therefore parameterize the friction of interest. The frictionless, complete-information benchmark is nested with $\lambda_c = \lambda_f = 1$, and a larger friction corresponds to lower values for the λ's.

We are now ready to study how variation in expectations of R_T (triggered by variation in z) translates into variation in equilibrium output. We start with the common-knowledge benchmark.

Proposition 6 (Forward Guidance with CK) *Suppose $\lambda_c = \lambda_f = 1$, which means that z is common knowledge. There exists a scalar $\phi_T^* > \sigma$ such that, for all realizations of uncertainty,*

$$y_0 = y_0^{trap} - \phi_T^* \cdot E_0[R_T], \tag{24}$$

where y_0^{trap} is the liquidity-trap level of output (i.e., the one obtained when the period-T nominal interest rate is fixed at the steady-state value). Furthermore, ϕ_T^* is strictly increasing in T and $\phi_T^* \to \infty$ as $T \to \infty$.

This result contains the predictions of the textbook New Keynesian model regarding the power of forward guidance during a liquidity trap: this power, as measured by ϕ_T^*, is predicted to increase without bound as the time of action is pushed further and further into the future. It is this prediction, along with its quantitative evaluation, that constitutes the forward-guidance puzzle.[30]

[28] Formally, this can be done by modeling the central bank communication itself as a public signal and by letting each agent have private information about the noise in that signal.

[29] Learning can be accommodated along the lines of Appendix B and was indeed allowed in the first NBER version of our paper, Angeletos and Lian (2016*b*).

[30] It is worth noting that the scalar ϕ_T^* is invariant to the ratio σ_z/σ_η, which parameterizes the precision of the available news:

Contrast this prediction with the PE effect of future interest rates: as evident in condition (5), the PE effect is decreasing in T, simply because of the discounting embedded in intertemporal preferences. The reason that ϕ_T^* exhibits the opposite pattern is because of the GE effects that run within and between the two blocks of the model. In particular, the feedback loop between aggregate spending and income implies that ϕ_T^* would stay constant with T even if we were to shut down the inflation response. The feedback loop between aggregate spending and inflation then explains why ϕ_T^* actually increases with T.

Let us elaborate. Reducing the interest rate at $t = T$ increases spending and causes inflation at $t = T$. Because the nominal interest rate is pegged prior to T, this translates to a low *real* interest rate between $T-1$ and T. This stimulates demand at $T-1$, contributing to even higher inflation at $T-1$, which feeds to even higher demand at $T-2$, and so on. A longer horizon therefore maps to a larger number of iterations in this feedback loop and, thereby, to a stronger cumulative effect at $t = 0$. Finally, as $T \to \infty$, this feedback loop explodes, which explains why ϕ_T^* increases without bound.

The fact that the forward-guidance puzzle is driven by GE effects is already recognized in the literature; see, for example, the discussions in Del Negro, Giannoni and Patterson (2015) and McKay, Nakamura and Steinsson (2016*b*). Our contribution is to explain how these GE effects depend on higher-order beliefs, which in turn explains why these effects are not as potent once one moves away from the common-knowledge benchmark.

Proposition 7 (Forward Guidance without CK) *Suppose that $\lambda_c < 1$ and/or $\lambda_f < 1$, which means that at least one group of agents lacks common knowledge of the news. There exists a scalar ϕ_T such that, for all realizations of uncertainty,*

$$y_0 = y_0^{trap} - \phi_T \cdot \bar{E}_0^c[R_T]. \tag{25}$$

Furthermore, the following properties hold:

(i) ϕ_T is bounded between the PE effect and the complete-information counterpart: $\sigma\beta^T < \phi_T < \phi_T^*$.

(ii) ϕ_T is strictly increasing in both λ_c and λ_f; the ratio ϕ_T/ϕ_T^* is strictly decreasing in T and converges to 0 as $T \to \infty$; finally, when λ_c is sufficiently low, ϕ_T also converges to 0 as $T \to \infty$.

Part (i) formalizes the sense in which the standard model "maximizes" the power of forward guidance: ϕ_T^* is an upper bound to the prediction that the analyst can make if she maintains the rational-expectations hypothesis but is agnostic about the degree of common knowledge in the economy. By varying the degree of common knowledge, we effectively span the range between this upper bound and the underlying PE effect.

Part (ii) is a version of our earlier horizon effect: by attenuating the feedback loop between inflation and spending, as well as the GE effects that run within each block of the model, lack of common knowledge reduces the power of forward guidance by a factor that increases with T. This attenuation increases without bound in the sense that ϕ_T becomes vanishingly small relative to ϕ_T^* as $T \to \infty$, even if the friction is small

varying the ratio σ_z/σ_η affects how much $E_0[R_T]$ varies with R_T, but does not affect how much either y_0 or π_0 vary with $E_0[R_T]$. This property extends to the incomplete-information scenario studied next and explains the sense in which both the puzzle and the resolution we offer are orthogonal to the question of how accurate or credible the news about the future interest rates might be.

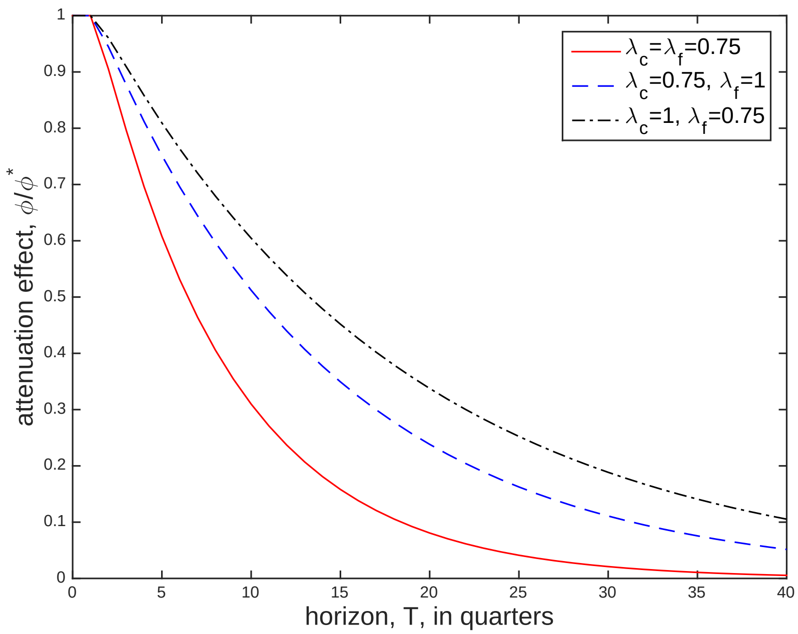

Figure 1: Attenuation effect at different horizons.

(i.e., if λ_c and λ_f are arbitrarily close to 1). Finally, if the friction is large enough, the documented effect can be strong enough that ϕ_T is decreasing in T, not only relative to ϕ_T^*, but also in absolute value.

A Numerical Illustration. We now use a numerical example to illustrate our findings. We interpret the period length as a quarter and adopt the calibration of the textbook New Keynesian model found in Galí (2008). That is, we set the discount factor to 0.99, the Frisch elasticity to 1, the elasticity of intertemporal substitution to 1, and the price revision rate to $1/3$.

What remains is to pick the values of λ_c and λ_f, that is, the departure from common knowledge. The existing literature offers little guidance on how to make this choice. In want of a better alternative, we let $\lambda_c = \lambda_f = 0.75$ and interpret this as a situation in which every agent who has heard the policy announcement believes that any other agent has failed to hear, or "trust," the announcement with a probability equal to 25 percent. This is arguably a modest "grain of doubt" in the minds of people about their ability to coordinate the adjustment in their beliefs and their behavior.[31,32]

The solid line in Figure 1 plots the resulting attenuation effect, as measured by the ratio ϕ_T/ϕ_T^*, against the horizon length, T. By setting $\lambda_c = \lambda_f$, this line assumes that the consumers and the firms are subject to the same informational friction. The dashed line isolates the friction in the consumer side ($\lambda_c = .75$) by shutting it down in the firm side ($\lambda_f = 1$). The dotted line does the converse.

The attenuation is strongest when the friction is present in both sides. Furthermore, the effect is quanti-

[31]The proposed re-interpretation of the informational friction is exact: if we let the signal of an agent be binary, revealing the true z with probability λ, we obtain exactly the same characterization for ϕ_T as the one under the signal structure assumed above.

[32]Although Coibion and Gorodnichenko (2012) provide evidence in support of the kind of informational friction we have accommodated here, that evidence does not directly relate to forward guidance. Yet, taking that evidence at face value could justify an even larger departure from common knowledge: whenever a shock hits the economy, the average forecast error in the expectations of inflation appears to be half as large as the actual response in inflation, which translates to a value for λ close to 0.5.

tatively significant. For example, at a horizon of 5 years ($T = 20$), the power of forward guidance is only *one tenth* of its common-knowledge counterpart. This is on top of any mechanical effect that the noise may have on the size of the shift of the expectations of future interest rates: by construction, the documented attenuation effect is normalized by the size of the variation in the first-order beliefs of R_T.[33]

This illustrates the following broader point: by anchoring the movements in the expectations of economic outcomes *relative* to the movements in the expectations of the policy instrument, our paper helps operationalize the idea that policy makers may have a harder time managing the former kind of expectations than the latter. This may help explain, for example, why forward guidance may trigger a large movement in the term structure without a commensurate movement in expectations of inflation and income.

Appendix C elaborates on the mechanics behind Figure 1. We first show how the documented effects can be understood through the lenses of a discounted Euler condition and a discounted NKPC, along the lines of our earlier, more abstract, result in Proposition 5. This representation helps connect our paper to McKay, Nakamura and Steinsson (2016*b*,*a*) and Gabaix (2016). It also helps explain why almost all of the effect seen in Figure 1 comes from the attenuation of two GE mechanisms: the price-setting complementarity inside the supply block; and the inflation-spending feedback between the two blocks. By contrast, the attenuation of the Keynesian multiplier that runs inside the demand block plays a small role.

Let us explain the last point. In the textbook version of the New Keynesian model, consumers have infinite horizons and their spending depends on aggregate spending *only* through the present value of permanent income. What is more, the discount rate is close to zero. This means that varying the expectations of income in the next, say, 5 years has a small effect on current spending. It follows that the attenuation of this particular GE mechanism cannot possibly be quantitatively significant in the textbook version of the model. But now note that the Keynesian multiplier can be reinforced by short horizons, hand-to-mouth consumers, counter-cyclical precautionary motives, and feedback effects between housing prices and consumer spending. By Corollary 2, we expect this to reinforce also the associated attenuation effect.[34]

The following clarification is also worth making. In our analysis, we have abstracted incomplete markets in the sense of borrowing constraints, but have allowed incomplete markets in the sense of ruling out risk sharing. Had we allowed the consumers to pool their resources at the end of the liquidity trap and to enjoy the same level of consumption thereafter, we would have effectively eliminated the coordination friction among them, thus also abstracting from the attenuation of the GE effect that runs inside the demand block of the model. This is the scenario considered in Wiederholt (2015).

[33] Estimated DSGE models typically assume a higher degree of price stickiness than the one assumed here. This tends to reduce the attenuation effect. For instance, raising the value of θ from $2/3$, the value in Galí (2008), to 0.85, the value in Christiano, Eichenbaum and Rebelo (2011), implies that the ratio ϕ_T/ϕ^*_T increases from about 0.1 to about 0.3 at $T = 20$. That said, note that menu-cost models calibrated to micro data indicate that the "right" value for θ is probably even smaller than 2/3. In any event, our goal here is only to illustrate; a more comprehensive quantitative evaluation is left for future work.

[34] Farhi and Werning (2017) make a similar point with regard to the interaction of borrowing constraints with Level-k Thinking.

7 Revisiting Fiscal Multipliers

We now introduce government spending in the model and study how aggregate income responds to news about government spending in the future. As usual, we treat the level of government spending as exogenous and assume that it is financed with lump-sum taxation. But unlike the standard practice, we remove common knowledge of the future level of government spending and of its macroeconomic effects.

The equilibrium characterization given in Proposition 1 still holds, except that the real marginal cost is now given by $mc_t = \Omega_c c_t + (1 - \Omega_c) g_t$, where g_t is the level of government spending and $\Omega_c \in (0, 1)$ is a constant that increases with the steady-state fraction of GDP absorbed by private consumption. Note then that g_t enters the joint dynamics of c_t and π_t only through the aforementioned formula for mc_t. An exogenous shock to government spending is therefore akin to an exogenous shock in real marginal costs.

Pick now any pair (T, T') such that $2 \leq T \leq T'$. We want T' to capture the length of the liquidity trap and T the period during which a fiscal stimulus is going to be enacted.[35] Accordingly, we let monetary policy satisfy Assumption 2, with T replaced by T'. We also fix $R_{T'} = 0$, so as to focus on fiscal policy. Finally, we assume that $g_t = 0$ for all $t \neq T$ so as to focus on beliefs about g_T, and specify the information structure in the same way as in Assumption 3. We can then show the following.

Proposition 8 *Propositions 6 and 7 hold true with ϕ_T^* and ϕ_T reinterpreted as, respectively, the complete- and the incomplete-information slope of y_0 with respect to $\bar{E}_0[g_T]$.*

The part of this result that regards ϕ_T^* is the fiscal analogue of the forward-guidance puzzle: in the common-knowledge benchmark, the current impact of a fiscal stimulus of a given size increases as the stimulus is pushed further into the future. The underlying logic is essentially the same as before: the PE effect of a fiscal stimulus, which is its direct effect on the concurrent real marginal costs, decays with the horizon T, but the GE effects that run within and between the two blocks of the model overturn this property. The other part of the result, which regards ϕ_T, contains the policy lesson of this section: once we relax the assumption that there is common knowledge of the size and the consequences of the fiscal stimulus, the fiscal multiplier is reduced at every T, and the more so the larger T is.

In short, while the standard model predicts that fiscal stimuli must be back-loaded in order to pile up the feedback effects between aggregate spending and inflation, our approach offers a rationale for front-loading them. Such front-loading eases the coordination friction in the economy: the sooner the fiscal stimulus is enacted, the less concerned the agents would be about the beliefs and the responses of others at long horizons, and, thereby, the easier it is to coordinate on a large response today.[36] An interesting

[35] Similarly to Christiano, Eichenbaum and Rebelo (2011), Woodford (2011), and Werning (2012), the fiscal multipliers we are concerned with are those that obtain when the ZLB constraint binds. The investigation of how our insights extend, when the economy is away from this constraint and monetary policy follows a Taylor rule, is the subject of ongoing work.

[36] Of course, whether the provided rationale for front-loading is sufficiently strong to offset the standard back-loading property depends on the severity of the informational friction and all the parameters that govern the magnitude of the GE feedback loops, such as κ, the slope of the NKPC. Letting λ or κ be small enough guarantees that the fiscal multiplier decreases with T, not only relative to its common-knowledge counterpart, but also in absolute value.

question, which we leave open for future research, is how this insight interacts with borrowing constraints and hand-to-mouth consumers, which is the more standard rationale for the front-loading of fiscal stimuli.

8 Conclusion

Modern macroeconomics assigns a crucial role to forward-looking expectations, such as consumer expectations of future income and future real interest rates or firm expectations of future inflation and future real marginal costs. This property seems desirable and realistic. However, by assuming common knowledge along with rational expectations, the dominant modeling practice hardwires a certain kind of perfection in the ability of economic agents to understand what happens around them, to align their forward-looking beliefs, and to coordinate their responses to any exogenous impulse. In so doing, it also maximizes the general-equilibrium multipliers on fiscal and monetary policies.

Conversely, allowing for higher-order uncertainty helps accommodate a realistic friction in the ability of economic agents to forecast, or comprehend, the macroeconomic effects of policy news. This in turn arrests the underlying general-equilibrium feedback loops and reduces the ability of policy makers to steer the economy, especially when the policy news regards longer horizons.

We first formalized these ideas within an abstract framework, which was flexible enough to nest the demand and the supply block of the New Keynesian model along with other applications. We next showed how these ideas lessen the forward-guidance puzzle and offer a rationale for the front-loading of fiscal stimuli. As shown in Appendix C, the same logic also helps moderate the paradox of flexibility, namely the prediction that, under certain conditions, a higher degree of price flexibility can amplify demand shocks and raise the effectiveness of monetary policy.

By anchoring the movements in the expectations of economic outcomes *relative* to the movements in the expectations of the policy instrument, our paper helps operationalize the idea that policy makers may have a harder time managing the former kind of expectations than the latter. But it also hints to the following possibility: especially when it comes to longer horizons, it may be more important for the policy maker to communicate her intended path for the economy (e.g., for output and inflation) than her intended path for the policy instrument. What is more, these communications must be loud and clear, not only in the ears of Wall Street (financial markets), but also in the ears of Main Street (firms and consumers). We leave the exploration of these ideas open for future research.

Another interesting research question, which we are pursuing in ongoing work, is the investigation of the implications of our insights for equilibrium selection and for the "neo-Fisherian" effects discussed in, inter alia, Garcıa-Schmidt and Woodford (2015) and Cochrane (2016, 2017). As already noted, these issues hinge on the same kind of higher-order beliefs as the ones we have unearthed in this paper.

Finally, it is worth clarifying the following point. In the present paper, we focused on environments in which GE effects reinforce PE effects. In this case, GE attenuation translates to under-reaction. In other contexts, GE effects mitigate PE effects. For example, this is the case when agents compete for limited resources and their actions are strategic substitutes. In this case, our insights remain valid, but their empirical implication is reversed: GE attenuation now translates to *over*-reaction. In a companion paper (Angeletos and Lian, 2016*a*) we discuss how this observation offers a unified explanation to seemingly disparate phenomena.

Appendix A: Proofs

In this Appendix, we prove the results stated in the main text. For all the proofs that regard the New-Keynesian model (as opposed to the abstract analysis in Section 5), we use a tilde over a variable to denote the log-deviation of this variable from its steady-state counterpart, and reserve the non-tilde notation for the original variables. The only exception to this rule is that we let $\tilde{a}_{i,t} \equiv \frac{a_{i,t}}{c^*}$, where $a_{i,t}$ is consumer i's initial asset position at period-t and c^* is steady-state spending. This takes care of the issue that the log-deviation of the asset position is not well defined because the steady-state value is $a^* = 0$ and is standard in the literature (e.g., Woodford, 2011).

Proof of Proposition 1. We proceed in four steps, starting with the behavior of the consumers, proceeding with the behavior of the firms, and concluding with market clearing and with the derivation of the two beauty contests shown in the main text.

Step 1: Consumers. Consider an arbitrary consumer $i \in \mathcal{I}_c$. Let $a_{i,t} = R_{t-1}s_{i,t-1}/\pi_t$ denote consumer i's initial asset position at period-t. By condition (2), the following intertemporal budget constraint holds in all periods and all states of Nature:[37]

$$\sum_{k=0}^{+\infty} \left\{ \left[\prod_{j=1}^{k} \left(\frac{R_{t+j-1}}{\pi_{t+j}} \right)^{-1} \right] c_{i,t+k} \right\} = a_{i,t} + \sum_{k=0}^{+\infty} \left\{ \left[\prod_{j=1}^{k} \left(\frac{R_{t+j-1}}{\pi_{t+j}} \right)^{-1} \right] (w_{i,t+k}n_{i,t+k} + e_{i,t+k}) \right\} . \tag{26}$$

Taking the log-linear approximation of the above around the steady state, we get the following:

$$\sum_{k=0}^{+\infty} \beta^k \tilde{c}_{i,t+k} = \tilde{a}_{i,t} + \sum_{k=0}^{+\infty} \beta^k \left\{ \Omega \left(\tilde{w}_{i,t+k} + \tilde{n}_{i,t+k} \right) + (1-\Omega)\, \tilde{e}_{i,t+k} \right\}, \tag{27}$$

where Ω is the ratio of labor income to total income in steady state. The consumer's optimality conditions, on the other hand, can be expressed as follows:

$$\tilde{n}_{i,t} = \frac{1}{\epsilon} \left(\tilde{w}_{i,t} - \frac{1}{\sigma} \tilde{c}_{i,t} \right), \tag{28}$$

$$\tilde{c}_{i,t} = E_{i,t} \left[\tilde{c}_{i,t+1} - \sigma \left(\tilde{R}_t - \tilde{\pi}_{t+1} \right) \right] = E_{i,t} \left[\tilde{c}_{i,t+1} - \sigma \tilde{r}_{t+1} \right], \tag{29}$$

where $E_{i,t}[\cdot]$ is the expectation of consumer i in period t. The first condition describes optimal labor supply; the second is the *individual-level* Euler condition, which describes optimal consumption and saving.

At this point, it is worth emphasizing that our analysis preserves the *standard* Euler condition at the individual level. This contrasts with McKay, Nakamura and Steinsson (2016*b*,*a*) and Werning (2015), where liquidity constraints cause this condition to be violated for *some* agents, as well as with Gabaix (2016), where a cognitive friction causes this condition to be violated for *every* agent. We revisit this point in Appendix C,

[37] One should think of the state of Nature as a realization of the exogenous payoff relevant shocks along with the cross-sectional distribution of the exogenous signals (information) received by the agents.

when we show that our analysis rationalizes a *discounted* Euler condition at the aggregate level, *in spite of* the preservation of the standard condition at the individual level.

Combining conditions (27), (28) and (29), we obtain the optimal expenditure of consumer i in period t as a function of the current and the expected future values of wages, dividends, and real interest rates:

$$\tilde{c}_{i,t} = \tfrac{(1-\beta)\epsilon\sigma}{\epsilon\sigma+\Omega}\tilde{a}_{i,t} - \sigma\sum_{k=1}^{+\infty}\beta^k E_{i,t}\left[\tilde{r}_{t+k}\right] \\ + (1-\beta)\left[\tfrac{(\epsilon+1)\sigma\Omega}{\epsilon\sigma+\Omega}\tilde{w}_{i,t} + \tfrac{\epsilon\sigma(1-\Omega)}{\epsilon\sigma+\Omega}\tilde{e}_{i,t}\right] + (1-\beta)\sum_{k=1}^{+\infty}\beta^k E_{i,t}\left[\tfrac{(\epsilon+1)\sigma\Omega}{\epsilon\sigma+\Omega}\tilde{w}_{i,t+k} + \tfrac{\epsilon\sigma(1-\Omega)}{\epsilon\sigma+\Omega}\tilde{e}_{i,t+k}\right]. \tag{30}$$

This condition, which is a variant of the consumption function seen in textbook treatments of the Permanent Income Hypothesis,[38] contains two elementary insights. First, all future variables—wages, dividends, and real interest rates—are discounted. Second, the *current* spending of a consumer depends on the present value of her income, which in turn depends, in equilibrium, on the *future* spending of other consumers.

The first property guarantees that the decision-theoretic, or partial-equilibrium, effect of forward guidance diminishes with the horizon at which interest rates are changed; the second represents a dynamic strategic complementarity, which is the modern reincarnation what was known as the "income multiplier" in the IS-LM framework. We elaborate on these two points more in the main text. For the time being, we aggregate condition (30), and use the facts that assets average to zero and that future idiosyncratic shocks are unpredictable, to obtain the following condition for aggregate spending:

$$\tilde{c}_t = -\sigma\sum_{k=1}^{+\infty}\beta^k \bar{E}_t^c\left[\tilde{r}_{t+k}\right] + (1-\beta)\left[\tfrac{(\epsilon+1)\sigma\Omega}{\epsilon\sigma+\Omega}\tilde{w}_t + \tfrac{\epsilon\sigma(1-\Omega)}{\epsilon\sigma+\Omega}\tilde{e}_t\right] \\ + (1-\beta)\sum_{k=1}^{+\infty}\beta^k \bar{E}_t^c\left[\tfrac{(\epsilon+1)\sigma\Omega}{\epsilon\sigma+\Omega}\tilde{w}_{t+k} + \tfrac{\epsilon\sigma(1-\Omega)}{\epsilon\sigma+\Omega}\tilde{e}_{t+k}\right], \tag{31}$$

where $\bar{E}_t^c[\cdot]$ henceforth denotes the *average* expectation of the consumers in period t.

Step 2: Firms. Consider a firm $j \in \mathcal{I}_f$ that gets the chance to reset its price during period t. The optimal reset price, denoted by p_t^{j*}, is given by the following:

$$\tilde{p}_t^{j*} = (1-\beta\theta)\left\{(\tilde{mc}_t^j + \tilde{p}_t) + \sum_{k=1}^{+\infty}(\beta\theta)^k E_{j,t}\left[\tilde{mc}_{t+k}^j + \tilde{p}_{t+k}\right]\right\} + (1-\beta\theta)\,\tilde{\mu}_t^j, \tag{32}$$

where $E_{j,t}^f[\cdot]$ denotes the firm's expectations in period t, $\tilde{mc}_t^j = \tilde{w}_t^j$ is its *real* marginal cost in period t, and $\tilde{\mu}_t^j$ is the corresponding markup shock. The interpretation of this condition is familiar: the optimal "reset"

[38] To see this more clearly, suppose that initial assets are zero, that the real interest rate is expected to equal the discount rate at all periods, and that labor supply is fixed ($\epsilon \to \infty$). Condition (30) then reduces to $\tilde{c}_{i,t} = (1-\beta)\left[\Omega\tilde{w}_{i,t} + (1-\Omega)\tilde{e}_{i,t}\right] + (1-\beta)\sum_{k=1}^{+\infty}\beta^k E_{i,t}\left[\Omega\tilde{w}_{i,t+k} + (1-\Omega)\tilde{e}_{i,t+k}\right]$, which means that optimal consumption equals "permanent income" (the annuity value of current and future income). Relative to this benchmark, condition (30) adjusts for three factors: for the endogeneity of labor supply, which explains the different weights on wages and dividends; for initial assets, which explains the first term in condition (30); and for the potential gap between the real interest rate and the subjective discount rate, which explains the second term.

price is given by the expected nominal marginal cost over the expected lifespan of the new price, plus the markup.[39] Aggregating the above condition, using the fact that the past price level is known and that inflation is given by $\tilde{\pi}_t = (1-\theta)(\tilde{p}_t^* - \tilde{p}_{t-1})$, where $\tilde{p}_t^* \equiv \int_{\mathcal{I}_f} \tilde{p}_t^{j*} dj$, we obtain the following condition for the level of inflation in period t:

$$\tilde{\pi}_t = \varkappa \tilde{mc}_t + \varkappa \sum_{k=1}^{+\infty} (\beta\theta)^k \bar{E}_t^f [\tilde{mc}_{t+k}] + \tfrac{1-\theta}{\theta} \sum_{k=1}^{+\infty} (\beta\theta)^k \bar{E}_t^f [\tilde{\pi}_{t+k}] + \varkappa \tilde{\mu}_t, \tag{33}$$

where $\varkappa \equiv \frac{(1-\theta)(1-\beta\theta)}{\theta}$ and $\bar{E}_t^f[\cdot]$ henceforth denotes the average expectation of the firms. The latter may or may not be the same as the average expectation of the consumers.

Step 3: Market Clearing, Wages, and Profits. Because the final-good sector is competitive and observes all the relevant prices,[40] and because the technology satisfies (3) and (4), we have that $\tilde{p}_t = \int_{\mathcal{I}_f} \tilde{p}_t^j dj$ and $\tilde{y}_t = \int_{\mathcal{I}_f} \tilde{y}_t^j dj = \int_{\mathcal{I}_f} \tilde{l}_t^j dj$. The latter, together with market clearing in the labor market, gives $\tilde{y}_t = \tilde{n}_t \equiv \int_{\mathcal{I}_c} \tilde{n}_{i,t} di$. Market clearing in the market for the final good, on the other hand, gives

$$\tilde{y}_t = \tilde{c}_t \equiv \int_{\mathcal{I}_c} \tilde{c}_{i,t} di.$$

Finally, note that the real profit of monopolist j at period t is given by $e_t^j = \left(\frac{p_t^j}{p_t} - w_t^j\right) y_t^j$. Log-linearizing and aggregating it gives $\tilde{e}_t = -\frac{\Omega}{1-\Omega}\tilde{w}_t + \tilde{y}_t$. Combining all these facts with (28), the optimality condition for labor supply, we arrive at the following characterization of the aggregate wages and the profits:

$$\tilde{w}_t = \tilde{mc}_t = \left(\epsilon + \frac{1}{\sigma}\right)\tilde{y}_t, \qquad \tilde{e}_t = \left[1 - \frac{\Omega\left(\epsilon + \frac{1}{\sigma}\right)}{1-\Omega}\right]\tilde{y}_t, \quad \text{and} \quad \tfrac{(\epsilon+1)\sigma\Omega}{\epsilon\sigma+\Omega}\tilde{w}_t + \tfrac{\epsilon\sigma(1-\Omega)}{\epsilon\sigma+\Omega}\tilde{e}_t = \tilde{y}_t. \tag{34}$$

Step 4: Beauty Contests. Condition (31), which follows merely from consumer optimality, pins down aggregate spending as a function of the average beliefs of wages, profits, interest rates, and inflation. As we impose REE, a consumer can infer that (34) holds, aggregate spending can then be expressed as a function of the consumers' average beliefs of interest rates, of inflation, and of aggregate spending itself. This is condition (5), the consumption beauty contest. Similarly, combining (33) and (34), we can express aggregate inflation as a function of the firms' average beliefs of aggregate spending and of inflation itself. This is condition (6), the inflation beauty contest.

Proof of Proposition 2. Because Θ_t is zero for all $t > T$, a_t is also zero for all $t > T$.[41] Using this fact along with the fact that Θ_t is zero also for $t < T$, and iterating on condition (12), we can obtain a_t for all $t < T$ as a linear function of the average first- and higher-order beliefs about Θ_T; see, e.g., Lemma 2 below for an explicit characterization in the case without learning. When information is complete, all agents share the same first-order beliefs about Θ_T with probability one, and this fact is itself common knowledge. It

[39] Note that future markups are unpredictable.
[40] Recall that the we have allowed the entire price vector, $(p_t^j)_{j\in[0,1]}$, to be common knowledge at period t.
[41] As mentioned in main text, we assume $\lim_{k\to\infty}\gamma^k E_{i,t}[a_{t+k}] = 0$ and rule out "extrinsic bubbles."

follows that higher-order beliefs collapse to first-order beliefs and, therefore, a_t becomes a linear function of $E_t[\Theta_T]$, the commonly shared expectation of Θ_T. Now take any $t < \tau \leq T$ and any pair of agents i, j. Complete information guarantees that $E_{i,t}[E_\tau[\Theta_T]] = E_t[\Theta_T] = E_{j,t}[E_\tau[\Theta_T]]$ with probability one. And since we already argued that, in equilibrium, a_τ is a known linear function of $E_\tau[\Theta_T]$, it is also the case $E_{i,t}[a_\tau] = E_{j,t}[a_\tau]$. That is, complete information (in the sense of Definition 1) rules out imperfect consensus (in the sense of Definition 2).

Proof of Lemma 1. Lemma 1 directly follows from the argument in main text.

Proof of Lemma 2. We prove the following stronger result: there exists positively-valued coefficients $\{\chi_{h,k}\}_{k\geq 1, 1\leq h\leq k}$, such that, for any $t \leq T-1$,

$$a_t = \sum_{h=1}^{T-t} \left\{ \chi_{h,T-t} \bar{E}_t^h \left[\Theta_T\right] \right\}, \tag{35}$$

where each $\chi_{h,k}$ is a function of (α, γ, h, k) and $\bar{E}_t^h[\cdot]$ is defined recursively by $\bar{E}_t^1[\cdot] = \bar{E}_t[\cdot]$ and $\bar{E}_t^h[\cdot] = \bar{E}_t\left[\bar{E}_t^{h-1}[\cdot]\right]$ for every $h \geq 2$. We now prove this claim by induction. First, consider $t = T-1$. From $a_T = \Theta_T$[42] and condition (16), we have $a_{T-1} = (\gamma + \alpha)\bar{E}_{T-1}[\Theta_T]$. It follows that condition (35) holds for

$$\chi_{1,1} = \gamma + \alpha.$$

Now, pick an arbitrary $t \leq T-2$, assume that condition (35) holds for all $\tau \in \{t+1, ..., T-1\}$, and let us prove that it also holds for t. From condition (16), we have

$$\begin{aligned} a_t &= \gamma^{T-t-1}(\gamma+\alpha)\bar{E}_t[\Theta_T] + \alpha \sum_{k=1}^{T-t-1} \gamma^{k-1} \bar{E}_t\left[\sum_{h=1}^{T-t-k} \left\{\chi_{h,T-t-k} \bar{E}_{t+k}^h [\Theta_T]\right\}\right] \\ &= \gamma^{T-t-1}(\gamma+\alpha)\bar{E}_t[\Theta_T] + \sum_{h=1}^{T-t-1}\sum_{k=1}^{T-t-h} \left(\alpha\gamma^{k-1}\chi_{h,T-t-k}\right) \bar{E}_t^{h+1}[\Theta_T], \end{aligned} \tag{36}$$

where the second line uses Assumption 1 (no learning). As a result, condition (35) holds for

$$\chi_{1,T-t} = \gamma^{T-t-1}(\gamma+\alpha) \quad \text{and} \quad \chi_{h+1,T-t} = \sum_{k=1}^{T-t-h} \alpha\gamma^{k-1}\chi_{h,T-t-k} \;\; h \in \{1, \cdots T-t-1\}. \tag{37}$$

This finishes the proof.

Proof of Theorem 1. This theorem builds on Proposition 3 and Theorem 2, which are proved in the sequel. We invite the reader to read first the proofs of these two results. Here, we prove Theorem 1 taking for granted these results.

Part (i) follows directly from projecting a_0 on $\bar{E}_0[\Theta_T]$ and letting ϕ_T be the coefficient of this projection and ϵ the residual.

[42] As mentioned in main text, we assume $\lim_{k\to\infty} \gamma^k E_{i,t}[a_{t+k}] = 0$ and rule out "extrinsic bubbles." Together with the fact Θ_t is zero for all $t > T$, a_t is also zero for all $t > T$. As a result, $a_T = \Theta_T$ from condition (12).

To prove part (ii), note that from Lemma 2, we have $\phi_T = \sum_{h=1}^{T} \chi_{h,T}\beta_h$, which is condition (20) in the main text. Together with the expression of ϕ_T^*, condition (18), and the fact that $\beta_h < 1$ for all $h \geq 2$ (from Proposition 3), we have $\phi_T/\phi_T^* < 1$ for all $T \geq 2$.

To prove part (iii), from condition (20), we have $\phi_T/\phi_T^* = \left[\sum_{h=1}^{T-1} s_{h,T}\left(\beta_h - \beta_{h+1}\right) + s_{T,T}\beta_T\right]/s_{T,T}$ and $\phi_{T+1}/\phi_{T+1}^* = \left[\sum_{h=1}^{T-1} s_{h,T+1}\left(\beta_h - \beta_{h+1}\right) + s_{T,T+1}\left(\beta_T - \beta_{T+1}\right) + s_{T+1,T+1}\beta_{T+1}\right]/s_{T+1,T+1}$. From Proposition 3 we know $\beta_h > \beta_{h+1}$ for all h. Together with Theorem 2, we have, for all $T \geq 1$,

$$\begin{aligned}
\phi_{T+1}/\phi_{T+1}^* &< \left[\sum_{h=1}^{T-1} s_{h,T+1}\left(\beta_h - \beta_{h+1}\right) + s_{T+1,T+1}\left(\beta_T - \beta_{T+1}\right) + s_{T+1,T+1}\beta_{T+1}\right]/s_{T+1,T+1} \\
&= \left[\sum_{h=1}^{T-1} s_{h,T+1}\left(\beta_h - \beta_{h+1}\right) + s_{T+1,T+1}\beta_T\right]/s_{T+1,T+1} \\
&\leq \left[\sum_{h=1}^{T-1} s_{h,T}\left(\beta_h - \beta_{h+1}\right) + s_{T,T}\beta_T\right]/s_{T,T} = \phi_T/\phi_T^*.
\end{aligned}$$

Proof of Proposition 3. Note that every agent i's information set at period 0 is drawn i.i.d. from the aggregate state of the Nature (for simplicity, we call this property "symmetry" in the rest of this proof), we have, for all $h \geq 2$,

$$\begin{aligned}
Cov\left(\bar{E}_0^h[\Theta_T], \bar{E}_0^1[\Theta_T]\right) &= Cov\left(E_{i,0}\left[\bar{E}_0^{h-1}[\Theta_T]\right], \bar{E}_0^1[\Theta_T]\right) = Cov\left(E_{i,0}\left[\bar{E}_0^{h-1}[\Theta_T]\right], E_{i,0}\left[\bar{E}_0^1[\Theta_T]\right]\right), \\
&= Cov\left(\bar{E}_0^{h-1}[\Theta_T], E_{i,0}\left[\bar{E}_0^1[\Theta_T]\right]\right) = Cov\left(\bar{E}_0^{h-1}[\Theta_T], \bar{E}_0^2[\Theta_T]\right),
\end{aligned}$$

where the second and the third equality come from the law of iterated expectations. By the same argument, we have, for all $h \geq 2$ and $j \in \{1, 2, \cdots h-1\}$,

$$Cov\left(\bar{E}_0^h[\Theta_T], \bar{E}_0^1[\Theta_T]\right) = Cov\left(\bar{E}_0^{h-j}[\Theta_T], \bar{E}_0^{1+j}[\Theta_T]\right). \tag{38}$$

From the previous condition, for $k \geq 1$, we have[43]

$$\beta_{2k} = \frac{Cov\left(\bar{E}_0^k[\Theta_T], \bar{E}_0^{k+1}[\Theta_T]\right)}{Var\left(\bar{E}_0^1[\Theta_T]\right)} = \frac{Cov\left(\bar{E}_0^k[\Theta_T], E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right]\right)}{Var\left(\bar{E}_0^1[\Theta_T]\right)} = \frac{Var\left(E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right]\right)}{Var\left(\bar{E}_0^1[\Theta_T]\right)} \geq 0 \;\; \forall i, \tag{39}$$

where the second equation follows from symmetry and the last equation follows from the law of iterated expectations. Similarly, for $k \geq 1$, we have

$$\beta_{2k-1} = \frac{Cov\left(\bar{E}_0^k[\Theta_T], \bar{E}_0^k[\Theta_T]\right)}{Var\left(\bar{E}_0^1[\Theta_T]\right)} = \frac{Var\left(\bar{E}_0^k[\Theta_T]\right)}{Var\left(\bar{E}_0^1[\Theta_T]\right)} \geq 0. \tag{40}$$

[43] Note that under incomplete information, we have $Var\left(\bar{E}_0^1[\Theta_T]\right) > 0$, so all β_k is well defined. To prove it, note that if $Var\left(\bar{E}_0^1[\Theta_T]\right) = 0$, together with the fact that the mean of Θ_T is zero, we have $\bar{E}_0^1[\Theta_T] = 0$ almost surely. As a result, we have $Var\left(E_{i,0}\left[\Theta_T\right]\right) = Cov\left(\Theta_T, E_{i,0}\left[\Theta_T\right]\right) = Cov\left(\Theta_T, \bar{E}_0\left[\Theta_T\right]\right) = 0$, and $E_{i,0}\left[\Theta_T\right] = 0 = E_{j,0}\left[\Theta_T\right]$ almost surely for all i, j. This is inconsistent with the definition about incomplete information in Definition 1.

Now, note that for any random variable X, and any information set I, according to the law of total variance, we have:

$$Var\left(\mathbb{E}[X|I]\right) \leq Var\left(X\right).$$

As a result, $Var\left(\bar{E}_0^{k+1}[\Theta_T]\right) = Var\left(E\left[E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right]|s\right]\right) \leq Var\left(E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right]\right)$ and $Var\left(E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right]\right) = Var\left(E\left[\bar{E}_0^k[\Theta_T]|\omega_i\right]\right) \leq Var\left(\bar{E}_0^k[\Theta_T]\right)$, where, as a reminder, s is the aggregate state of the Nature and ω_i is the information set of agent i. Together with conditions (39) and (40), we know that, for all $k \geq 1$, $\beta_{2k+1} \leq \beta_{2k} \leq \beta_{2k-1}$. This proves that, for all $h \geq 2$, $\beta_h \in [0,1]$ and is weakly decreasing in h.

Now we try to prove β_h is strictly decreasing in h. Note that from condition (39), $\beta_2 = \frac{Var\left(E_{i,0}\left[\bar{E}_0[\Theta_T]\right]\right)}{Var\left(\bar{E}_0^1[\Theta_T]\right)} \leq \beta_1 = 1$. If $\beta_2 = \beta_1 = 1$, we have $Var\left(E_{i,0}\left[\bar{E}_0[\Theta_T]\right]\right) = Var\left(\bar{E}_0[\Theta_T]\right)$ for all i. This means that

$$\begin{aligned} Var\left(E_{i,0}\left[\bar{E}_0[\Theta_T]\right] - \bar{E}_0[\Theta_T]\right) &= Var\left(E_{i,0}\left[\bar{E}_0[\Theta_T]\right]\right) + Var\left(\bar{E}_0[\Theta_T]\right) - 2Cov\left(E_{i,0}\left[\bar{E}_0[\Theta_T]\right], \bar{E}_0[\Theta_T]\right) \\ &= Var\left(\bar{E}_0[\Theta_T]\right) - Var\left(E_{i,0}\left[\bar{E}_0[\Theta_T]\right]\right) = 0, \end{aligned} \tag{41}$$

where the second equality follows from the law of iterated expectations. As a result, for all i, $E_{i,0}\left[\bar{E}_0[\Theta_T]\right] = \bar{E}_0[\Theta_T]$ almost surely. We henceforth have that

$$\begin{aligned} Var\left(\bar{E}_0\left[\Theta_T\right]\right) &= Cov\left(\Theta_T, \bar{E}_0^2\left[\Theta_T\right]\right) = Cov\left(\Theta_T, E_{i,0}\left[\bar{E}_0\left[\Theta_T\right]\right]\right) \\ &= Cov\left(\Theta_T, \bar{E}_0\left[\Theta_T\right]\right) = Cov\left(\Theta_T, E_{i,0}\left[\Theta_T\right]\right) = Var\left(E_{i,0}\left[\Theta_T\right]\right), \end{aligned}$$

where the first equality follows a similar argument as condition (38), the second and fourth equalities follow from symmetry, the third equality follows from $E_{i,0}\left[\bar{E}_0[\Theta_T]\right] = \bar{E}_0[\Theta_T]$ almost surely, and the last equality follows from the law of iterated expectations. This means that

$$\begin{aligned} Var\left(E_{i,0}\left[\Theta_T\right] - \bar{E}_0\left[\Theta_T\right]\right) &= Var\left(E_{i,0}\left[\Theta_T\right]\right) + Var\left(\bar{E}_0\left[\Theta_T\right]\right) - Cov\left(E_{i,0}\left[\Theta_T\right], \bar{E}_0\left[\Theta_T\right]\right) \\ &= Var\left(E_{i,0}\left[\Theta_T\right]\right) - Var\left(\bar{E}_0\left[\Theta_T\right]\right) = 0, \end{aligned} \tag{42}$$

where the second equality follows from symmetry. As a result, $E_{i,0}\left[\Theta_T\right] = \bar{E}_0\left[\Theta_T\right]$ almost surely for all i, and $E_{i,0}\left[\Theta_T\right] = E_{j,0}\left[\Theta_T\right]$ almost surely for all i, j. This is contradictory to the definition of incomplete information. As a result, $\beta_2 < \beta_1 = 1$.

Now, suppose it is not the case that β_h is strictly decreasing in h. Then there exists a smallest $h^* > 1$ such that $\beta_{h^*+1} = \beta_{h^*}$.

If $h^* = 2k$ for some $k \geq 1$. From conditions (39) and (40), we have $Var\left(E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right]\right) = Var\left(\bar{E}_0^{k+1}[\Theta_T]\right)$. Following a similar argument as condition (42), we have $E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right] = \bar{E}_0^{k+1}[\Theta_T]$ almost surely. We henceforth have

$$\begin{aligned} Cov\left(\bar{E}_0^k[\Theta_T], \bar{E}_0^{k+1}[\Theta_T]\right) &= Cov\left(E_{i,0}\left[\bar{E}_0^{k-1}[\Theta_T]\right], \bar{E}_0^{k+1}[\Theta_T]\right) = Cov\left(E_{i,0}\left[\bar{E}_0^{k-1}[\Theta_T]\right], E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right]\right) \\ &= Cov\left(E_{i,0}\left[\bar{E}_0^{k-1}[\Theta_T]\right], \bar{E}_0^k[\Theta_T]\right) = Cov\left(\bar{E}_0^k[\Theta_T], \bar{E}_0^k[\Theta_T]\right), \end{aligned}$$

where the first and the last equalities follow from symmetry, the second equality follows from the fact that $E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right]=\bar{E}_0^{k+1}[\Theta_T]$ almost surely, and the third equality follows from the law of iterated expectations. This expression means $\beta_{h^*-1}=\beta_{2k-1}=\beta_{2k}=\beta_{h^*}$, which contradicts the fact that h^* is the smallest h such that $\beta_{h^*+1}=\beta_{h^*}$.

If $h^*=2k-1$ for some $k\geq 2$, from conditions (39) and (40), we have $Var\left(E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right]\right)=Var\left(\bar{E}_0^k[\Theta_T]\right)$. Following a similar argument as condition (41) for all i, $E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right]=\bar{E}_0^k[\Theta_T]$ almost surely. We henceforth have

$$\begin{aligned}Var\left(\bar{E}_0^k[\Theta_T]\right)&=Cov\left(\bar{E}_0^{k-1}[\Theta_T],\bar{E}_0^{k+1}[\Theta_T]\right)=Cov\left(\bar{E}_0^{k-1}[\Theta_T],E_{i,0}\left[\bar{E}_0^{k}[\Theta_T]\right]\right)\\&=Cov\left(\bar{E}_0^{k-1}[\Theta_T],\bar{E}_0^{k}[\Theta_T]\right)=Cov\left(\bar{E}_0^{k-1}[\Theta_T],E_{i,0}\left[\bar{E}_0^{k-1}[\Theta_T]\right]\right)\\&=Var\left(E_{i,0}\left[\bar{E}_0^{k-1}[\Theta_T]\right]\right),\end{aligned}$$

where the first equality follows a similar argument as condition (38), the second and forth equalities follow from symmetry, the third equality follows from $E_{i,0}\left[\bar{E}_0^k[\Theta_T]\right]=\bar{E}_0^k[\Theta_T]$ almost surely, and the last equality follows from the law of iterated expectations. This expression means that $\beta_{h^*-1}=\beta_{2k-2}=\beta_{2k-1}=\beta_{h^*}$, which contradicts the fact that h^* is the smallest h such that $\beta_{h^*+1}=\beta_{h^*}$.

As a result, β_h is strictly decreasing in h. This implies that $\beta_h<\beta_1=1,\ \forall h\geq 2$. It also means that $\beta_h>0\ \forall h$. If not, there exists a h^* such that $\beta_{h^*}=0$. From strict monotonicity, we then have $\beta_{h^*+1}<\beta_{h^*}=0$, which contradicts $\beta_{h^*+1}\geq 0$. This finishes the proof of Proposition 3.

Proof of Corollary 1. Corollary 1 follows directly from part (ii) of Theorem 1.

Proof of Theorem 2. To simplify notation, we extend the definition of $s_{h,\tau}=\sum_{r=1}^{h}\chi_{r,\tau}$ for all $h>\tau$. In the case that $h>\tau$, from Lemma 2, we have $\chi_{h,\tau}=0$. As a result, $s_{h,\tau}=s_{\tau,\tau}$ for all $h>\tau$. We also define $s_{0,\tau}=0$ for all $\tau\geq 1$.
From condition (16), we have

$$s_{h,\tau}=\gamma^{\tau-1}(\gamma+\alpha)+\sum_{l=1}^{\tau-1}\alpha\gamma^{l-1}s_{h-1,\tau-l}\quad\forall h\geq 1\ and\ \tau\geq 1. \tag{43}$$

Now, for all $\tau\geq 1$, as $\chi_{h,\tau}=0$ for $h>\tau$, we can use $d_\tau=s_{\tau,\tau}$ denote the combined effect of beliefs of all different orders. From condition (43), we have

$$d_\tau=\gamma^{\tau-1}(\gamma+\alpha)+\sum_{l=1}^{\tau-1}\alpha\gamma^{l-1}d_{\tau-l}\quad\forall\tau\geq 1, \tag{44}$$

where we use the fact that $s_{h,\tau}=s_{\tau,\tau}$ for all $h>\tau$. From condition (44), we can easily verify, by induction, that

$$d_\tau=(\gamma+\alpha)^\tau\quad\forall\tau\geq 1. \tag{45}$$

For any $h\geq 1$, we now prove that $s_{h,\tau}/s_{\tau,\tau}=s_{h,\tau}/d_\tau$, strictly decreases with $\tau\geq h$. Notice that from condition (43), we have, for all $\tau \geq h \geq 1$,

$$s_{h,\tau+1} = \gamma^{\tau}(\gamma+\alpha) + \alpha s_{h-1,\tau} + \sum_{l=1}^{\tau-1} \alpha\gamma^{l} s_{h-1,\tau-l}$$
$$= \gamma s_{h,\tau} + \alpha s_{h-1,\tau} < (\gamma+\alpha) s_{h,\tau}. \quad (46)$$

Also note that from condition (45), we have $s_{\tau+1,\tau+1} = d_{\tau+1} = (\gamma+\alpha) d_{\tau} = (\gamma+\alpha) s_{\tau,\tau}$. Together, we have $s_{h,\tau+1}/s_{\tau+1,\tau+1} < s_{h,\tau}/s_{\tau,\tau}$ for all $\tau \geq h \geq 1$.

Finally, we prove that, for any $h \geq 1$, $s_{h,\tau}/s_{\tau,\tau} \to 0$ as $\tau \to +\infty$. Because $s_{1,\tau} = \gamma^{\tau-1}(\gamma+\alpha)$ and $s_{\tau,\tau} = (\gamma+\alpha)^{\tau}$, $\lim_{\tau\to\infty} s_{h,\tau}/s_{\tau,\tau} \to 0$ holds for $h = 1$. Suppose there is some h such that $\lim_{\tau\to\infty} s_{h,\tau}/s_{\tau,\tau} \to 0$ does not hold, let $h^* > 1$ be the smallest of such h. As $s_{h^*,\tau}/s_{\tau,\tau}$ is strictly decreasing in τ, there exists $\Gamma > 0$ such that $\lim_{\tau\to\infty} s_{h^*,\tau}/s_{\tau,\tau} \to \Gamma$. From conditions (45) and (46), we have $\frac{s_{h^*,\tau+1}}{s_{\tau+1,\tau+1}} = \frac{\gamma}{\gamma+\alpha}\frac{s_{h^*,\tau}}{s_{\tau,\tau}} + \frac{\alpha}{\gamma+\alpha}\frac{s_{h^*-1,\tau}}{s_{\tau,\tau}}$. Let $\tau \to +\infty$, we have $\Gamma = \frac{\gamma}{\gamma+\alpha}\Gamma$. This cannot be true as $\alpha, \gamma > 0$. As a result, $\lim_{\tau\to\infty} s_{h,\tau}/s_{\tau,\tau} \to 0$ for all $h \geq 1$.

Proof of Proposition 4. By Theorem 1, the ratio $\frac{\phi_T}{\phi_T^*}$ is strictly decreasing in T and bounded in $(0,1)$. It follows that $\frac{\phi_T}{\phi_T^*}$ necessarily converges to some $\varphi \in [0,1)$ as $T \to \infty$. Similarly, by Proposition 3, β_h is strictly decreasing in T and bounded in $(0,1)$. It follows that β_h necessarily converges to some $\underline{\beta} \in [0,1)$ as $T \to \infty$.

We first prove $\varphi = \underline{\beta} \equiv \lim_{h\to\infty} \beta_h$. We note that for, any $\vartheta > 0$, there exists a h^*, such that $\left|\beta_h - \underline{\beta}\right| < \frac{\vartheta}{2}$ for all $h \geq h^*$. From Theorem 2, we can then find $T^* \in \mathbb{N}_+$ such that, for all $T \geq T^*$, $\frac{s_{h^*-1,T}}{s_{T,T}} \leq \frac{\vartheta}{2}$. Together with conditions (18) and (20), we have, for all $T \geq \max\{h^*, T^*\}$,

$$\left|\frac{\phi_T}{\phi_T^*} - \underline{\beta}\right| = \left|\frac{\sum_{h=1}^{T}\chi_{h,T}\left(\beta_h - \underline{\beta}\right)}{s_{T,T}}\right| = \left|\frac{\sum_{h=1}^{h^*-1}\chi_{h,T}\left(\beta_h - \underline{\beta}\right) + \sum_{h=h^*}^{T}\chi_{h,T}\left(\beta_h - \underline{\beta}\right)}{s_{T,T}}\right|$$
$$\leq \frac{\sum_{h=1}^{h^*-1}\chi_{h,T}}{s_{T,T}} + \frac{\sum_{h=h^*}^{T}\chi_{h,T}\frac{\vartheta}{2}}{s_{T,T}}$$
$$\leq \frac{\vartheta}{2} + \frac{\vartheta}{2} = \vartheta,$$

where the first inequality we use the fact that $\left|\beta_h - \underline{\beta}\right| \leq 1$ and the second inequality uses the fact that $\frac{\sum_{h=h^*}^{T}\chi_{h,T}}{s_{T,T}} \leq \frac{s_{T,T}}{s_{T,T}} = 1$. As a result, $\varphi \equiv lim_{T\to\infty}\frac{\phi_T}{\phi_T^*} = \underline{\beta}$.

Finally, from condition (40), we know $\beta_{2h-1} = \frac{Var\left(\bar{E}_0^h[\Theta_T]\right)}{Var\left(\bar{E}_0^1[\Theta_T]\right)}$. If $\lim_{h\to\infty} Var\left(\bar{E}^h[\Theta_T]\right) = 0$, we have $\lim_{h\to\infty}\beta_{2h-1} = 0$. As β_h is decreasing in h, we also have

$$\underline{\beta} = \lim_{h\to\infty}\beta_h = 0. \quad (47)$$

As a result, $lim_{T\to\infty}\frac{\phi_T}{\phi_T^*} = 0$.

Proof of Proposition 5. Under the assumed information structure, we have for any $h \in \{1, ..., T\}$ and $0 \leq t_1 < t_2 < \cdots < t_h < T$,

$$\bar{E}_{t_1}[\bar{E}_{t_2}[...[\bar{E}_{t_h}[\Theta_T]...]] = \lambda^h z. \quad (48)$$

Now we prove by induction that, for all $t \leq T-1$,

$$a_t = (\gamma + \alpha) \left\{ \Pi_{\tau=t+1}^{T-1} (\gamma + \lambda\alpha) \bar{E}_t[\Theta_T] \right\}. \tag{49}$$

Since $\Theta_t = 0$ for all $t \neq T$, together with condition (16), we have $a_T = \Theta_T$ and $a_{T-1} = (\gamma + \alpha) \bar{E}_{T-1}[\Theta_T]$. As a result, condition (49) holds for $t = T-1$. Now, pick a $t \leq T-2$, assume that the claim holds for all $\tau \in \{t+1, ..., T-1\}$, and let us prove that it also holds for t. Using the claim for all $\tau \in \{t+1, ..., T-1\}$, condition (16), and condition (48), we have, for $t \leq T-2$,

$$a_t = \gamma^{T-t} \bar{E}_t [\Theta_T] + \alpha \bar{E}_t [a_{t+1}] + \alpha \sum_{k=2}^{T-t} \gamma^{k-1} \bar{E}_t [a_{t+k}],$$

$$\gamma \bar{E}_t [a_{t+1}] = \lambda \gamma^{T-t} \bar{E}_t [\Theta_T] + \lambda\alpha \sum_{k=2}^{T-t} \gamma^{k-1} \bar{E}_t [a_{t+k}].$$

As a result, we have $a_t = \left(\frac{\gamma}{\lambda} + \alpha\right) \bar{E}_t [a_{t+1}]$. Together with condition (49) for $t+1$, we have

$$a_t = (\gamma + \alpha) \left\{ \Pi_{\tau=t+1}^{T-1} (\gamma + \lambda\alpha) \bar{E}_t[\Theta_T] \right\}.$$

This proves condition (49) for all $t \leq T-1$. As a result, $\phi_T = (\gamma + \alpha) \Pi_{t=1}^{T-1} (\gamma + \lambda\alpha)$. This proves Proposition 5.

Proof of Corollary 2. Note that $\delta - \delta' = \alpha (1 - \lambda)$ increases with α for any given $\lambda < 1$.

Proof of Lemma 3. As firms have complete information, the canonical NKPC in condition (10) holds. Substituting it into the consumption beauty contest, condition (5), and using the fact that future markup shocks are unpredictable, we have

$$\tilde{y}_t = -\sigma \tilde{R}_t - \sigma \sum_{k=1}^{\infty} \beta^k \bar{E}_t^c[\tilde{R}_{t+k}] + \sum_{k=1}^{\infty} (1 - \beta + k\sigma\kappa) \beta^{k-1} \bar{E}_t^c[\tilde{y}_{t+k}].$$

Proof of Proposition 6. Let $\left\{\tilde{y}_t^{trap}, \tilde{\pi}_t^{trap}\right\}_{t=0}^{T}$ denote the liquidity-trap level of output and inflation (i.e., the one obtained when the period-T nominal interest rate is fixed at the steady-state value, $\tilde{R}_T = 0$). From conditions (9) and (10), we have, for all $t \leq T-1$,

$$\tilde{y}_t - \tilde{y}_t^{trap} = \sigma E_t[\tilde{\pi}_{t+1} - \tilde{\pi}_{t+1}^{trap}] + E_t[\tilde{y}_{t+1} - \tilde{y}_{t+1}^{trap}], \tag{50}$$

$$\tilde{\pi}_t - \tilde{\pi}_t^{trap} = \kappa \left(\tilde{y}_t - \tilde{y}_t^{trap}\right) + \beta E_t[\tilde{\pi}_{t+1} - \tilde{\pi}_{t+1}^{trap}]. \tag{51}$$

Now we will prove the following stronger result, which nests the representation in condition (24): there exists positive scalars $\{\phi_\tau^*, \varpi_\tau^*\}_{\tau \geq 0}$ such that, whenever Assumptions 2 hold and z is commonly known, the equilibrium spending and inflation at any $t \leq T$ are given by

$$\tilde{y}_t - \tilde{y}_t^{trap} = -\phi_{T-t}^* \cdot E_t[\tilde{R}_T], \tag{52}$$

$$\tilde{\pi}_t - \tilde{\pi}_t^{trap} = \kappa \left(\tilde{y}_t - \tilde{y}_t^{trap}\right) - \varpi_{T-t}^* \cdot E_t[\tilde{R}_T]. \tag{53}$$

We prove this result by induction, starting with $t = T$ and proceeding backwards. When $t = T$, under Assumption 2, we have $\tilde{y}_T - \tilde{y}_T^{trap} = -\sigma \tilde{R}_T$ and $\tilde{\pi}_T - \tilde{\pi}_T^{trap} = \kappa \left(\tilde{y}_T - \tilde{y}_T^{trap}\right)$. This verifies (52) and (53) for $t = T$, with

$$\phi_0^* = \sigma \text{ and } \varpi_0^* = 0. \tag{54}$$

Now suppose that the result holds for arbitrary $t \in \{1, ..., T\}$ and let's prove that it also holds for $t-1$. By the assumption that (52) and (53) hold at t along with the Law of Iterated Expectations, we have

$$E_{t-1}[\tilde{y}_t - \tilde{y}_t^{trap}] = -\phi_{T-t}^* \cdot E_{t-1}[\tilde{R}_T],$$

$$E_{t-1}[\tilde{\pi}_t - \tilde{\pi}_t^{trap}] = -\left(\kappa \phi_{T-t}^* + \varpi_{T-t}^*\right) \cdot E_{t-1}[\tilde{R}_T].$$

Using the above together with conditions (50) and (51) verifies that (52) and (53) hold also for $t-1$, with

$$\phi_{T-t+1}^* = (1 + \sigma\kappa)\, \phi_{T-t}^* + \sigma \varpi_{T-t}^*, \tag{55}$$

$$\varpi_{T-t+1}^* = \beta\kappa \phi_{T-t}^* + \beta \varpi_{T-t}^*. \tag{56}$$

This completes the proof of conditions (50) and (51), and gives a recursive formula that can be used to compute ϕ_T^*.

Now we prove the Proposition. From conditions (55) and (56), we have that, for all $\tau \geq 0$,

$$\phi_{\tau+1}^* = (1 + \sigma\kappa)\, \phi_\tau^* + \sigma \varpi_\tau^*, \tag{57}$$

$$\varpi_{\tau+1}^* = \beta\kappa \phi_\tau^* + \beta \varpi_\tau^*. \tag{58}$$

Together with condition (54), we know, as $\kappa > 0$, $\varpi_\tau^* > 0, \;\; \forall \tau \geq 1$. Then, from condition (57), we have $\phi_\tau^* > \sigma, \;\; \forall \tau \geq 1$, and ϕ_τ^* is strictly increasing in τ. Moreover, as $1 + \sigma\kappa > 1$, we know ϕ_τ^* explodes to infinity as $\tau \to \infty$ from condition (57).

Finally, we prove a few more results useful for the rest of the paper. First, we prove a recursive relationship about $\{\phi_\tau^*\}_{\tau \geq 0}$.

$$\frac{\phi_{\tau+1}^*}{\phi_\tau^*} + \beta \frac{\phi_{\tau-1}^*}{\phi_\tau^*} = 1 + \beta + \sigma\kappa \quad \forall \tau \geq 1. \tag{59}$$

From condition (57), we have, for all $\tau \geq 1$,

$$\beta \phi_\tau^* = \beta (1 + \sigma\kappa)\, \phi_{\tau-1}^* + \sigma \beta \varpi_{\tau-1}^*.$$

Together with conditions (57) and (58), we arrive at condition (59).

Second, we prove that, when $\kappa > 0$,

$$\frac{\phi^*_\tau}{\phi^*_{\tau-1}} \text{ is strictly increasing in } \tau \geq 1. \tag{60}$$

From conditions (57) and (58), we have $\phi^*_1 = \sigma\left(1+\sigma\kappa\right)$ and $\phi^*_2 = \sigma\left(\left(1+\sigma\kappa\right)^2 + \sigma\kappa\beta\right)$. As a result, when $\kappa > 0$,

$$\frac{\phi^*_2}{\phi^*_1} = 1 + \sigma\kappa + \frac{\sigma\kappa\beta}{1+\sigma\kappa} > \frac{\phi^*_1}{\phi^*_0}.$$

Now we proceed by induction. Suppose that, for $\tau \geq 1$, we have $\frac{\phi^*_{\tau+1}}{\phi^*_\tau} > \frac{\phi^*_\tau}{\phi^*_{\tau-1}}$. Using condition (59) for τ and $\tau+1$, we have $\frac{\phi^*_{\tau+2}}{\phi^*_{\tau+1}} > \frac{\phi^*_{\tau+1}}{\phi^*_\tau}$. This proves (60).

Finally, from condition (59), we know $\frac{\phi^*_\tau}{\phi^*_{\tau-1}}$ is bounded above. Together with (60), $\frac{\phi^*_\tau}{\phi^*_{\tau-1}}$ must converge to $\Gamma^* > 0$, as $\tau \to \infty$. From condition (59) again, we know Γ^* satisfy

$$\Gamma^* + \beta\frac{1}{\Gamma^*} = 1 + \beta + \sigma\kappa. \tag{61}$$

Proof of Proposition 7. With $\left\{\tilde{y}^{trap}_t, \tilde{\pi}^{trap}_t\right\}^T_{t=0}$ defined as in the proof of Proposition 6, along with the fact that it is common knowledge monetary policy replicates flexible-price allocations from $T+1$ and on, we can rewrite the two beauty contests as follows:

$$\tilde{y}_t - \tilde{y}^{trap}_t = -\sigma\beta^{T-t}\bar{E}^c_t[\tilde{R}_T] + \sum_{k=1}^{T-t}\sigma\beta^{k-1}\bar{E}^c_t\left[\tilde{\pi}_{t+k} - \tilde{\pi}^{trap}_{t+k}\right] + (1-\beta)\sum_{k=1}^{T-t}\beta^{k-1}\bar{E}^c_t\left[\tilde{y}_{t+k} - \tilde{y}^{trap}_{t+k}\right], \tag{62}$$

$$\tilde{\pi}_t - \tilde{\pi}^{trap}_t = \kappa\left(\tilde{y}_t - \tilde{y}^{trap}_t\right) + \kappa\sum_{k=1}^{T-t}\left(\beta\theta\right)^k\bar{E}^f_t\left[\tilde{y}_{t+k} - \tilde{y}^{trap}_{t+k}\right] + \tfrac{1-\theta}{\theta}\sum_{k=1}^{T-t}\left(\beta\theta\right)^k\bar{E}^f_t\left[\tilde{\pi}_{t+k} - \tilde{\pi}^{trap}_{t+k}\right]. \tag{63}$$

Consider the following claim, which nests the representation in condition (25): under Assumption 3, there exists functions $\phi, \varpi : (0,1] \times (0,1] \times \mathbb{N} \to \mathbb{R}_+$ such that, for any $t \leq T-1$,

$$\tilde{y}_t - \tilde{y}^{trap}_t = -\phi\left(\lambda_c, \lambda_f, T-t\right)\bar{E}^c_t[\tilde{R}_T], \tag{64}$$

$$\tilde{\pi}_t - \tilde{\pi}^{trap}_t = \kappa\left(\tilde{y}_t - \tilde{y}^{trap}_t\right) - \varpi\left(\lambda_c, \lambda_f, T-t\right)\bar{E}^f_t[\tilde{R}_T]. \tag{65}$$

We now establish this claim by induction.

First, consider $t = T$, as $\tilde{R}_T$ becomes common known at period T, , we have

$$\tilde{y}_T - \tilde{y}^{trap}_T = -\sigma\tilde{R}_T \qquad \text{and} \qquad \tilde{\pi}_T - \tilde{\pi}^{trap}_T = \kappa\left(\tilde{y}_T - \tilde{y}^{trap}_T\right).$$

Then, consider $t = T-1$. From conditions (62) and (63), we have

$$\tilde{y}_{T-1} - \tilde{y}_{T-1}^{trap} = -\sigma(1+\sigma\kappa)\bar{E}_{T-1}^{c}[\tilde{R}_T],$$
$$\tilde{\pi}_{T-1} - \tilde{\pi}_{T-1}^{trap} = \kappa\left(\tilde{y}_{T-1} - \tilde{y}_{T-1}^{trap}\right) - \sigma\kappa\beta\bar{E}_{T-1}^{f}[\tilde{R}_T].$$

It follows that the claim holds for $t = T-1$ with

$$\phi(\lambda_c, \lambda_f, 1) = \sigma\left(1+\sigma\kappa\right) \qquad \text{and} \qquad \varpi(\lambda_c, \lambda_f, 1) = \sigma\kappa\beta. \tag{66}$$

Now, pick an arbitrary $t \leq T-2$, assume that conditions (64) and (65) hold for all $\tau \in \{t+1, ..., T-1\}$, and let us prove that it also holds for t. Since the claim holds for $\tau \in \{t+1, ...T-1\}$, and since $\tilde{y}_T - \tilde{y}_T^{trap} = -\sigma\tilde{R}_T$ and $\tilde{\pi}_T - \tilde{\pi}_T^{trap} = -\kappa\sigma\tilde{R}_T$, from condition (62), we have

$$\tilde{y}_t - \tilde{y}_t^{trap} = -\sigma\beta^{T-t-1}\left(1+\sigma\kappa\right)\bar{E}_t^c[\tilde{R}_T] - \left(1-\beta+\sigma\kappa\right)\sum_{k=1}^{T-t-1}\beta^{k-1}\phi\left(\lambda_c, \lambda_f, T-t-k\right)\bar{E}_t^c[\bar{E}_{t+k}^c[\tilde{R}_T]]$$
$$-\sigma\sum_{k=1}^{T-t-1}\beta^{k-1}\varpi\left(\lambda_c, \lambda_f, T-t-k\right)\bar{E}_t^c\left[\bar{E}_{t+k}^f[\tilde{R}_T]\right].$$

As a result, we have

$$\tilde{y}_t - \tilde{y}_t^{trap} = -\sigma\beta^{T-t-1}\left(1+\sigma\kappa\right)\bar{E}_t^c[\tilde{R}_T]$$
$$-\sum_{k=1}^{T-t-1}\beta^{k-1}\left[\left(1-\beta+\sigma\kappa\right)\lambda_c\phi\left(\lambda_c, \lambda_f, T-t-k\right) + \sigma\lambda_f\varpi\left(\lambda_c, \lambda_f, T-t-k\right)\right]\bar{E}_t^c[\tilde{R}_T],$$

where we have used the fact that, under Assumption 3, for $1 \leq k \leq T-t-1$,

$$\bar{E}_t^c[\bar{E}_{t+k}^c[\tilde{R}_T]] = \lambda_c\bar{E}_t^c[\tilde{R}_T] \quad \text{and} \quad \bar{E}_t^c[\bar{E}_{t+k}^f[\tilde{R}_T]] = \lambda_f\bar{E}_t^c[\tilde{R}_T]. \tag{67}$$

This proves the part of the claim that regards output, condition (64), with

$$\phi\left(\lambda_c, \lambda_f, T-t\right) = \beta^{T-t-1}\left(\sigma+\sigma^2\kappa\right) + \sum_{k=1}^{T-t-1}\beta^{k-1}\left[\left(1-\beta+\sigma\kappa\right)\lambda_c\phi\left(\lambda_c, \lambda_f, T-t-k\right) + \sigma\lambda_f\varpi\left(\lambda_c, \lambda_f, T-t-k\right)\right]. \tag{68}$$

Similarly, the inflation beauty contest in condition (63) gives

$$\tilde{\pi}_t - \tilde{\pi}_t^{trap} = \kappa\left(\tilde{y}_t - \tilde{y}_t^{trap}\right) - \sigma\tfrac{\kappa}{\theta}\left(\beta\theta\right)^{T-t}\bar{E}_t^f[\tilde{R}_T] - \tfrac{\kappa}{\theta}\sum_{k=1}^{T-t-1}\left(\beta\theta\right)^k\phi\left(\lambda_c,\lambda_f,T-t-k\right)\bar{E}_t^f[\bar{E}_{t+k}^c[\tilde{R}_T]]$$
$$- \tfrac{1-\theta}{\theta}\sum_{k=1}^{T-t-1}\left(\beta\theta\right)^k\varpi\left(\lambda_c,\lambda_f,T-t-k\right)\bar{E}_t^f[\bar{E}_{t+k}^f[\tilde{R}_T]].$$

As a result, we have

$$\tilde{\pi}_t - \tilde{\pi}_t^{trap} = \kappa\left(\tilde{y}_t - \tilde{y}_t^{trap}\right)$$
$$- \left\{\sigma\tfrac{\kappa}{\theta}\left(\beta\theta\right)^{T-t} + \sum_{k=1}^{T-t-1}\left(\beta\theta\right)^k\left[\tfrac{\kappa\lambda_c}{\theta}\phi\left(\lambda_c,\lambda_f,T-t-k\right) + \tfrac{(1-\theta)\lambda_f}{\theta}\varpi\left(\lambda_c,\lambda_f,T-t-k\right)\right]\right\}\bar{E}_t^f[\tilde{R}_T],$$

where we have used the fact that, similarly to the consumers' case, for $1 \leq k \leq T-t-1$,

$$\bar{E}_t^f[\bar{E}_{t+k}^c[\tilde{R}_T]] = \lambda_c\bar{E}_t^f[\tilde{R}_T] \quad \text{and} \quad \bar{E}_t^f[\bar{E}_{t+k}^f[\tilde{R}_T]] = \lambda_f\bar{E}_t^f[\tilde{R}_T]. \tag{69}$$

This proves the part of the claim that regards inflation, condition (65) with

$$\varpi\left(\lambda_c,\lambda_f,T-t\right) = \sigma\frac{\kappa}{\theta}\left(\beta\theta\right)^{T-t} + \sum_{k=1}^{T-t-1}\left(\beta\theta\right)^k\left(\tfrac{\kappa\lambda_c}{\theta}\phi\left(\lambda_c,\lambda_f,T-t-k\right) + \tfrac{(1-\theta)\lambda_f}{\theta}\varpi\left(\lambda_c,\lambda_f,T-t-k\right)\right). \tag{70}$$

We finally provide a recursive formula for computing $\phi(\lambda_c,\lambda_f,T-t)$ and $\varpi\left(\lambda_c,\lambda_f,T-t\right)$, which will be useful later. From condition (68), we have, for $t \leq T-2$,

$$\phi\left(\lambda_c,\lambda_f,T-t\right) = \beta\phi\left(\lambda_c,\lambda_f,T-t-1\right) + \left(1-\beta+\sigma\kappa\right)\lambda_c\phi\left(\lambda_c,\lambda_f,T-t-1\right) + \sigma\lambda_f\varpi\left(\lambda_c,\lambda_f,T-t-1\right)$$
$$= \left(\beta + \left(1-\beta+\sigma\kappa\right)\lambda_c\right)\phi\left(\lambda_c,\lambda_f,T-t-1\right) + \sigma\lambda_f\varpi\left(\lambda_c,\lambda_f,T-t-1\right). \tag{71}$$

Similarly, from condition (70), we have, for $t \leq T-2$,

$$\varpi\left(\lambda_c,\lambda_f,T-t\right) = \beta\theta\varpi\left(\lambda_c,\lambda_f,T-t-1\right) + \beta\theta\left(\tfrac{\kappa\lambda_c}{\theta}\phi\left(\lambda_c,\lambda_f,T-t-1\right) + \tfrac{(1-\theta)\lambda_f}{\theta}\varpi\left(\lambda_c,\lambda_f,T-t-1\right)\right)$$
$$= \kappa\beta\lambda_c\phi\left(\lambda_c,\lambda_f,T-t-1\right) + \beta\left[\theta + \left(1-\theta\right)\lambda_f\right]\varpi\left(\lambda_c,\lambda_f,T-t-1\right). \tag{72}$$

From now on, to simplify notation, we use ϕ_τ and ϖ_τ as shortcuts for, respectively, $\phi\left(\lambda_c,\lambda_f,\tau\right)$ and $\varpi\left(\lambda_c,\lambda_f,\tau\right)$.

We first prove part (i) of Proposition 7. From condition (68), we know $\phi_\tau > \sigma\beta^\tau$. The fact that $\phi_\tau < \phi_\tau^*$ is a direct corollary from the monotonicity of ϕ_τ with respect to λ_c and λ_f, which will be proved shorty.

We then prove part (ii) of Proposition 7. As $\kappa > 0$, from conditions (66), (71) and (72), we know that

$\phi_\tau, \varpi_\tau > 0$ for all $\tau \geq 1$.

We will first prove, for $\tau \geq 2$, $\phi_\tau = \phi(\lambda_c, \lambda_f, \tau)$ is strictly increasing in both λ_c and λ_f. We will proceed by induction on τ. For $\tau = 2$, from (66), (71) and (72), we have ϕ_2 and ϖ_2 is strictly increasing in both λ_c and λ_f. Suppose for $\tau \geq 2$, ϕ_τ, ϖ_τ is strictly increasing in both λ_c and λ_f. From conditions (71) and (72), we know $\phi_{\tau+1}$ and $\varpi_{\tau+1}$ are strictly increasing in both λ_c and λ_f, where we use the fact that $\phi_\tau, \varpi_\tau > 0$. This proves that, for $\tau \geq 2$, $\phi_\tau = \phi(\lambda_c, \lambda_f, \tau)$ is strictly increasing in both λ_c and λ_f. Because of the strict monotonicity, we have, for $\tau \geq 2$, whenever $\lambda_c < 1$ and/or $\lambda_f < 1$, $\frac{\phi_\tau}{\phi_\tau^*} = \frac{\phi(\lambda_c,\lambda_f,\tau)}{\phi(1,1,\tau)} < 1$.

We now prove that, whenever $\lambda_c < 1$ and/or $\lambda_f < 1$, the ratio $\frac{\phi_\tau}{\phi_\tau^*} = \frac{\phi(\lambda_c,\lambda_f,\tau)}{\phi_\tau^*}$ is strictly decreasing in $\tau \geq 1$. We start by noticing, from the proof of Proposition 6, we have, for $\tau \geq 3$,

$$\frac{\phi_\tau^*}{\phi_{\tau-1}^*} + \beta\frac{\phi_{\tau-2}^*}{\phi_{\tau-1}^*} = 1 + \beta + \sigma\kappa \quad . \tag{73}$$

Now we prove that ϕ_τ satisfies an inequality with a similar form as (73):

$$\frac{\phi_\tau}{\phi_{\tau-1}} + \beta\frac{\phi_{\tau-2}}{\phi_{\tau-1}} \leq 1 + \beta + \sigma\kappa\lambda_c \leq 1 + \beta + \sigma\kappa \qquad \forall\tau \geq 3. \tag{74}$$

From condition (71), we have, for $\tau \geq 3$,

$$\phi_\tau = (\beta + (1-\beta)\lambda_c)\phi_{\tau-1} + \sigma\kappa\lambda_c\phi_{\tau-1} + \sigma\lambda_f\varpi_{\tau-1},$$
$$\tfrac{\beta}{\beta+(1-\beta)\lambda_c}\phi_{\tau-1} = \beta\phi_{\tau-2} + \tfrac{\sigma\beta\kappa\lambda_c}{\beta+(1-\beta)\lambda_c}\phi_{\tau-2} + \tfrac{\sigma\beta\lambda_f}{\beta+(1-\beta)\lambda_c}\varpi_{\tau-2}.$$

From the previous two conditions, we have, for $\tau \geq 3$,

$$\begin{aligned}\phi_\tau + \beta\phi_{\tau-2} = {} & (\beta + (1-\beta)\lambda_c)\phi_{\tau-1} + \sigma\kappa\lambda_c\phi_{\tau-1} + \sigma\lambda_f\varpi_{\tau-1} \\ & + \tfrac{\beta}{\beta+(1-\beta)\lambda_c}\phi_{\tau-1} - \tfrac{\sigma\beta\kappa\lambda_c}{\beta+(1-\beta)\lambda_c}\phi_{\tau-2} - \tfrac{\sigma\beta\lambda_f}{\beta+(1-\beta)\lambda_c}\varpi_{\tau-2}.\end{aligned} \tag{75}$$

Note that, for $\tau \geq 3$ and $\lambda_c, \lambda_f \in (0, 1]$, we have

$$\left[(\beta + (1-\beta)\lambda_c) + \sigma\kappa\lambda_c + \tfrac{\beta}{\beta+(1-\beta)\lambda_c}\right]\phi_{\tau-1} \leq (1 + \beta + \sigma\kappa\lambda_c)\phi_{\tau-1},$$

and, from condition (72),

$$\begin{aligned}& \sigma\lambda_f\varpi_{\tau-1} - \tfrac{\sigma\beta\kappa\lambda_c}{\beta+(1-\beta)\lambda_c}\phi_{\tau-2} - \tfrac{\sigma\beta\lambda_f}{\beta+(1-\beta)\lambda_c}\varpi_{\tau-2} \\ = & \sigma\lambda_f(\kappa\beta\lambda_c\phi_{\tau-2} + \beta[\theta + (1-\theta)\lambda_f]\varpi_{\tau-2}) - \tfrac{\sigma\beta\kappa\lambda_c}{\beta+(1-\beta)\lambda_c}\phi_{\tau-2} - \tfrac{\sigma\beta\lambda_f}{\beta+(1-\beta)\lambda_c}\varpi_{\tau-2} \\ = & \sigma\kappa\beta\lambda_c\left(\lambda_f - \tfrac{1}{\beta+(1-\beta)\lambda_c}\right)\phi_{\tau-2} + \sigma\beta\lambda_f\left[\theta + (1-\theta)\lambda_f - \tfrac{1}{\beta+(1-\beta)\lambda_c}\right]\varpi_{\tau-2} \\ \leq & 0.\end{aligned}$$

Together with condition (75), we arrive at condition (74).

Now we can prove that, whenever $\lambda_c < 1$ and/or $\lambda_f < 1$, $\frac{\phi_\tau}{\phi^*_\tau}$ is strictly decreasing in τ. We already prove $\frac{\phi_2}{\phi^*_2} < 1 = \frac{\phi_1}{\phi^*_1}$. We proceed by induction on τ. If $\frac{\phi_\tau}{\phi^*_\tau} < \frac{\phi_{\tau-1}}{\phi^*_{\tau-1}}$ for $\tau \geq 2$, we have $\frac{\phi_{\tau-1}}{\phi_\tau} > \frac{\phi^*_{\tau-1}}{\phi^*_\tau}$. From (73) and (74), we have $\frac{\phi_{\tau+1}}{\phi_\tau} < \frac{\phi^*_{\tau+1}}{\phi^*_\tau}$ and thus $\frac{\phi_{\tau+1}}{\phi^*_{\tau+1}} < \frac{\phi_\tau}{\phi^*_\tau}$. This finishes the proof that $\frac{\phi_\tau}{\phi^*_\tau}$ is strictly decreasing in $\tau \geq 1$, whenever $\lambda_c < 1$ and/or $\lambda_f < 1$.

Now we prove that, whenever $\lambda_c < 1$ and/or $\lambda_f < 1$, $\frac{\phi_\tau}{\phi^*_\tau}$ converges to 0 as $\tau \to \infty$. Because $\frac{\phi_\tau}{\phi^*_\tau} > 0$ is strictly decreasing in $\tau \geq 1$, there exists $\Gamma \in [0, 1)$ such that $\frac{\phi_\tau}{\phi^*_\tau} \to \Gamma$ as $\tau \to \infty$. We next prove by contradiction that $\Gamma = 0$.

Suppose first that $\lambda_c < 1$. If $\Gamma > 0$, we have $\frac{\phi_\tau}{\phi^*_\tau}\frac{\phi^*_{\tau-1}}{\phi_{\tau-1}} \to 1$ as $\tau \to \infty$. Because $\frac{\phi^*_\tau}{\phi^*_{\tau-1}} \to \Gamma^*$, we have $\frac{\phi_\tau}{\phi_{\tau-1}} \to \Gamma^*$ and $\frac{\phi_{\tau-2}}{\phi_{\tau-1}} \to \frac{1}{\Gamma^*}$ as $\tau \to \infty$. From condition (61), we have $\frac{\phi_\tau}{\phi_{\tau-1}} + \beta\frac{\phi_{\tau-2}}{\phi_{\tau-1}} \to 1 + \beta + \sigma\kappa$ as $\tau \to \infty$. However, this is inconsistent with (74) when $\lambda_c < 1$. As a result, $\Gamma = 0$.

Suppose next that $\lambda_c = 1$ but $\lambda_f < 1$. We prove a stronger version of (74):

$$\frac{\phi_\tau}{\phi_{\tau-1}} + (1 + \sigma\kappa(1 - \lambda_f))\,\beta\frac{\phi_{\tau-2}}{\phi_{\tau-1}} \leq 1 + \beta + \sigma\kappa \quad \forall\tau \geq 3. \tag{76}$$

From conditions (71) and (72), we have, for $\tau \geq 3$,

$$\begin{aligned}
\phi_\tau &= (1 + \sigma\kappa)\,\phi_{\tau-1} + \sigma\lambda_f\varpi_{\tau-1}, \\
\beta\phi_{\tau-1} &= \beta\phi_{\tau-2} + \beta\sigma\kappa\phi_{\tau-2} + \beta\sigma\lambda_f\varpi_{\tau-2}, \\
\varpi_{\tau-1} &= \kappa\beta\phi_{\tau-2} + \beta\left[\theta + (1-\theta)\,\lambda_f\right]\varpi_{\tau-2}.
\end{aligned}$$

As a result, for $\tau \geq 3$,

$$\begin{aligned}
\phi_\tau + \beta\phi_{\tau-2} &= (1 + \sigma\kappa + \beta)\,\phi_{\tau-1} + \sigma\lambda_f\varpi_{\tau-1} - \beta\sigma\kappa\phi_{\tau-2} - \beta\sigma\lambda_f\varpi_{\tau-2} \\
&\leq (1 + \sigma\kappa + \beta)\,\phi_{\tau-1} + \sigma(\lambda_f - 1)\,\kappa\beta\phi_{\tau-2}.
\end{aligned}$$

This proves (76).

Now, if $\Gamma > 0$, similarly, we have $\frac{\phi_\tau}{\phi^*_\tau}\frac{\phi^*_{\tau-1}}{\phi_{\tau-1}} \to 1$ as $\tau \to \infty$. Because $\frac{\phi^*_\tau}{\phi^*_{\tau-1}} \to \Gamma^*$, we have $\frac{\phi_\tau}{\phi_{\tau-1}} \to \Gamma^*$ and $\frac{\phi_{\tau-2}}{\phi_{\tau-1}} \to \frac{1}{\Gamma^*}$ as $\tau \to \infty$. From condition (61), we have $\frac{\phi_\tau}{\phi_{\tau-1}} + (1 + \sigma\kappa(1-\lambda_f))\,\beta\frac{\phi_{\tau-2}}{\phi_{\tau-1}} \to 1 + \beta + \sigma\kappa + \sigma\kappa(1 - \lambda_f)\,\beta\frac{1}{\Gamma^*}$ as $\tau \to \infty$. However, this is inconsistent with equation (76) when $\lambda_f < 1$. As a result, $\Gamma = 0$ when $\lambda_c = 1$, but $\lambda_f < 1$.

Finally, we prove that, when λ_c is sufficiently low, $\phi(\lambda_c, \lambda_f, \tau)$ converges to zero as $\tau \to \infty$. The eigenvalues of the dynamic system (ϕ_τ, ϖ_τ) based on conditions (71) and (72) are

$$\begin{aligned}
m_1 &= \frac{\beta + (1 - \beta + \sigma\kappa)\,\lambda_c + \beta\left[(1-\theta)\,\lambda_f + \theta\right] - \sqrt{\left(\beta + (1 - \beta + \sigma\kappa)\,\lambda_c - \beta\left[(1-\theta)\,\lambda_f + \theta\right]\right)^2 + 4\sigma\beta\lambda_f\lambda_c\kappa}}{2} > 0; \\
m_2 &= \frac{\beta + (1 - \beta + \sigma\kappa)\,\lambda_c + \beta\left[(1-\theta)\,\lambda_f + \theta\right] + \sqrt{\left(\beta + (1 - \beta + \sigma\kappa)\,\lambda_c - \beta\left[(1-\theta)\,\lambda_f + \theta\right]\right)^2 + 4\sigma\beta\lambda_f\lambda_c\kappa}}{2} > m_1.
\end{aligned}$$

Note that $\lim_{\lambda_c \to 0} m_2 = \beta < 1$. As a result, when λ_c is sufficiently low, both eigenvalues are below 1, which means that $\phi(\lambda_c, \lambda_f, \tau)$ converges to zero as $\tau \to \infty$.

Proof of Proposition 8. We use $\tilde{g}_t$ to denote the amount of government spending at period t. As mentioned in main text, the government spending $\tilde{g}_t$ is financed by lump sum tax at period t, $\tilde{t}_t = \tilde{g}_t$. Similar to the analysis for monetary policy, we assume $\tilde{g}_t$ becomes commonly known at period t and only allow higher-order uncertainty about future $\tilde{g}$.

Similar to the main text, now we start to work with log-linearized representation. Because the introduction of lump-sum tax, the individual budget constraint becomes

$$\sum_{k=0}^{+\infty} \beta^k \tilde{c}_{i,t+k} = \tilde{a}_{i,t} + \sum_{k=0}^{+\infty} \beta^k \left\{ \Omega_1 \left(\tilde{w}_{i,t+k} + \tilde{n}_{i,t+k} \right) + \Omega_2 \tilde{e}_{i,t+k} - (\Omega_1 + \Omega_2 - 1) \tilde{t}_t \right\},$$

where Ω_1 is the ratio of labor income to total income (net of tax) in steady state, Ω_2 is the ratio of dividend income to total income (net of tax) in steady state, and $\Omega_1 + \Omega_2 - 1$ is the ratio of lump sum tax to total income (net of tax) in steady state. On the other hand, the individual optimal labor supply and Euler equation, conditions (28) and (29), still hold here. Together, this gives rise to the optimal expenditure of consumer $i \in \mathcal{I}_c$ in period t,

$$\tilde{c}_{i,t} = \tfrac{(1-\beta)\epsilon\sigma}{\epsilon\sigma+\Omega_1} \tilde{a}_{i,t} - \sigma \sum_{k=1}^{+\infty} \beta^k E_{i,t} \left[\tilde{r}_{t+k} \right] + (1-\beta) \left[\tfrac{(\epsilon+1)\sigma\Omega_1}{\epsilon\sigma+\Omega_1} \tilde{w}_{i,t} + \tfrac{\epsilon\sigma\Omega_2}{\epsilon\sigma+\Omega_1} \tilde{e}_{i,t} - \tfrac{\epsilon\sigma(\Omega_1+\Omega_2-1)}{\epsilon\sigma+\Omega_1} \tilde{t}_t \right] \tag{77}$$
$$+ (1-\beta) \sum_{k=1}^{+\infty} \beta^k E_{i,t} \left[\tfrac{(\epsilon+1)\sigma\Omega_1}{\epsilon\sigma+\Omega_1} \tilde{w}_{i,t+k} + \tfrac{\epsilon\sigma\Omega_2}{\epsilon\sigma+\Omega_1} \tilde{e}_{i,t+k} - \tfrac{\epsilon\sigma(\Omega_1+\Omega_2-1)}{\epsilon\sigma+\Omega_1} \tilde{t}_{t+k} \right].$$

Using the fact that assets average to zero and that future idiosyncratic shocks are unpredictable, we obtain the following condition for aggregate spending:

$$\tilde{c}_t = -\sigma \sum_{k=1}^{+\infty} \beta^k \bar{E}^c_t \left[\tilde{r}_{t+k} \right] + (1-\beta) \left[\tfrac{(\epsilon+1)\sigma\Omega_1}{\epsilon\sigma+\Omega_1} \tilde{w}_t + \tfrac{\epsilon\sigma\Omega_2}{\epsilon\sigma+\Omega_1} \tilde{e}_t - \tfrac{\epsilon\sigma(\Omega_1+\Omega_2-1)}{\epsilon\sigma+\Omega_1} \tilde{t}_t \right] \tag{78}$$
$$+ (1-\beta) \sum_{k=1}^{+\infty} \beta^k \bar{E}^c_t \left[\tfrac{(\epsilon+1)\sigma\Omega_1}{\epsilon\sigma+\Omega_1} \tilde{w}_{t+k} + \tfrac{\epsilon\sigma\Omega_2}{\epsilon\sigma+\Omega_1} \tilde{e}_{t+k} - \tfrac{\epsilon\sigma(\Omega_1+\Omega_2-1)}{\epsilon\sigma+\Omega_1} \tilde{t}_{t+k} \right].$$

The firm side, on the other hand, is essentially same as the case without government spending, as a result, condition (6) still holds, but the formula for marginal cost are different. In particular, from the production function (3) and the optimal labor supply condition (28), we have

$$\tilde{mc}_t = \tilde{w}_t = \epsilon \int_{\mathcal{I}_c} \tilde{n}_{i,t} di + \frac{1}{\sigma} \tilde{c}_t = \epsilon \tilde{y}_t + \frac{1}{\sigma} \tilde{c}_t = \left(\epsilon \Omega_3 + \frac{1}{\sigma} \right) \tilde{c}_t + \epsilon \left(1 - \Omega_3 \right) \tilde{g}_t, \tag{79}$$

where $\tilde{y}_t = \Omega_3 \tilde{c}_t + (1 - \Omega_3) \tilde{g}_t$, $\Omega_3 = \frac{1}{\Omega_1 + \Omega_2}$ is the steady state consumption to output ratio, and $1 - \Omega_3 =$

$\frac{\Omega_1+\Omega_2-1}{\Omega_1+\Omega_2}$ is the steady state government spending to output ratio.[44] As a result, the inflation beauty contest in condition (6) can be written as

$$\tilde{\pi}_t = \kappa\left(\Omega_c \tilde{c}_t + (1-\Omega_c)\,\tilde{g}_t\right) + \kappa \sum_{k=1}^{+\infty} (\beta\theta)^k\, \bar{E}_t^f[\Omega_c \tilde{c}_{t+k} + (1-\Omega_c)\,\tilde{g}_{t+k}] + \tfrac{1-\theta}{\theta} \sum_{k=1}^{+\infty} (\beta\theta)^k\, \bar{E}_t^f[\tilde{\pi}_{t+k}] + \varkappa \tilde{\mu}_t, \quad (80)$$

where $\Omega_c = \frac{\epsilon\Omega_3 + \frac{1}{\sigma}}{\epsilon + \frac{1}{\sigma}}$.

Finally, note that the real profit of monopolist j at period t is given by $e_t^j = \left(\frac{p_t^j}{p_t} - w_t^j\right) y_t^j$. After log-linearization, we have $\tilde{e}_t = -\frac{\frac{\Omega_1}{\Omega_1+\Omega_2}}{1-\frac{\Omega_1}{\Omega_1+\Omega_2}} \tilde{w}_t + \tilde{y}_t = -\frac{\Omega_1}{\Omega_2}\tilde{w}_t + \tilde{y}_t$.[45] Together with condition (79), we have, for all t,

$$\begin{aligned}
\tfrac{(\epsilon+1)\sigma\Omega_1}{\epsilon\sigma+\Omega_1}\tilde{w}_t + \tfrac{\epsilon\sigma\Omega_2}{\epsilon\sigma+\Omega_1}\tilde{e}_t - \tfrac{\epsilon\sigma(\Omega_1+\Omega_2-1)}{\epsilon\sigma+\Omega_1}\tilde{t}_t &= \tfrac{\sigma\Omega_1}{\epsilon\sigma+\Omega_1}\tilde{w}_t + \tfrac{\epsilon\sigma\Omega_2}{\epsilon\sigma+\Omega_1}\tilde{y}_t - \tfrac{\epsilon\sigma(\Omega_1+\Omega_2-1)}{\epsilon\sigma+\Omega_1}\tilde{g}_t \\
&= \tfrac{\epsilon\sigma(\Omega_1+\Omega_2)}{\epsilon\sigma+\Omega_1}\tilde{y}_t + \tfrac{\Omega_1}{\epsilon\sigma+\Omega_1}\tilde{c}_t - \tfrac{\epsilon\sigma(\Omega_1+\Omega_2-1)}{\epsilon\sigma+\Omega_1}\tilde{g}_t \\
&= \tfrac{\epsilon\sigma(\Omega_1+\Omega_2)}{\epsilon\sigma+\Omega_1}\left[\tfrac{1}{\Omega_1+\Omega_2}\tilde{c}_t + \tfrac{\Omega_1+\Omega_2-1}{\Omega_1+\Omega_2}\tilde{g}_t\right] + \tfrac{\Omega_1}{\epsilon\sigma+\Omega_1}\tilde{c}_t - \tfrac{\epsilon\sigma(\Omega_1+\Omega_2-1)}{\epsilon\sigma+\Omega_1}\tilde{g}_t = \tilde{c}_t.
\end{aligned}$$

Substitute it into condition (78), we have

$$\tilde{c}_t = -\sigma \sum_{k=1}^{+\infty} \beta^{k-1} \bar{E}_t^c[\tilde{r}_{t+k}] + \tfrac{1-\beta}{\beta}\left\{\sum_{k=1}^{+\infty} \beta^k \bar{E}_t^c\left[\tilde{c}_{t+k}\right]\right\}. \quad (81)$$

This is exactly the same form of the consumption beauty contest, as condition (5).

Now let us state Proposition 8 formally here. Similar to Assumption 2, we assume $\tilde{g}_T = z + \eta$, where z and η are random variables, independent of one another and of any other shock in the economy, with $z \sim N(0, \sigma_z^2)$ and $\eta \sim N(0, \sigma_\eta^2)$. The former is realized at $t = 0$, and could be interpreted as news about government spending; the latter is realized at $t = T$ and is unpredictable prior to that point.

First consider the complete information outcome. Suppose z is commonly known starting at $t = 0$, we can find a scalar $\phi_{g,T}^*$ such that $\tilde{c}_0 - \tilde{c}_0^{trap} = \phi_{g,T}^* E_0[\tilde{g}_T]$, where $\tilde{c}_0^{trap}$ denotes the "liquidity trap" level of consumption (i.e., the one obtained when it is common knowledge that $\tilde{g}_T = 0$.) We have, when $\kappa > 0$,

$$\phi_{g,T}^* > 0, \text{ is strictly increasing in } T, \text{ and diverges to infinity as } T \to \infty. \quad (82)$$

Now consider the case in which z is not common knowledge. Similar to Section 6, we consider the information structure specified in Assumption 3, in which let each agent receives a private signal about z at period 0. We can then find a scalar $\phi_{g,T}$ such that $\tilde{c}_0 - \tilde{c}_0^{trap} = \phi_{g,T}\bar{E}_0^c[\tilde{g}_T]$. We have, as long as $\kappa > 0$ and

[44] In steady state, the ratio of government spending to consumption will be equal to ratio of lump sum tax to total income (net of tax), $\Omega_1 + \Omega_2 - 1$. This explains the formula for Ω_3.

[45] This expression is equivalent to $\frac{\Omega_2}{\Omega_1+\Omega_2}\tilde{e}_t + \frac{\Omega_1}{\Omega_1+\Omega_2}(\tilde{w}_t + \tilde{y}_t) = \frac{\Omega_2}{\Omega_1+\Omega_2}\tilde{e}_t + \frac{\Omega_1}{\Omega_1+\Omega_2}\left(\tilde{w}_t + \int_{\mathcal{I}_f} \tilde{l}_t^j dj\right) = \tilde{y}_t$. The last equation is true because $\frac{\Omega_1}{\Omega_1+\Omega_2}$ is steady state labor income to total income ratio (before deducting tax) and $\frac{\Omega_2}{\Omega_1+\Omega_2}$ is steady state dividend income to total income ratio (before deducting tax).

information is incomplete, that is $\lambda_c < 1$,[46]

$$\phi_{g,T} \in \left(0, \phi^*_{g,T}\right), \text{ is strictly increasing in } \lambda_c \text{ and } \lambda_f;$$
$$\text{the ratio } \phi_{g,T}/\phi^*_{g,T} \text{ is strictly decreasing in } T \text{ and converges to 0 as } T \to \infty;$$
$$\text{finally, when } \lambda_c \text{ is sufficiently low, } \phi_{g,T} \text{ also converges to 0 as } T \to \infty. \quad (83)$$

We start from the proof of condition (82). Similar to the proof Proposition 6, we can establish that there exists non-negative scalars $\left\{\phi^*_{g,\tau}, \varpi^*_{g,\tau}\right\}_{\tau\geq 0}$ such that, when z is commonly known, the equilibrium spending and inflation at any $t \leq T$ are given by

$$\tilde{c}_t - \tilde{c}_t^{trap} = \phi^*_{g,T-t} \cdot E_t[\tilde{g}_T], \quad (84)$$
$$\tilde{\pi}_t - \tilde{\pi}_t^{trap} = \kappa \left(\Omega_c \left(\tilde{c}_t - \tilde{c}_t^{trap}\right) + (1-\Omega_c)\,\tilde{g}_t\right) + \varpi^*_{g,T-t} \cdot E_t[\tilde{g}_T], \quad (85)$$

where $\tilde{c}_t^{trap}$ and $\tilde{\pi}_t^{trap}$ denotes the "liquidity trap" level of consumption and inflation (i.e., the one obtained when it is common knowledge that $\tilde{g}_T = 0$.) Note that the Euler condition and NKPC with government spending under complete information can be written as:

$$\tilde{c}_t = -\sigma\left\{\tilde{R}_t - E_t\left[\tilde{\pi}_{t+1}\right]\right\} + E_t\left[\tilde{c}_{t+1}\right], \quad (86)$$
$$\tilde{\pi}_t = \kappa\left(\Omega_c\tilde{c}_t + (1-\Omega_c)\,\tilde{g}_t\right) + \beta E_t\left[\tilde{\pi}_{t+1}\right] + \varkappa\tilde{\mu}_t. \quad (87)$$

Using the above expressions, similar to the proof of Proposition 6, we can establish that $\phi^*_{g,0} = 0$, $\varpi^*_{g,0} = 0$, $\phi^*_{g,1} = \sigma\kappa\left(1-\Omega_c\right)$, $\varpi^*_{g,1} = \beta\kappa\left(1-\Omega_c\right)$, and for all $\tau \geq 1$,

$$\phi^*_{g,\tau+1} = \left(1+\sigma\kappa\Omega_c\right)\phi^*_{g,\tau} + \sigma\varpi^*_{g,\tau}, \quad (88)$$
$$\varpi^*_{g,\tau+1} = \beta\kappa\Omega_c\phi^*_{g,\tau} + \beta\varpi^*_{g,\tau}. \quad (89)$$

From condition (88), we can see when $\kappa > 0$, $\phi^*_{g,\tau}$ is positive, strictly increasing in τ and diverges to infinity. This proves condition (82). Similar to condition (59, one can also prove the following recursive relationship about ϕ^*_g:

$$\frac{\phi^*_{g,\tau+1}}{\phi^*_{g,\tau}} + \beta\frac{\phi^*_{g,\tau-1}}{\phi^*_{g,\tau}} = 1 + \beta + \sigma\kappa\Omega_c \quad \forall\tau \geq 1. \quad (90)$$

Moreover, as condition (61), we know $\frac{\phi^*_{g,\tau}}{\phi^*_{g,\tau-1}}$ must converge to $\Gamma^*_g > 0$, as $\tau \to \infty$:

$$\Gamma^*_g + \beta\frac{1}{\Gamma^*_g} = 1 + \beta + \sigma\kappa\Omega_c. \quad (91)$$

We now turn to the case of incomplete information and establish the proof of condition (83). Similar to

[46]For simplicity here, we always remove common knowledge about z among consumer here. We allow $\lambda_f \in (0,1]$. In other words, we nest the case in which firms have perfect knowledge about z.

the proof of Proposition 7, we can find $\phi_g, \varpi_g : (0,1] \times (0,1] \times \mathbb{N} \to \mathbb{R}_{\geq 0}$ such that, for any $t \leq T-1$,

$$\tilde{c}_t - \tilde{c}_t^{trap} = \phi_g\left(\lambda_c, \lambda_f, T-t\right) \bar{E}_t^c[\tilde{g}_T], \tag{92}$$

$$\tilde{\pi}_t - \tilde{\pi}_t^{trap} = \kappa\left(\Omega_c\left(\tilde{c}_t - \tilde{c}_t^{trap}\right) + \left(1-\Omega_c\right)\tilde{g}_t\right) + \varpi_g\left(\lambda_c, \lambda_f, T-t\right) \bar{E}_t^f[\tilde{g}_T]. \tag{93}$$

Using conditions (80) and (81), we have $\phi_g\left(\lambda_c, \lambda_f, 1\right) = \sigma\kappa\left(1-\Omega_c\right)$, $\varpi_g\left(\lambda_c, \lambda_f, 1\right) = \beta\kappa\left(1-\Omega_c\right)$ and, for all $\tau \geq 2$,

$$\phi_g\left(\lambda_c, \lambda_f, \tau\right) = \sigma\beta^{\tau-1}\left(1-\Omega_c\right)\kappa + \sum_{k=1}^{\tau-1}\beta^{k-1}\left[\left(1-\beta+\sigma\kappa\Omega_c\right)\lambda_c\phi_g\left(\lambda_c, \lambda_f, \tau-k\right) + \sigma\lambda_f\varpi_g\left(\lambda_c, \lambda_f, \tau-k\right)\right]; \tag{94}$$

$$\varpi_g\left(\lambda_c, \lambda_f, \tau\right) = \left(1-\Omega_c\right)\frac{\kappa}{\theta}\left(\beta\theta\right)^{T-t} + \sum_{k=1}^{\tau-1}\left(\beta\theta\right)^k\left(\tfrac{\kappa\lambda_c\Omega_c}{\theta}\phi_g\left(\lambda_c, \lambda_f, \tau-k\right) + \tfrac{(1-\theta)\lambda_f}{\theta}\varpi_g\left(\lambda_c, \lambda_f, \tau-k\right)\right). \tag{95}$$

Together, we can establish, for all $\tau \geq 2$,

$$\phi_g\left(\lambda_c, \lambda_f, \tau\right) = \left(\beta + \left(1-\beta+\sigma\kappa\Omega_c\right)\lambda_c\right)\phi_g\left(\lambda_c, \lambda_f, \tau-1\right) + \sigma\lambda_f\varpi_g\left(\lambda_c, \lambda_f, \tau-1\right); \tag{96}$$

$$\varpi_g\left(\lambda_c, \lambda_f, \tau\right) = \kappa\beta\lambda_c\Omega_c\phi_g\left(\lambda_c, \lambda_f, \tau-1\right) + \beta\left[\theta + \left(1-\theta\right)\lambda_f\right]\varpi_g\left(\lambda_c, \lambda_f, \tau-1\right). \tag{97}$$

From the above conditions, we can see that for for all $\tau \geq 2$, $\phi_{g,\tau} = \phi_g\left(\lambda_c, \lambda_f, \tau\right)$ is strictly increasing in λ_c and λ_f. As $\phi_{g,\tau}^* = \phi_g\left(1, 1, T\right)$, we also have $\phi_{g,\tau} \in \left(0, \phi_{g,\tau}^*\right)$.

Now, we now prove that, whenever $\lambda_c < 1$, the ratio $\frac{\phi_{g,\tau}}{\phi_{g,\tau}^*} = \frac{\phi_g(\lambda_c, \lambda_f, \tau)}{\phi_{g,\tau}^*}$ is strictly decreasing in $\tau \geq 1$ and converges to 0 as $\tau \to \infty$. To this goal, similar to condition (74), we try to establish that

$$\frac{\phi_{g,\tau}}{\phi_{g,\tau-1}} + \beta\frac{\phi_{g,\tau-2}}{\phi_{g,\tau-1}} \leq 1 + \beta + \sigma\kappa\Omega_c\lambda_c < 1 + \beta + \sigma\kappa\Omega_c \quad \forall \tau \geq 3. \tag{98}$$

From condition (96), we have, for $\tau \geq 3$,

$$\phi_{g,\tau} = \left(\beta + \left(1-\beta\right)\lambda_c\right)\phi_{g,\tau-1} + \sigma\kappa\Omega_c\lambda_c\phi_{g,\tau-1} + \sigma\lambda_f\varpi_{g,\tau-1},$$
$$\tfrac{\beta}{\beta+(1-\beta)\lambda_c}\phi_{g,\tau-1} = \beta\phi_{g,\tau-2} + \tfrac{\sigma\beta\kappa\Omega_c\lambda_c}{\beta+(1-\beta)\lambda_c}\phi_{g,\tau-2} + \tfrac{\sigma\beta\lambda_f}{\beta+(1-\beta)\lambda_c}\varpi_{g,\tau-2}.$$

From the previous two conditions, we have, for $\tau \geq 3$,

$$\begin{aligned}\phi_{g,\tau} + \beta\phi_{g,\tau-2} = {} & \left(\beta + \left(1-\beta\right)\lambda_c\right)\phi_{g,\tau-1} + \sigma\kappa\Omega_c\lambda_c\phi_{g,\tau-1} + \sigma\lambda_f\varpi_{g,\tau-1} \\ & + \tfrac{\beta}{\beta+(1-\beta)\lambda_c}\phi_{g,\tau-1} - \tfrac{\sigma\beta\kappa\Omega_c\lambda_c}{\beta+(1-\beta)\lambda_c}\phi_{g,\tau-2} - \tfrac{\sigma\beta\lambda_f}{\beta+(1-\beta)\lambda_c}\varpi_{g,\tau-2}.\end{aligned} \tag{99}$$

Note that, for $\tau \geq 3$ and $\lambda_c, \lambda_f \in (0, 1]$, we have

$$\left[(\beta + (1-\beta)\lambda_c) + \sigma\kappa\Omega_c\lambda_c + \tfrac{\beta}{\beta+(1-\beta)\lambda_c}\right]\phi_{g,\tau-1} \leq (1 + \beta + \sigma\kappa\Omega_c\lambda_c)\,\phi_{g,\tau-1},$$

and from condition (97), we have

$$\begin{aligned}
&\sigma\lambda_f\varpi_{g,\tau-1} - \tfrac{\sigma\beta\kappa\Omega_c\lambda_c}{\beta+(1-\beta)\lambda_c}\phi_{g,\tau-2} - \tfrac{\sigma\beta\lambda_f}{\beta+(1-\beta)\lambda_c}\varpi_{g,\tau-2} \\
=&\sigma\lambda_f\left(\kappa\beta\lambda_c\Omega_c\phi_{g,\tau-2} + \beta\left[\theta + (1-\theta)\lambda_f\right]\varpi_{g,\tau-2}\right) - \tfrac{\sigma\beta\kappa\Omega_c\lambda_c}{\beta+(1-\beta)\lambda_c}\phi_{g,\tau-2} - \tfrac{\sigma\beta\lambda_f}{\beta+(1-\beta)\lambda_c}\varpi_{g,\tau-2} \\
=&\sigma\kappa\beta\Omega_c\lambda_c\left(\lambda_f - \tfrac{1}{\beta+(1-\beta)\lambda_c}\right)\phi_{g,\tau-2} + \sigma\beta\lambda_f\left[\theta + (1-\theta)\lambda_f - \tfrac{1}{\beta+(1-\beta)\lambda_c}\right]\varpi_{g,\tau-2} \\
\leq&0.
\end{aligned}$$

Together with condition (99), we reach at condition (98). To prove $\frac{\phi_{g,\tau}}{\phi^*_{g,\tau}}$ is strictly decreasing in τ, note that we already prove that $\frac{\phi_{g,2}}{\phi^*_{g,2}} < 1 = \frac{\phi_{g,1}}{\phi^*_{g,1}}$. We proceed by induction on τ. If $\frac{\phi_{g,\tau}}{\phi^*_{g,\tau}} < \frac{\phi_{g,\tau-1}}{\phi^*_{g,\tau-1}}$ for $\tau \geq 2$, we have $\frac{\phi_{g,\tau-1}}{\phi_{g,\tau}} > \frac{\phi^*_{g,\tau-1}}{\phi^*_{g,\tau}}$. From (90) and (98), we have $\frac{\phi_{g,\tau+1}}{\phi_{g,\tau}} < \frac{\phi^*_{g,\tau+1}}{\phi^*_{g,\tau}}$ and thus $\frac{\phi_{g,\tau+1}}{\phi^*_{g,\tau+1}} < \frac{\phi_{g,\tau}}{\phi^*_{g,\tau}}$. This finishes the proof that $\frac{\phi_{g,\tau}}{\phi^*_{g,\tau}}$ is strictly decreasing in $\tau \geq 1$.

To prove that $\frac{\phi_{g,\tau}}{\phi^*_{g,\tau}}$ converges to 0 as $\tau \to \infty$. Because $\frac{\phi_{g,\tau}}{\phi^*_{g,\tau}} > 0$ is strictly decreasing in $\tau \geq 1$, there exists $\Gamma_g \in [0, 1)$ such that $\frac{\phi_{g,\tau}}{\phi^*_{g,\tau}} \to \Gamma_g$ as $\tau \to \infty$. If $\Gamma_g > 0$, we have $\frac{\phi_{g,\tau}}{\phi^*_{g,\tau}}\frac{\phi^*_{g,\tau-1}}{\phi_{g,\tau-1}} \to 1$ as $\tau \to \infty$. Because $\frac{\phi^*_{g,\tau}}{\phi^*_{g,\tau-1}} \to \Gamma^*_g$, we have $\frac{\phi_{g,\tau}}{\phi_{g,\tau-1}} \to \Gamma^*_g$ and $\frac{\phi_{g,\tau-2}}{\phi_{g,\tau-1}} \to \frac{1}{\Gamma^*_g}$ as $\tau \to \infty$. From condition (91), we have $\frac{\phi_{g,\tau}}{\phi_{g,\tau-1}} + \beta\frac{\phi_{g,\tau-2}}{\phi_{g,\tau-1}} \to 1 + \beta + \sigma\kappa\Omega_c$ as $\tau \to \infty$. However, this is inconsistent with (98) as $\lambda_c < 1$ and $\kappa > 0$. As a result, $\Gamma_g = 0$.

Finally, we prove that, when λ_c is sufficiently low, $\phi_g(\lambda_c, \lambda_f, \tau)$ converges to zero as $\tau \to \infty$. The eigenvalues of the dynamic system $(\phi_{g,\tau}, \varpi_{g,\tau})$ based on conditions (96) and (97) are

$$\begin{aligned}
m_1 =& \frac{\beta + (1-\beta+\sigma\kappa\Omega_c)\lambda_c + \beta\left[(1-\theta)\lambda_f + \theta\right]}{2} \\
&- \frac{\sqrt{\left(\beta + (1-\beta+\sigma\kappa\Omega_c)\lambda_c - \beta\left[(1-\theta)\lambda_f + \theta\right]\right)^2 + 4\sigma\beta\lambda_f\lambda_c\Omega_c\kappa}}{2} > 0, \\
m_2 =& \frac{\beta + (1-\beta+\sigma\kappa\Omega_c)\lambda_c + \beta\left[(1-\theta)\lambda_f + \theta\right]}{2} \\
&+ \frac{\sqrt{\left(\beta + (1-\beta+\sigma\kappa\Omega_c)\lambda_c - \beta\left[(1-\theta)\lambda_f + \theta\right]\right)^2 + 4\sigma\beta\lambda_f\lambda_c\Omega_c\kappa}}{2} > m_1.
\end{aligned}$$

Note that $\lim_{\lambda_c \to 0} m_2 = \beta < 1$. As a result, when λ_c is sufficiently low, both eigenvalues are below 1, which means that $\phi_g(\lambda_c, \lambda_f, \tau)$ converges to zero as $\tau \to \infty$.

Appendix B. Rational Inattention and Learning

In this appendix, we first sketch how the friction we consider can be recast as the product of rational inattention. We next extend Theorem 1 to two leading forms of learning studied in the literature. We finally prove an asymptotic version of our horizon effect for arbitrary forms of learning.

The Friction as the Product of Rational Inattention. We now briefly sketch how the friction we consider can be recast as the product of rational inattention and, in this sense, a form of bounded rationality.[47]

We let Θ_T be Normally distributed and, to sharpen the exposition, we assume that Θ_T is realized at $t = 0$ (think of Θ_T as being itself the news). The typical agent is nevertheless unable to observe Θ_T perfectly. Instead, for every t, the action $a_{i,t}$ must be measurable in $\omega_i^t \equiv \{\omega_{i,\tau}\}_{\tau \leq t}$, where $\omega_{i,\tau}$ is a noisy signal obtained in period τ. The noise is assumed to be independent across the agent.[48] This guarantees that all aggregate outcomes are functions of Θ_T and, therefore, we can reduce the rational-inattention problem faced by each agent to the choice of a sequence of signals about Θ_T. We finally let these signals be chosen optimally, that is, so as to maximize the agent's ex ante payoff, subject to the following constraint:

$$\mathcal{I}\left(\omega_{i,t}, \Theta_T \mid \omega_i^{t-1} \right) \leq \kappa^{RI}, \tag{100}$$

where $\mathcal{I}\left(\omega_{i,t}, \Theta_T | \omega_i^{t-1}\right)$ is the (entropy-based) information flow between the period-t signal and Θ_T, conditional on the agent's past information, and $\kappa^{RI} > 0$ is an exogenous scalar.

The usual interpretation of constraint (100) is that it captures the agent's limited cognitive capacity in tracking Θ_T. But since beliefs about Θ_T map, in equilibrium, to beliefs of future outcomes, one can also think of (100) as a constraint on the agent's ability to figure out the likely effects of the underlying variation in Θ_T. This echoes Tirole (2015), who interprets rational inattention in games as a form of "costly contemplation."

As long as the prior about Θ_T is Gaussian and the objective function is quadratic, which is the case here by assumption, the optimal signal is also Gaussian. Furthermore, the noise in the signal has to be independent across periods, or else the agent could economize on cognitive costs, that is, relax the constraint in (100). These arguments are standard; see, e.g., Mackowiak, Matejka and Wiederholt (2017). The case studied here is actually far simpler than the one studied in the literature, because the relevant fundamental (Θ_T) does not vary as time passes. We infer that the optimal signal at every $t \leq T-1$ is given by $\omega_{i,t} = \Theta_T + v_{i,t}$, where the noise $v_{i,t}$ is orthogonal to both Θ_T and $\{v_{i,\tau}\}_{\tau<t}$. Letting τ_t denotes the precision (i.e., the reciprocal of the variance) of this noise, we have that the period-t information flow is given by

$$\mathcal{I}\left(\omega_{i,t}, \Theta_T \mid \omega_i^{t-1} \right) = \tfrac{1}{2} \log_2 \left(1 + \frac{\tau_t}{\varsigma_t} \right),$$

[47] For the rational inattention problem to be well-defined, we need to specify a payoff structure behind condition (11). For example, we can think player i's payoff is $\mathcal{U}_i = \sum_t \beta^t U(a_{i,t}, a_{i,t+1}; \Theta_t, a_t)$, where U is a reverse-engineered quadratic utility function so that the player's best-response condition is given by (11).

[48] Although this assumption is separate the information-flow constraint (100), it is standard in the literature (e.g., Woodford, 2003*a*, Mackowiak and Wiederholt, 2009, Luo et al., 2017) and seems appealing if one interprets the noise as the product of cognitive limitations. It is also broadly consistent with experimental evidence (e.g., Khaw, Stevens and Woodford, 2016).

where ς_t denotes the precision of the agent's prior in the beginning of period t; the latter is defined recursively by $\varsigma_0 = \sigma_\theta^{-2}$ and $\varsigma_{t+1} = \varsigma_t + \tau_t$ for $t \in \{0, ..., T-1\}$. It follows that the information constraint (100) pins down the sequence $\{\tau_t\}_{t=0}^{T-1}$ as a function of κ^{RI} and σ_θ^2 alone. All in all, the setting we have considered here is therefore nested in the cases with learning studied in the next part of this appendix, for which Theorem 1 applies. This completes the rational-inattention interpretation of our results.[49]

The Horizon Effect with Sticky Information or Noisy Private Learning. We now extend Theorem 1 to two leading forms of learning studied in the literature. Let $\Theta_T \sim N\left(0, \sigma_\theta^2\right)$ and consider the following two cases of learning.

Case 1. Agents become gradually aware of Θ_T, as in Mankiw and Reis (2002) and Wiederholt (2015). In particular, at each $t \in \{0, ..., T-1\}$, a fraction λ_{sticky} of agents who have not become aware about Θ_T become aware about Θ_T. Θ_T becomes commonly known at period T.

Case 2. Agents receive a new private signal each period, as in Woodford (2003*a*), Nimark (2008), and Mackowiak and Wiederholt (2009). In particular, at at each $t \in \{0, ..., T-1\}$, agent i's new information about Θ_T is summarized in the private signal $s_{i,t} = \Theta_T + v_{i,t}$, where $v_{i,t} \sim N\left(0, \sigma_{v,t}^2\right)$ is i.i.d across i and t, and independent of Θ_T. Θ_T becomes commonly known at period T.

In both cases, there exists $\{\lambda_t\}_{t=0}^{T-1}$ such that, for all t, $\lambda_t \in (0,1)$ and, for any $h \in \{1, ..., T\}$ and $0 \le t_1 < t_2 < \cdots < t_h < T$,

$$\bar{E}_{t_1}[\bar{E}_{t_2}[...[\bar{E}_{t_h}[\Theta_T]...]] = \lambda_{t_1} \cdots \lambda_{t_h} \Theta_T, \tag{101}$$

In case 1, $\lambda_t = 1 - (1 - \lambda_{sticky})^{t+1}$. In case 2, $\lambda_t = \frac{\sum_{\tau=0}^{t} \sigma_{v,\tau}^{-2}}{\sum_{\tau=0}^{t} \sigma_{v,\tau}^{-2} + \sigma_\theta^{-2}}$.

Now we prove by induction that, for all $t \le T-1$,

$$a_t = (\gamma + \alpha) \left\{ \Pi_{\tau=t+1}^{T-1} (\gamma + \lambda_\tau \alpha) \bar{E}_t[\Theta_T] \right\}. \tag{102}$$

Since $\Theta_t = 0$ for all $t \neq T$, together with condition (16), we have $a_T = \Theta_T$ and $a_{T-1} = (\gamma + \alpha) \bar{E}_{T-1}[\Theta_T]$. As a result, condition (102) holds for $t = T-1$. Now, pick a $t \le T-2$, assume that the claim holds for all $\tau \in \{t+1, ..., T-1\}$, and let us prove that it also holds for t. Using the claim for all $\tau \in \{t+1, ..., T-1\}$, condition (16), and condition (101), we have, for $t \le T-2$,

$$a_t = \gamma^{T-t} \bar{E}_t\left[\Theta_T\right] + \alpha \bar{E}_t\left[a_{t+1}\right] + \alpha \sum_{k=2}^{T-t} \gamma^{k-1} \bar{E}_t\left[a_{t+k}\right];$$

$$\gamma \bar{E}_t\left[a_{t+1}\right] = \lambda_{t+1} \gamma^{T-t} \bar{E}_t\left[\Theta_T\right] + \alpha \lambda_{t+1} \sum_{k=2}^{T-t} \gamma^{k-1} \bar{E}_t\left[a_{t+k}\right].$$

[49] Two remarks are worth making. First, suppose that the agents have a limited cognitive capacity to allocate, not per period, but across the entire horizon. In this case, the series of per-period constraints seen in condition (100) are replaced by a single constraint over the entire horizon, namely $\mathcal{I}\left(\{\omega_{i,t}\}_{t=0}^{T-1}, \Theta_T\right) \le \kappa^{RI}$. It then becomes optimal to allocate all capacity to the period-0 signal, which means that this case can justify to our baseline analysis, which assumes away learning. Second, suppose that the news about the fundamental of interest (say, monetary policy) come at the same time with news about another fundamental (say, TFP) and that the agents can economize on cognitive effort by obtaining a joint signal of all the news. In this case, rational inattention can contribute to confounding of one kind of news with another, a scenario not considered here.

As a result, we have $a_t = \left(\frac{\gamma}{\lambda_{t+1}} + \alpha\right) \bar{E}_t\left[a_{t+1}\right]$. Together with condition (102) for $t+1$, we have

$$a_t = (\gamma+\alpha)\left\{\Pi_{\tau=t+1}^{T-1}\left(\gamma+\lambda_\tau\alpha\right)\bar{E}_t[\Theta_T]\right\}.$$

This proves condition (102) for all $t \leq T-1$. As a result, $\phi_T = (\gamma+\alpha)\,\Pi_{t=1}^{T-1}\left(\gamma+\lambda_t\alpha\right)$. Together with the fact that $\lambda_t \in (0,1)$ and $\phi_T^* = (\gamma+\alpha)^T$, we prove Theorem 1 for the case with learning.

The Limit Property with Arbitrary Learning. As noted in the main text, it is possible to prove, under quite general conditions, an *asymptotic* version of our horizon effect: as long as the higher-order uncertainty is bounded away from zero (in a sense we make precise now), the scalar ϕ_T becomes vanishingly small relative to ϕ_T^* as $T\to\infty$.

For any $t \leq T-1$ and any $k \in \{1,...,T-t\}$, we henceforth let B_t^k collect all the relevant k-order beliefs, as of period t :

$$B_t^k \equiv \left\{x \;:\; \exists(t_1,t_2,\cdots,t_k),\text{ with } t=t_1<t_2<\cdots<t_k\leq T-1,\text{ such that } x=\bar{E}_{t_1}[\bar{E}_{t_2}[\cdots\bar{E}_{t_k}[\Theta_T]\cdots]]\right\}.$$

We next introduce the following assumption.

Assumption 4 (Non-Vanishing Higher-Order Uncertainty) *There exists an $\epsilon>0$ such that:*

(i) For all $t\in\{0,....,T-1\}$, there exists at least a mass ϵ of agents such that

$$Var\left(E_t[x]|\omega_i^t\right) \geq \epsilon Var\left(E_t\left[x\right]\right),$$

for all $x\in B_\tau^k\cup\{\Theta_T\}$, $\tau\in\{t+1,...,T-1\}$, and $k\in\{1,...,T-\tau\}$, where ω_i^t is agent i's information set at period t and $E_t\left[x\right]$ denotes the rational expectation of variable x conditional on the union of information sets of all agents in the economy available at period t.

(ii) $Var\left(\bar{E}_0[\Theta_T]\right)\geq\epsilon$.

To interpret this assumption, note that complete information imposes that $E_t\left[x\right]$ is known to every agent, and therefore that $Var\left(E_t\left[x\right]]|\omega_i^t\right)=0$, regardless of how volatile $E_t\left[x\right]$ itself is. By contrast, letting $Var\left(E_t\left[x\right]|\omega_i^t\right)>0$ whenever $Var\left(E_t\left[x\right]\right)>0$ is essentially *tautological* to assuming that agents have incomplete information or, equivalently, that they face higher-order uncertainty. Relative to this tautology, part (i) introduces an arbitrarily small bound on the level of higher-order uncertainty. This bound guarantees that the higher-order uncertainty does not vanish as we let T go to infinity. Part (ii), on the other hand, means simply that there is non-trivial variation in first-order beliefs in the first place. The next result then formalizes our point that our horizon effect, at least in its limit form, holds for arbitrary forms of learning.

Proposition 9 (Limit) *Under Assumption 4, the ratio $\frac{\phi_T}{\phi_T^*}$ converges to zero as $T\to\infty$.*

Proof of Proposition 9. We first prove that, under Assumption (4),

$$Var\left(y\right)\leq\left(1-\epsilon^2\right)^k Var\left(\Theta_T\right). \tag{103}$$

for any $t \leq T-1$ and $y = \bar{E}_t \bar{E}_{t_2} ... \bar{E}_{t_k}[\Theta_T] \in B_t^k$.

To simplify notation, let $x = \bar{E}_{t_2} ... \bar{E}_{t_k}[\Theta_T]$ for $k \geq 2$, and $x = \Theta_T$ for $k = 1$. From Assumption 4, we have, there is at least a mass ϵ of agents such that

$$Var\left(E_t\left[x\right]|\omega_i^t\right) \geq \epsilon Var\left(E_t\left[x\right]\right).$$

As a result,

$$E\left[Var\left(E_t\left[x\right]|\omega_i^t\right)|\Omega_t\right] \geq \epsilon^2 Var\left(E_t\left[x\right]\right),$$

where Ω_t is the cross-sectional distribution of agent's information set ω_i^t at period t. Using the law of total variance, we have

$$Var\left(E_t\left[x\right]\right) = E\left[Var\left(E_t\left[x\right]|\omega_i^t\right)|\Omega_t\right] + Var\left(E\left[E_t\left[x\right]|\omega_i^t\right]\right) = E\left[Var\left(E_t\left[x\right]|\omega_i^t\right)|\Omega_t\right] + Var\left(E\left[x|\omega_i^t\right]\right).$$

As a result, we have[50]

$$\begin{aligned} Var\left(y\right) = Var\left(\bar{E}_t \bar{E}_{t_2} ... \bar{E}_{t_k}[\Theta_T]\right) = Var\left(\bar{E}_t\left[x\right]\right) \leq Var\left(E\left[x|\omega_i^t\right]\right) \leq \left(1-\epsilon^2\right) Var\left(E_t\left[x\right]\right) \\ \leq \left(1-\epsilon^2\right) Var\left(x\right) = \left(1-\epsilon^2\right) Var\left(\bar{E}_{t_2} ... \bar{E}_{t_k}[\Theta_T]\right). \end{aligned}$$

Iterating the previous condition proves (103).

Condition (103) provides an upper bound for the variance of all k-th order belief. Together with the fact that $\phi_T^* = s_{T,T}$ and, for any random variables X, Y and scalars $a, b \geq 0$,

$$\begin{aligned} Var(aX+bY) &= a^2 Var(X) + 2ab Cov\left(X,Y\right) + b^2 Var\left(Y\right) \\ &\leq a^2 Var(X) + 2ab\sqrt{Var\left(X\right)Var\left(Y\right)} + b^2 Var\left(Y\right) \\ &= \left(a\sqrt{Var\left(X\right)} + b\sqrt{Var\left(Y\right)}\right)^2, \end{aligned} \tag{104}$$

we have

$$\begin{aligned} \left(\frac{\phi_T}{\phi_T^*}\right)^2 &= \left(\frac{Cov\left(a_0, \bar{E}_0\left[\Theta_T\right]\right)}{\phi_T^* Var\left(\bar{E}_0\left[\Theta_T\right]\right)}\right)^2 \leq \frac{Var\left(a_0\right)}{\left[\phi_T^*\right]^2 Var\left(\bar{E}_0[\Theta_T]\right)} \\ &\leq \frac{1}{Var\left(\bar{E}_0[\Theta_T]\right)}\left[\sum_{k=1}^{T}\left(\frac{\chi_{k,T}}{s_{T,T}}\left(1-\epsilon^2\right)^{\frac{k}{2}}\sqrt{Var\left(\Theta_T\right)}\right)\right]^2 \\ &= \left[\sum_{k=1}^{T}\left(\frac{\chi_{k,T}}{s_{T,T}}\left(1-\epsilon^2\right)^{\frac{k}{2}}\right)\right]^2 \frac{Var\left(\Theta_T\right)}{Var\left(\bar{E}_0[\Theta_T]\right)}. \end{aligned} \tag{105}$$

Further note that, for any $\vartheta > 0$, there exists $h \in \mathbb{N}_+$ such that $\frac{\left(1-\epsilon^2\right)^{\frac{h}{2}}}{1-\left(1-\epsilon^2\right)^{\frac{1}{2}}} \leq \frac{\vartheta}{2}$. From Theorem 2, there

[50] We use the fact that for any random variable X, and any information set I, $Var\left(E\left[X|I\right]\right) \leq Var\left(X\right)$. We also use the fact that $\bar{E}_t\left[\cdot\right] = E\left[E\left[\cdot|\omega_i^t\right]|\Omega_t\right]$, where Ω_t is the cross sectional distribution of ω_i^t at time t,

exists $T^* \in \mathbb{N}_+$ such that, for all $T \geq T^*$, $\sum_{k=1}^{h-1} \frac{\chi_{k,T}}{s_{T,T}} \leq \frac{\vartheta}{2}$. As a result, for all $T \geq \max\{T^*, h\}$,

$$\sum_{k=1}^{T} \frac{\chi_{k,T}}{s_{T,T}} \left(1-\epsilon^2\right)^{\frac{k}{2}} \leq \sum_{k=1}^{h-1} \frac{\chi_{k,T}}{s_{T,T}} + \sum_{k=h}^{T} \left(1-\epsilon^2\right)^{\frac{k}{2}} \leq \frac{\vartheta}{2} + \frac{\left(1-\epsilon^2\right)^{\frac{h}{2}}}{1-(1-\epsilon^2)^{\frac{1}{2}}} \leq \vartheta.$$

This proves

$$\sum_{k=1}^{T} \left(\frac{\chi_{k,T}}{s_{T,T}} \left(1-\epsilon^2\right)^{\frac{k}{2}} \right) \to 0 \text{ as } T \to +\infty.$$

Together with (105) and the fact that $Var\left(\bar{E}_0[\Theta_T]\right) \geq \epsilon$, the proof of Proposition 9 is completed.

Appendix C. Additional Results for the New Keynesian Model

In this Appendix, we provide a few additional results regarding the application of our insights in the context of a liquidity trap. We first explain how our results regarding the forward-guidance puzzle can be understood under the lenses of a discounted Euler condition and a discounted NKPC, and draw certain connections to the literature. We next show how our insights help lessen the paradox of flexibility. We finally show that all the results of Section 5 extend to the new type of beauty contest seen in condition (22).

Discounted Euler Condition and Discounted NKPC. Proposition 5 has already indicated how the lack of common knowledge is akin to introducing additional discounting in the forward-looking equations of a macroeconomic model. We now illustrate how this helps recast our results regarding forward guidance and fiscal multipliers under the lenses of a discounted Euler condition and a discounted NKPC.

For the present purposes, we make a minor modification to the setting used in Section 6: for $t \leq T$, we let the firms lack knowledge of the concurrent level of marginal cost. For simplicity, we also let the firms and the consumers face the same level of friction, that is, we set $\lambda_c = \lambda_f = \lambda$. These modifications are not strictly needed but sharpen the representation offered below.[51]

Proposition 10 *The power of forward guidance in the absence of common knowledge, ϕ_T, is the same as that in a representative-agent variant in which the Euler condition and the NKPC are modified as follows, for all $t \leq T-1$:*

$$\tilde{y}_t = -\sigma \left\{ \tilde{R}_t - \lambda E_t \left[\tilde{\pi}_{t+1} \right] \right\} + M_c E_t \left[\tilde{y}_{t+1} \right] \quad (106)$$

$$\tilde{\pi}_t = \kappa' \tilde{y}_t + \beta M_f E_t \left[\tilde{\pi}_{t+1} \right] + \varkappa \tilde{\mu}_t, \quad (107)$$

where $M_c \equiv \beta + (1-\beta)\lambda \in (\beta, 1]$, $M_f \equiv \theta + (1-\theta)\lambda \in (\theta, 1]$, and $\kappa' \equiv \kappa\lambda$.[52]

This result, which is analogous to Proposition 5 in our abstract setting, maps the incomplete-information ϕ_T of the economy under consideration to the complete-information ϕ_T^* of a *variant* economy, in which the Euler condition and the NKPC have been "discounted" in the manner described above. When we remove common knowledge, it is *as if* the representative consumer discounts her expectations of next period's aggregate income and inflation by a factor equal to, respectively, M_c and λ; and it is *as if* the representative firm discounts the future inflation by a factor equal to M_f.[53]

Consider first the Euler condition. When β is close to 1, the discount on future consumption, M_c, is close to 1, even if λ is close to zero. This underscores that the multiplier inside the demand block—which gets attenuated by the absence of common knowledge—is weak in the textbook version of the New Keynesian model. As mentioned in the main text, short horizons, counter-cyclical precautionary savings, and

[51] Without these modifications, the obtained representation is a bit less elegant, but the essence remains the same; see Proposition 10 in the first NBER version of our paper, Angeletos and Lian (2016*b*).

[52] To be precise, condition (106) holds with $M_c = 1$ for $t = T-1$.

[53] The change in the slope of the NKPC, from κ to κ', is of relative little interest to us, because the effect of the informational friction through this slope cannot be identified separately from that of a higher Frisch elasticity or less steep marginal costs.

feedback effects between housing prices and consumer spending tend to reinforce this multiplier, thereby also increasing the discounting caused by the absence of common knowledge. Also note that, while future consumption is discounted by M_c, future inflation is discounted by λ. Clearly, this can have a significant effect on the joint dynamics of spending and inflation even when M_c is close to 1.

Consider next the NKPC. For the textbook parameterization of the degree of price stickiness (meaning a price revision rate, $1-\theta$, equal to $1/3$), the effective discount factor, M_f, falls from 1 to $.9$ as we move from common knowledge ($\lambda = 1$) to the level of imperfection assumed in our numerical example ($\lambda = .75$). The magnitude of this discount helps explain the sizable effects seen in Figure 1. Under the parameterization we consider, the *actual* response of inflation to news about future demand is greatly reduced relative the common-knowledge benchmark. The fact that the average consumer underestimates the inflation response, as well as the spending of other consumers, reinforces this effect and helps further attenuate the feedback loop between inflation and spending.

Relation to Gabaix (2016) and McKay, Nakamura and Steinsson (2016*a*). Related forms of discounting appear in McKay, Nakamura and Steinsson (2016*a*), for the Euler condition, and in Gabaix (2016), for both the Euler condition and the NKPC. In this regards, these papers and ours are complementary to one another. However, the underlying theory and its empirical manifestations are different.

McKay, Nakamura and Steinsson (2016*a*) obtain a discounted Euler condition at the aggregate level by introducing a specific combination of heterogeneity and market incompleteness that forces some agents to hit their borrowing constraints and breaks the individual-level Euler condition. This theory therefore ties the resolution of forward guidance to microeconomic evidence about the response of individual consumption to idiosyncratic shocks. By contrast, our theory ties the resolution of forward guidance to survey evidence about the response of average forecast errors to the underlying policy news. The two theories can therefore be quantified independently from one another—and it's an open question which is one is more relevant in the context of forward guidance.

Gabaix (2016) on the other hand, assumes two kinds of friction. The first is that agents are less responsive to any variation in interest rates and incomes due to "sparsity" (a form of adjustment cost). The second is that agents underestimate the response of future *aggregate* outcomes to exogenous shocks. The first is of purely decision-theoretic nature and, as the one in McKay, Nakamura and Steinsson (2016*a*), amounts to a distortion of the individual-level Euler condition. The second is more closely related to the one we have obtained here: by anchoring expectations of *aggregate* outcomes, it gives rise to discounting only at the aggregate level. In this regard, Gabaix's theory and ours have a similar empirical implication: they both let the average forecast of future inflation and income respond less than the complete-information, rational-expectations, benchmark. Yet, our theory makes the following distinct prediction, which is consistent with the evidence in Coibion and Gorodnichenko (2012): the forecast errors, and the associated discounting, ought to decrease as time passes, agents accumulate more information, and higher-order beliefs converge to first-order beliefs.

Proof of Proposition 10. Let us first focus on the incomplete-information ϕ_T. When firms lack common knowledge of the concurrent level of marginal cost, condition (62) continues to hold but condition (63) becomes, for any $t \leq T$,

$$\tilde{\pi}_t - \tilde{\pi}_t^{trap} = \kappa \sum_{k=0}^{T-t} (\beta\theta)^k \, \bar{E}_t^f \left[\tilde{y}_{t+k} - \tilde{y}_{t+k}^{trap} \right] + \tfrac{1-\theta}{\theta} \sum_{k=1}^{T-t} (\beta\theta)^k \, \bar{E}_t^f \left[\tilde{\pi}_{t+k} - \tilde{\pi}_{t+k}^{trap} \right].$$

Slightly different from conditions (64) and (65),[54] we can find functions $\phi, \omega : (0,1] \times \mathbb{N} \to \mathbb{R}_{\geq 0}$ such that, for any $t \leq T-1$,

$$\tilde{y}_t - \tilde{y}_t^{trap} = -\phi(\lambda, T-t)\, \bar{E}_t^c[\tilde{R}_T], \tag{108}$$

$$\tilde{\pi}_t - \tilde{\pi}_t^{trap} = -\omega(\lambda, T-t)\, \bar{E}_t^f[\tilde{R}_T], \tag{109}$$

where $\phi(\lambda, 1) = \sigma(1+\sigma\lambda\kappa)$, $\omega(\lambda, 1) = \kappa\lambda\sigma(1+\sigma\lambda\kappa) + \kappa\beta(\theta + (1-\theta)\lambda)\sigma$, and for $t \leq T-2$,

$$\phi(\lambda, T-t) = (\beta + (1-\beta)\lambda)\,\phi(\lambda, T-t-1) + \sigma\lambda\omega(\lambda, T-t-1), \tag{110}$$

$$\omega(\lambda, T-t) = \beta(\theta + (1-\theta)\lambda)\,\omega(\lambda, T-t-1) + \kappa\lambda\phi(\lambda, T-t). \tag{111}$$

We now derive the complete-information ϕ_T^* and ω_T^* of a *variant* economy, where they denote how the output and inflation at $t = 0$ responds to shocks to the representative agent's belief about $\tilde{R}_T$ at $t = 0$. From conditions (106), (107) and footnote 52 in the appendix, we have $\phi_1^* = \sigma(1+\sigma\lambda\kappa)$, $\omega_1^* = \kappa\lambda\sigma(1+\sigma\lambda\kappa) + \kappa\beta(\theta + (1-\theta)\lambda)\sigma$, and, for $t \leq T-2$,

$$\phi_{T-t}^* = (\beta + (1-\beta)\lambda)\,\phi_{T-t-1}^* + \sigma\lambda\omega_{T-t-1}^*, \tag{112}$$

$$\omega_{T-t}^* = \beta(\theta + (1-\theta)\lambda)\,\omega_{T-t-1}^* + \kappa\lambda\phi_{T-t}^*. \tag{113}$$

The previous conditions coincide with conditions (110) and (111), and prove Proposition 10.

On the Paradox of Flexibility. We now consider the implications of our insights for the paradox of flexibility. In the standard model, the power of forward guidance and the fiscal multiplier vis-a-vis future government spending increase with the degree of price flexibility: ϕ_T^* increases with κ.[55] This property is directly related to the "paradox of flexibility" (Eggertsson and Krugman, 2012). The next result proves, in effect, that the mechanism identified in our paper helps diminish this paradox as well.

Proposition 11 (Price Flexibility) *Let ϕ_T^* be the scalar characterized in either Proposition 7 or Proposition 8 and set $\lambda_f = 1$. We have $\frac{\partial \phi_T^*}{\partial \kappa} > 0$ and $\frac{\partial}{\partial \lambda_c}\left(\frac{\partial \phi_T^*}{\partial \kappa}\right) > 0$. That is, the power of forward guidance and the fiscal multiplier vis-a-vis future government spending increase with the degree of price flexibility, but at a rate that*

[54] There are two differences compared to conditions (64) and (65). First, as we impose $\lambda_c = \lambda_f = \lambda$ here, ϕ and ω are functions of λ, the common parameter characterizing the degree of information friction. Second, as firms lack common knowledge of the concurrent level of marginal cost, it is easier to let ω measure how inflation as a whole responds to the average firm's belief about $\tilde{R}_T$.

[55] Whenever we vary κ, we vary θ while keeping the Frisch elasticity constant, which means that variation in κ maps one-to-one to variation in the degree of price flexibility.

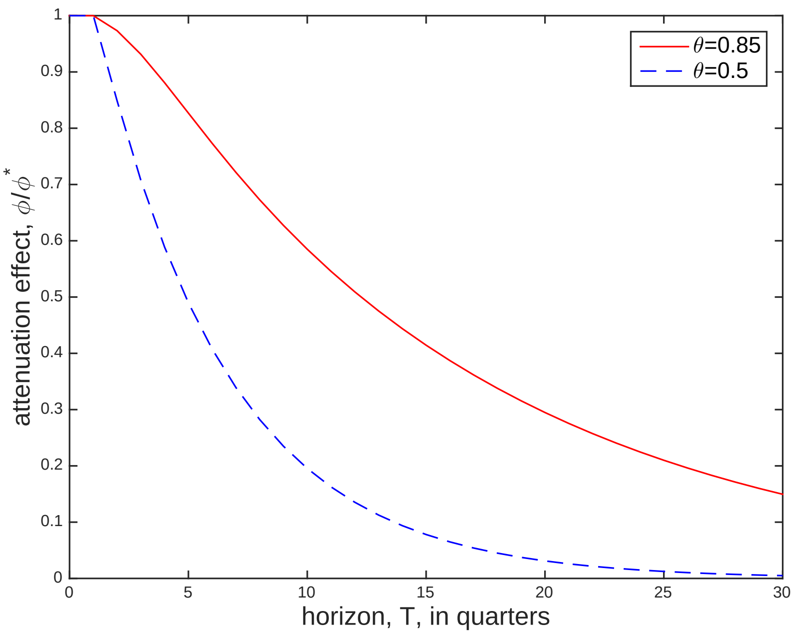

Figure 2: Varying the degree of price flexibility.

is slower the greater the departure from common knowledge.

This finding is an example of how lack of common knowledge reduces the paradox of flexibility more generally. In the standard model, a higher degree of price flexibility raises the GE effects of all kinds of demand shocks—whether these come in the form of forward guidance, discount rates, or borrowing constraints—because it intensifies the feedback loop between aggregate spending and inflation. By intensifying this kind of macroeconomic complementarity, however, a higher degree of price flexibility also raises the relative importance of higher-order beliefs, which in turn contributes to stronger attenuation effects of the type we have documented in this paper. In a nutshell, the very same mechanism that creates the paradox of flexibility within the New Keynesian framework also helps contain that paradox once we relax the common-knowledge assumption of that framework.

Note that we have proved the above result only under the restriction $\lambda_f = 1$, which means that only the consumers lack common knowledge. Whenever $\lambda_f < 1$, there is a conflicting effect, which is that higher price flexibility reduces the strategic complementarity that operates within the supply block, thereby also reducing the role of λ_f itself. For the numerical example considered earlier, however, the overall effect of higher price flexibility is qualitatively the same whether $\lambda_f = 1$ or $\lambda_f = \lambda_c$.

We illustrate this in Figure 2. We let $\lambda_f = \lambda_c = 0.75$, use the same parameter values as those used in Figure 1, and plot the relation between the ratio ϕ_T/ϕ_T^* and the horizon T under two values for θ. The solid red line corresponds to a higher value for θ, while the dashed blue line corresponds to a lower value for θ, that is, to more price flexibility. As evident in the figure, more price flexibility maps, not only to a lower ratio ϕ_T/ϕ_T^* (i.e., stronger attenuation) for any given T, but also to a more rapid decay in that ratio as we raise T.

Proof of Proposition 11. Consider first the environment studied in Section 6 and let us study the cross-partial derivative of the power of forward guidance with respect to κ and λ_c. To simplify notation, we use ϕ_τ and ϖ_τ as shortcuts for, respectively, $\phi(\lambda_c, \lambda_f, \tau)$ and $\varpi(\lambda_c, \lambda_f, \tau)$, where the functions ϕ and ϖ are defined as in the proof of Proposition 7. From conditions (66), we have

$$\frac{\partial \phi_1}{\partial \kappa} = \frac{\partial \phi(\lambda_c, 1, 1)}{\partial \kappa} = \sigma^2 > 0 \qquad \text{and} \qquad \frac{\partial \varpi_1}{\partial \kappa} = \frac{\partial \varpi(\lambda_c, 1, 1)}{\partial \kappa} = \sigma\beta > 0. \tag{114}$$

For any $\tau \geq 2$, when $\lambda_f = 1$, conditions (71) and (72) become

$$\phi_\tau = (\beta + (1 - \beta + \sigma\kappa)\lambda_c)\phi_{\tau-1} + \sigma\varpi_{\tau-1} \qquad \text{and} \qquad \varpi_\tau = \kappa\beta\lambda_c\phi_{\tau-1} + \beta\varpi_{\tau-1}.$$

As a result, for all $\tau \geq 2$, we have

$$\frac{\partial \phi_\tau}{\partial \kappa} = (\beta + (1 - \beta + \sigma\kappa)\lambda_c)\frac{\partial \phi_{\tau-1}}{\partial \kappa} + \sigma\lambda_c\phi_{\tau-1} + \sigma\frac{\partial \varpi_{\tau-1}}{\partial \kappa}, \tag{115}$$

$$\frac{\partial \varpi_\tau}{\partial \kappa} = \kappa\beta\lambda_c\frac{\partial \phi_{\tau-1}}{\partial \kappa} + \beta\lambda_c\phi_{\tau-1} + \beta\frac{\partial \varpi_{\tau-1}}{\partial \kappa}. \tag{116}$$

From conditions (114), (115) and (116), $\frac{\partial \phi_\tau}{\partial \kappa}$ and $\frac{\partial \varpi_\tau}{\partial \kappa}$ are strictly positive for any $\tau \geq 1$ by induction. Moreover, from conditions (66), (114), (115) and (116), we have that $\frac{\partial \phi_2}{\partial \kappa}$ and $\frac{\partial \varpi_2}{\partial \kappa}$ are strictly increasing in λ_c. Then, from conditions (115), (116) and the fact that ϕ_τ itself is strictly increasing in λ_c for all $\tau \geq 2$, we have $\frac{\partial \phi_\tau}{\partial \kappa}$ and $\frac{\partial \varpi_\tau}{\partial \kappa}$ are strictly increasing in λ_c for all $\tau \geq 2$ by induction.

Consider now the environment studied in Section 7 and let us study the cross-partial derivative of the relevant fiscal multiplier with respect to κ and λ_c when $\lambda_f = 1$. From the proof of Proposition 8, similarly to conditions (114), (115) and (116), we have

$$\frac{\partial \phi_{g,1}}{\partial \kappa} = \sigma(1 - \Omega_c) > 0 \qquad \text{and} \qquad \frac{\partial \varpi_{g,1}}{\partial \kappa} = \beta(1 - \Omega_c) > 0,$$
$$\frac{\partial \phi_{g,\tau}}{\partial \kappa} = (\beta + (1 - \beta + \sigma\kappa\Omega_c)\lambda_c)\frac{\partial \phi_{g,\tau-1}}{\partial \kappa} + \sigma\Omega_c\lambda_c\phi_{g,\tau-1} + \sigma\frac{\partial \varpi_{g,\tau-1}}{\partial \kappa},$$
$$\frac{\partial \varpi_{g,\tau}}{\partial \kappa} = \kappa\beta\lambda_c\Omega_c\frac{\partial \phi_{g,\tau-1}}{\partial \kappa} + \beta\lambda_c\Omega_c\phi_{g,\tau-1} + \beta\frac{\partial \varpi_{g,\tau-1}}{\partial \kappa}.$$

The result then follows from the same argument as before.

Extension of Lemma 2 and Theorems 1 and 2. Here we show that Lemma 2, Theorem 2 and, by implication, Theorem 1 extend to the kind of multi-layer beauty contest seen in condition (22) of Lemma 3.

Similar to Section 5, we impose Assumption 1. Similar to the proof of Lemma 2, we can find positively-valued coefficients $\{\chi_{h,\tau}\}_{\tau\geq 1, 1\leq h\leq \tau}$, such that, for any $t \leq T-1$,

$$\tilde{y}_t - \tilde{y}_t^{trap} = \sum_{h=1}^{T-t}\left\{\chi_{h,T-t}\bar{E}_t^h\left[\tilde{R}_T\right]\right\}, \tag{117}$$

with $\tilde{y}_t^{trap}$ defined as in the proof of Proposition 6 and

$$\chi_{1,\tau} = \sigma\left(1+\tau\sigma\kappa\right)\beta^{\tau-1} \quad \forall\tau \geq 1, \tag{118}$$

$$\chi_{k,\tau} = \sum_{l=1}^{\tau-k+1}\left(1-\beta+l\sigma\kappa\right)\beta^{l-1}x_{k-1,\tau-l} \quad \forall k\geq 2 \; and \; \tau \geq k. \tag{119}$$

We can then characterize the combined effect of beliefs of order up to k on spending, $s_{k,\tau}$, as[56]

$$s_{k,\tau} = \sigma\left(1+\tau\sigma\kappa\right)\beta^{\tau-1} + \sum_{l=1}^{\tau-1}\left(1-\beta+l\sigma\kappa\right)\beta^{l-1}s_{k-1,\tau-l} \quad \forall k\geq 1 \; and \; \tau\geq 1. \tag{120}$$

Let $d_\tau = s_{\tau,\tau}$ denote the combined effect of beliefs of all different orders on spending. Similar to condition (18), $d_\tau = \phi^*_\tau$. Following the proof of Proposition 6, we have

$$d_0 = \sigma \qquad \text{and} \qquad d_1 = \sigma\left(1+\sigma\kappa\right),$$

$$\frac{d_\tau}{d_{\tau-1}} + \beta\frac{d_{\tau-2}}{d_{\tau-1}} = 1+\beta+\sigma\kappa \quad \forall\tau\geq 2. \tag{121}$$

Now we prove $s_{k,\tau}$ satisfies an inequality with a similar form as condition (121):

$$\frac{s_{k,\tau}}{s_{k,\tau-1}} + \beta\frac{s_{k,\tau-2}}{s_{k,\tau-1}} \leq 1+\beta+\sigma\kappa \quad \forall\tau\geq 3 \; and \; k\geq 1. \tag{122}$$

From condition (120), we have

$$\beta s_{k,\tau-1} = \sigma\left(1+\left(\tau-1\right)\sigma\kappa\right)\beta^{\tau-1} + \sum_{l=2}^{\tau-1}\left(1-\beta+\left(l-1\right)\sigma\kappa\right)\beta^{l-1}s_{k-1,\tau-l} \quad \forall k\geq 1 \; and \; \tau\geq 2.$$

As a result, we have

$$s_{k,\tau} = \beta s_{k,\tau-1} + \left(1-\beta\right)s_{k-1,\tau-1} + \sigma^2\kappa\beta^{\tau-1} + \sigma\kappa\sum_{l=1}^{\tau-1}\beta^{l-1}s_{k-1,\tau-l} \quad \forall k\geq 1 \; and \; \tau\geq 2,$$

$$\beta s_{k,\tau-1} = \beta^2 s_{k,\tau-2} + \beta\left(1-\beta\right)s_{k-1,\tau-2} + \sigma^2\kappa\beta^{\tau-1} + \sigma\kappa\sum_{l=2}^{\tau-1}\beta^{l-1}s_{k-1,\tau-l} \quad \forall k\geq 1 \; and \; \tau\geq 3.$$

Using the previous two conditions, we have, for all $k\geq 1$ and $\tau\geq 3$,

$$s_{k,\tau} + \beta^2 s_{k,\tau-2} + \beta\left(1-\beta\right)s_{k-1,\tau-2} = 2\beta s_{k,\tau-1} + \left(1-\beta+\sigma\kappa\right)s_{k-1,\tau-1},$$

$$s_{k,\tau} + \beta s_{k,\tau-2} = \left(1+\beta+\sigma\kappa\right)s_{k,\tau-1} + \beta\left(1-\beta\right)\chi_{k,\tau-2} - \left(1-\beta+\sigma\kappa\right)\chi_{k,\tau-1}. \tag{123}$$

[56]Similar to the proof of Theorem 2, for notation simplicity, we extend the definition of $s_{h,\tau} = \sum_{r=1}^{h}\chi_{r,\tau}$ for all $h > \tau$. As for $h > \tau$, $\chi_{h,\tau} = 0$, we have $s_{h,\tau} = s_{\tau,\tau}$ for all $h > \tau$. We also define $s_{0,\tau} = 0$ for all $\tau \geq 1$.

To prove (122), we only need to prove:

$$\beta\left(1-\beta\right)\chi_{k,\tau-2}\leq\left(1-\beta+\sigma\kappa\right)x_{k,\tau-1}\quad\forall k\geq1\ and\ \tau\geq3. \tag{124}$$

In fact, we prove the following stronger result:

$$\beta\chi_{k,\tau-2}\leq x_{k,\tau-1}\quad\forall k\geq1\ and\ \tau\geq3. \tag{125}$$

From condition (118), we know that (125) is true for $k=1$ and $\tau\geq3$. From condition (119), we know that

$$\chi_{k,\tau-1}=\sum_{l=1}^{\tau-k}\left(1-\beta+\sigma l\kappa\right)\beta^{l-1}x_{k-1,\tau-1-l}\quad\forall k\geq2\ and\ \tau\geq k+1, \tag{126}$$

$$\beta\chi_{k,\tau-2}=\sum_{l=2}^{\tau-k}\left(1-\beta+\sigma\left(l-1\right)\kappa\right)\beta^{l-1}x_{k-1,\tau-1-l}\quad\forall k\geq2\ and\ \tau\geq k+2.$$

This proves $\beta\chi_{k,\tau-2}\leq x_{k,\tau-1}$ for $k\geq2$ and $\tau\geq k+2$. Together with the fact that, $\chi_{k,\tau-2}=0\ \forall k\geq\tau-1$, we prove (125) and thus (124). This finishes the proof of (122).

Based on (121) and (122), we can then establish a result akin to Theorem 2. That is, for any given $k\geq1$ and $\tau\geq k$, the relative contribution of the first k orders, $\frac{s_{k,\tau}}{s_{\tau,\tau}}=\frac{s_{k,\tau}}{d_\tau}$, strictly decreases with τ.

First, note that, for any given $k\geq1$,$1=\frac{s_{k,k}}{d_k}>\frac{s_{k,k+1}}{d_{k+1}}$, because $x_{k+1,k+1}>0$. Then, we can proceed by induction on $\tau\geq k$, for any fixed $k\geq1$. If we have $\frac{s_{k,\tau}}{d_\tau}>\frac{s_{k,\tau+1}}{d_{\tau+1}}$ for some $\tau\geq k$, we have $\frac{s_{k,\tau}}{s_{k,\tau+1}}>\frac{d_\tau}{d_{\tau+1}}$. Using (121) and (122), we have $\frac{s_{k,\tau+2}}{s_{k,\tau+1}}<\frac{d_{\tau+2}}{d_{\tau+1}}$, and thus $\frac{s_{k,\tau+1}}{d_{\tau+1}}>\frac{s_{k,\tau+2}}{d_{\tau+2}}$. This completes the proof that, for any $k\geq1$ and any $\tau\geq k$, the ratio $\frac{s_{k,\tau}}{s_{\tau,\tau}}$, strictly decreases with the horizon τ.

Finally, we prove that, for any $k\geq1$,

$$\frac{s_{k,\tau}}{s_{\tau,\tau}}\to0,\quad as\ \tau\to\infty. \tag{127}$$

In other words, we want to prove the relative contribution of the first k orders of beliefs to aggregate spending converges to zero when the horizon τ goes to infinity.

First note that, from condition (120), we have $s_{1,\tau}=\sigma\left(1+\sigma\tau\kappa\right)\beta^{\tau-1}\to0,$ as $\tau\to+\infty$. From the proof of Proposition 6, we know $s_{\tau,\tau}=d_\tau=\phi^*_\tau\geq\sigma$. As a result, (127) is true for $k=1$.

If there exists $k\geq2$ such that (127) does not hold, we let $k^*\geq2$ denote the smallest of such k. Then, (127) holds for $1\leq k\leq k^*-1$. Because we already prove that $\frac{s_{k^*,\tau}}{s_{\tau,\tau}}\geq0$ is decreasing with the horizon τ, there exists $0<\Gamma<1$ such that $\frac{s_{k^*,\tau}}{s_{\tau,\tau}}=\frac{s_{k^*,\tau}}{\phi^*_\tau}\to\Gamma$ as $\tau\to\infty$. As a result, $\frac{s_{k^*,\tau}}{\phi^*_\tau}\frac{\phi^*_{\tau-1}}{s_{k^*,\tau-1}}\to1$ as $\tau\to\infty$. Because we already prove that, in the proof of Proposition 6, $\frac{\phi^*_\tau}{\phi^*_{\tau-1}}\to\Gamma^*$, we have

$$\frac{s_{k^*,\tau}}{s_{k^*,\tau-1}}\to\Gamma^*\quad and\quad\frac{s_{k^*,\tau-2}}{s_{k^*,\tau-1}}\to\frac{1}{\Gamma^*}\quad as\ \tau\to\infty. \tag{128}$$

Note that since $s_{k^*,\tau}=s_{k^*-1,\tau}+\chi_{k^*,\tau}$ and $\frac{s_{k^*-1,\tau}}{s_{\tau,\tau}}\to0\ as\ \tau\to\infty$, we have $\frac{\chi_{k^*,\tau}}{s_{\tau,\tau}}=\frac{\chi_{k^*,\tau}}{\phi^*_\tau}\to\Gamma\ as\ \tau\to\infty.$

As a result,

$$\frac{\chi_{k^*,\tau}}{s_{k^*,\tau}} = \frac{\chi_{k^*,\tau}}{\phi_\tau^*}\frac{\phi_\tau^*}{s_{k^*,\tau}} \to 1 \quad as\ \tau \to \infty. \tag{129}$$

Now we prove a stronger version of (122)

$$\frac{s_{k,\tau}}{s_{k,\tau-1}} + \beta\frac{s_{k,\tau-2}}{s_{k,\tau-1}} + \sigma\kappa\frac{\chi_{k,\tau-1}}{s_{k,\tau-1}} \leq 1 + \beta + \sigma\kappa \quad \forall\tau \geq 3\ and\ k \geq 1. \tag{130}$$

This comes from the fact that (125) can be written as

$$\beta\left(1-\beta\right)\chi_{k,\tau-2} + \sigma\kappa\chi_{k,\tau-1} \leq \left(1-\beta+\sigma\kappa\right)x_{k,\tau-1} \quad \forall\tau \geq 3\ and\ k \geq 1. \tag{131}$$

Using (61), (128) and (129), we have

$$\frac{s_{k^*,\tau}}{s_{k^*,\tau-1}} + \beta\frac{s_{k^*,\tau-2}}{s_{k^*,\tau-1}} + \sigma\kappa\frac{\chi_{k^*,\tau-1}}{s_{k^*,\tau-1}} \to \Gamma^* + \beta\frac{1}{\Gamma^*} + \sigma\kappa = 1 + \beta + 2\sigma\kappa \quad as\ \tau \to \infty.$$

This contradicts (130) when $\kappa > 0$ and proves (127). This finishes the proof of the result akin to Theorem 2. Together with Proposition 3, we then establish the "horizon effect" akin to Theorem 1. Similarly, Proposition 4 also holds here.

References

Allen, Franklin, Stephen Morris, and Hyun Song Shin. (2006) "Beauty Contests and Iterated Expectations in Asset Markets." *Review of financial Studies*, 19(3): 719–752.

Andrade, Philippe, Gaballo Gaetano, Eric Mengus, and Benoit Mojon. (2015) "Forward Guidance and Heterogeneous Beliefs." *Banque De France mimeo*.

Angeletos, George-Marios, and Alessandro Pavan. (2007) "Efficient Use of Information and Social Value of Information." *Econometrica*, 75(4): 1103–1142.

Angeletos, George-Marios, and Chen Lian. (2016*a*) "Dampening General Equilibrium: From Micro to Macro." *NBER Working Paper No. 23379*.

Angeletos, George-Marios, and Chen Lian. (2016*b*) "Forward Guidance without Common Knowledge." *NBER Working Paper No. 22785*.

Angeletos, George-Marios, and Chen Lian. (2016*c*) "Incomplete Information in Macroeconomics: Accommodating Frictions in Coordination." *Handbook of Macroeconomics*, 2: 1065–1240.

Angeletos, George-Marios, and Jennifer La'O. (2010) "Noisy Business Cycles." In *NBER Macroeconomics Annual 2009, Volume 24*. 319–378. University of Chicago Press.

Angeletos, George-Marios, and Jennifer La'O. (2013) "Sentiments." *Econometrica*, 81(2): 739–779.

Bacchetta, Philippe, and Eric van Wincoop. (2006) "Can Information Heterogeneity Explain the Exchange Rate Determination Puzzle?" *The American Economic Review*, 96(3): 552.

Bergemann, Dirk, and Stephen Morris. (2013) "Robust Predictions in Games with Incomplete Information." *Econometrica*, 81(4): 1251–1308.

Campbell, Jeffrey R, Charles L Evans, Jonas DM Fisher, and Alejandro Justiniano. (2012) "Macroeconomic Effects of Federal Reserve Forward Guidance." *Brookings Papers on Economic Activity*, 2012(1): 1–80.

Chahrour, Ryan, and Gaetano Gaballo. (2017) "Learning from Prices: Amplification and Business Fluctuations." *Boston College Mimeo*.

Christiano, Lawrence, Martin Eichenbaum, and Sergio Rebelo. (2011) "When Is the Government Spending Multiplier Large?" *Journal of Political Economy*, 119(1): 78–121.

Chung, Hess, Edward Herbst, and Michael T Kiley. (2015) "Effective Monetary Policy Strategies in New Keynesian Models: A Reexamination." *NBER Macroeconomics Annual*, 29(1): 289–344.

Cochrane, John. (2016) "Do Higher Interest Rates Raise or Lower Inflation?" *Stanford University mimeo*.

Cochrane, John. (2017) "The New-Keynesian Liquidity Trap." *Journal of Monetary Economics*, 92: 47–63.

Coibion, Olivier, and Yuriy Gorodnichenko. (2012) "What Can Survey Forecasts Tell Us about Information Rigidities?" *Journal of Political Economy*, 120(1): 116–159.

Coibion, Olivier, and Yuriy Gorodnichenko. (2015) "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts." *American Economic Review*, 105(8): 2644–78.

Del Negro, Marco, Marc P Giannoni, and Christina Patterson. (2015) "The Forward Guidance Puzzle." *FRB of New York mimeo*.

Eggertsson, Gauti B, and Paul Krugman. (2012) "Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach." *The Quarterly Journal of Economics*, 127(3): 1469–1513.

Farhi, Emmanuel, and Iván Werning. (2017) "Monetary Policy, Bounded Rationality, and Incomplete Markets." *NBER Working Paper No. 23281*.

Gabaix, Xavier. (2016) "A Behavioral New Keynesian Model." *NBER Working Paper No. 22954*.

Galí, Jordi. (2008) *Monetary policy, Inflation, and the Business Cycle: an Introduction to the New Keynesian Framework and its Applications.* Princeton University Press.

Galí, Jordi, J David López-Salido, and Javier Vallés. (2007) "Understanding the Effects of Government Spending on Consumption." *Journal of the European Economic Association*, 5(1): 227–270.

Garcıa-Schmidt, Mariana, and Michael Woodford. (2015) "Are Low Interest Rates Deflationary? A Paradox of Perfect-Foresight Analysis." *NBER Working Paper No. 21614*.

Grossman, Sanford J, and Joseph E Stiglitz. (1980) "On the Impossibility of Informationally Efficient Markets." *The American Economic Review*, 70(3): 393–408.

Kaplan, Greg, Benjamin Moll, and Giovanni L Violante. (2016) "Monetary Policy According to HANK." *NBER Working Paper No. 21897*.

Khaw, Mel Win, Luminita Stevens, and Michael Woodford. (2016) "Discrete Adjustment to a Changing Environment: Experimental Evidence." *NBER Working Paper No. 22978*.

Kiley, Michael T. (2016) "Policy Paradoxes in the New Keynesian Model." *Review of Economic Dynamics*, 21: 1–15.

Lucas, Robert E. (1972) "Expectations and the Neutrality of Money." *Journal of Economic Theory*, 4(2): 103–124.

Luo, Yulei, Jun Nie, Gaowang Wang, and Eric R. Young. (2017) "Rational inattention and the dynamics of consumption and wealth in general equilibrium." *Journal of Economic Theory*, 172(C): 55 – 87.

Mackowiak, Bartosz, and Mirko Wiederholt. (2009) "Optimal Sticky Prices under Rational Inattention." *American Economic Review*, 99(3): 769–803.

Mackowiak, Bartosz, Filip Matejka, and Mirko Wiederholt. (2017) "The Rational Inattention Filter." *Goethe University Frankfurt mimeo*.

Mankiw, N Gregory, and Ricardo Reis. (2002) "Sticky Information Versus Sticky Prices: a Proposal to Replace the New Keynesian Phillips Curve." *The Quarterly Journal of Economics*, 117(4): 1295–1328.

McKay, Alisdair, Emi Nakamura, and Jón Steinsson. (2016*a*) "The Discounted Euler Equation: A Note." *NBER Working Paper No. 22129*.

McKay, Alisdair, Emi Nakamura, and Jón Steinsson. (2016*b*) "The Power of Forward Guidance Revisited." *American Economic Review*, 106(10): 3133–3158.

Morris, Stephen, and Hyun Song Shin. (1998) "Unique Equilibrium in a Model of Self-fulfilling Currency Attacks." *American Economic Review*, 587–597.

Morris, Stephen, and Hyun Song Shin. (2002) "Social Value of Public Information." *The American Economic Review*, 92(5): 1521–1534.

Morris, Stephen, and Hyun Song Shin. (2003) "Global games: theory and applications." In *Advances in Economics and Econometrics (Proceedings of the Eighth World Congress of the Econometric Society)*. Cambridge University Press.

Morris, Stephen, and Hyun Song Shin. (2006) "Inertia of Forward-looking Expectations." *The American Economic Review*, 152–157.

Nimark, Kristoffer. (2008) "Dynamic Pricing and Imperfect Common Knowledge." *Journal of Monetary Economics*, 55(2): 365–382.

Nimark, Kristoffer. (2017) "Dynamic Higher Order Expectations." *Cornell Univeristy mimeo*.

Samet, Dov. (1998) "Iterated Expectations and Common Priors." *Games and Economic Behavior*, 24(1): 131–141.

Sims, Christopher A. (2003) "Implications of Rational Inattention." *Journal of Monetary Economics*, 50(3): 665–690.

Singleton, Kenneth J. (1987) "Asset Prices in a Time-series Model with Disparately Informed, Competitive Traders." In *New Approaches to Monetary Economics*. , ed. William A. Barnett and Kenneth J. Singleton, 249–272. Cambridge University Press.

Tirole, Jean. (2015) "Cognitive Games and Cognitive Traps." *Toulouse School of Economics mimeo*.

Vellekoop, Nathanael, and Mirko Wiederholt. (2017) "Inflation Expectations and Choices of Households." *Goethe University Frankfurt mimeo*.

Werning, Iván. (2012) "Managing a Liquidity Trap: Monetary and Fiscal Policy." *NBER Working Paper No. 17344*.

Werning, Iván. (2015) "Incomplete Markets and Aggregate Demand." *NBER Working Paper No. 21448*.

Wiederholt, Mirko. (2015) "Empirical Properties of Inflation Expectations and the Zero Lower Bound." *Goethe University Frankfurt mimeo*.

Woodford, Michael. (2003*a*) "Imperfect Common Knowledge and the Effects of Monetary Policy." *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*.

Woodford, Michael. (2003*b*) *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton University Press.

Woodford, Michael. (2011) "Simple Analytics of the Government Expenditure Multiplier." *American Economic Journal: Macroeconomics*, 3(1): 1–35.