Managing Expectations: Instruments vs. Targets[*]

George-Marios Angeletos[†] **Karthik A. Sastry**[‡]

December 9, 2019

Abstract

Should a policymaker offer forward guidance in terms of the intended path for the policy instrument (e.g., keep the Federal Funds rate at a low level for τ periods) or a target for the equilibrium outcome of interest (e.g., bring unemployment down to y%)? We study how the optimal approach depends on plausible bounds on agents' depth of knowledge and rationality. Agents make mistakes in predicting the behavior of others and the GE effects of policy. An optimal communication strategy minimizes the welfare consequences of such mistakes. This is achieved by offering guidance in terms of a sharp outcome target if and only if the GE feedback is strong enough. Our results suggest that central banks should stop talking about interest rates and start talking about unemployment when faced with a prolonged liquidity trap, a steep Keynesian cross, or a large financial accelerator.

[*]An earlier version was entitled "Managing Expectations without Rational Expectations." Angeletos acknowledges the financial support of the National Science Foundation (Award #1757198). We also thank various seminar participants for comments.

[†]MIT and NBER; angelet@mit.edu.

[‡]MIT; ksastry@mit.edu.

1 Introduction

Forward guidance is rarely comprehensive. Even if a central bank can shape expectations about future interest rates, it remains up to the public to predict the consequences for aggregate demand and unemployment. Under what circumstances is it better to do the opposite, promising todo "whatever it takes" to achieve a specific outcome and leaving the public to ponder what policy will support this target?

We study how the answer to this question depends on bounded rationality, or the possibility that agents make mistakes in reasoning about the behavior of others and the equilibrium effects of policy. Our main lesson is that this distortion is minimized by the following strategy: offer guidance in terms of the outcome target ("bring unemployment down to Y%") rather than the instrument ("keep interest rates low for τ periods") whenever the GE feedback is sufficiently strong.

Context. The following example, nested in our framework, helps fix ideas. The economy is in a recession at the zero lower bound. Aggregate demand depends on expectations of future interest rates and aggregate income. Aggregate income in turn is demand-determined, forming an analogue of the Keynesian cross.[1] For simplicity, there is no uncertainty about the state of the economy and its future prospects, other than what the central bank plans to do in the future. The central bank can manage market expectations by offering a promise to keep interest rates either low for a certain amount of time, or for as long as it takes to reach a certain target for aggregate employment. In the first case, the central bank offers "clarity" about its policy instrument; in the latter, about its ultimate target.

Many central banks faced such a choice during the Great Recession. For instance, in December 2012 the US Federal Reserve transitioned from communicating the time at which it intended to start lifting interest rates to communicating a target for unemployment.[2] Our framework stylizes this situation as a switch from "instrument communication" to "target communication."[3]

The sharpest example of the kind of target communication we have in mind is ECB President Mario Draghi's famous proclamation in July 2012 to do "whatever it takes." This succeeded to restore market confidence in markets, despite an abundant lack of clarity about the policy mix to be used.

In this paper, we offer a new take on what "confidence" means and on when such communications may be most effective. We do not rely on the policymaker's words or actions selecting one of many equilibria. We also abstract from familiar considerations such as signaling effects (Morris and Shin, 2002; Campbell et al., 2012; Nakamura and Steinsson, 2018), commitment problems (Atkeson, Chari and Kehoe, 2007), or the best way of conditioning policy on the economy's fundamentals (Poole, 1970). We instead focus on the possibility that private agents are not fully rational or have difficulty reasoning about the economy—and on the resulting role for policy in minimizing this distortion.

[1] The Keynesian cross, or the feedback between income and spending, is an example of the kind of GE feedback we have in mind. Its strength is herein exogenously parameterized but in richer models is directly related to the severity of credit frictions and the expected length of the liquidity trap.

[2] Think of the former as communicating the *magnitude* of the policy intervention: in the textbook New Keynesian model, the Euler condition implies that reducing the interest rate by 1% for 2 periods is the same as reducing it by 2% for 1 period.

[3] Sections 2.2 and 2.3 contain a more detailed discussion of the relevant events, and of their mapping to our formal model.

A rational expectations benchmark and beyond. Consider the version of our framework that imposes a representative agent and rational expectations, as in textbook policy problems. In this benchmark, the combination of policies and outcomes that can be implemented are invariant to whether the policymaker communicates the value of the policy instrument or that of the targeted outcome. This invariance epitomizes a more general property of the Ramsey policy paradigm, the equivalence of primal and dual formulations (Lucas and Stokey, 1983; Chari and Kehoe, 1999).

This property depends critically not only on the elementary assumption that the representative agent is *herself* rational and aware of the policy communication, but also on the subtler assumption that such rationality and awareness is common knowledge ("I know that you know..."). Such common knowledge guarantees that every agent can make the same *correct* conjecture about the behavior of others and hence also about the equilibrium mapping between policy instruments and targeted outcome—which in turn guarantees that it makes no difference whether the policymaker picks the one or the other.

Our contribution starts by relaxing this assumption: we introduce a bound on the depth of the agents' knowledge and rationality, in the form of anchored higher-order beliefs and/or level-k thinking.[4] This operationalizes the idea that agents imperfectly reason about the behavior of others and, by extension, the GE effects of policy. The friction is taken for granted; our contribution is to study whether and how the policymaker can work around it.

Main lesson. Our main lesson is that, in the presence of this friction, offering guidance in terms of targets rather than instruments is preferable when and only when the GE feedback is sufficiently strong. In practical terms, policymakers should put more weight on communicating a commitment to an unemployment or output growth target when the economy enters a prolonged liquidity trap, when the Keynesian cross is steep, or when the financial accelerator is large. In a sense we make precise, such a policy "maximizes confidence," or minimizes the distortion caused by bounded rationality.

Logic. The above lesson combines two intermediate results, the first of which regards the policymaker's implementability constraint, or the equilibrium relation between the instrument and the outcome. As anticipated earlier, this relation is invariant to the form of forward guidance in our rational-expectations benchmark. Our first result explain *how* this invariance breaks once we add bounded rationality.

Instrument and target communication induce qualitatively different strategic interactions within the private sector. In the former case, agents play a game of strategic complements: when an agent expects the others to invest or spend less in response to the announcement, she responds less herself. In the latter case, everything flips: conditional on an announced employment or GDP target, a household that expects higher aggregate spending also expects a higher interest rate, which reduces the incentive to spend.

That these considerations balance out to the same equilibrium combination of policies and outcomes, or the same implementability constraint, is a knife-edge consequence of rational expectations.

[4] Our main specification borrows the relevant insight about higher-order beliefs from the literature on incomplete information (Angeletos and Lian, 2018; Morris and Shin, 2002, 2006; Woodford, 2003) but shuts down first-order uncertainty with the help of heterogeneous priors. Level-k Thinking (Farhi and Werning, 2019; Garcıa-Schmidt and Woodford, 2019; Nagel, 1995; Stahl, 1993) delivers essentially the same mechanism and the same results, modulo a minor "bug" that we discuss in due course.

Away from this benchmark, the aforementioned strategic considerations determine how a mis-specified belief about what others know or how others act translates into different equilibrium behaviors. And, because these strategic considerations depend on the communication strategy, so too does the set of implementable pairs for the instrument and the outcome.

Our second result relates to the interaction between the form of forward guidance and the underlying GE mechanism. On the one hand, the form of forward guidance regulates which object the agents have to forecast or reason about: fixing a value for the instrument burdens the agents with the task of predicting the outcome, setting a sharp target for the latter lets them ponder what the requisite policy will be. On the other hand, the GE feedback regulates which of these two objects is relatively more important in shaping actual behavior: when this feedback is weak, agents care relatively more about interest rates; and when it is strong, they care more about aggregate demand.

These observations, together, imply that target communication minimizes the bite of bounded rationality on implementability if and only if the GE feedback is sufficiently large. And since we focus on bounded rationality as the only source of a welfare loss relative to the first best, this yields our result that target communication is optimal if and only if the GE feedback is sufficiently large .

Monetary policy application. Let us return to our example of forward guidance about future monetary policy during a liquidity trap. A recent theoretical literature has studied the role of bounded rationality focusing exclusively on the case that we have called instrument communication; see Angeletos and Lian (2018), Farhi and Werning (2019) and Gabaix (2018). By contrast, the choice between instrument- and target-focused guidance was explicitly debated by policymakers during the Great Recession. Furthermore, concerns about “clarity” and “market interpretation” were loudly voiced, even if not fully articulated. These issues are covered extensively in retrospectives by Blinder (2018) and Feroli et al. (2017), and the discussions of the latter by Powell (2016) and Williams (2016).

In this context, we make three contributions. First, we qualify the lessons of the aforementioned theoretical literature by showing that the same mechanism that attenuates the power of forward guidance under instrument communication *amplifies* its power under target communication. Second, we offer a *formal* counterpart to the aforementioned policy concerns, relating “clarity” and “market interpretation” to bounds in people’s depth of knowledge and rationality. . And third, we tie the optimal form of forward guidance to the strength of the GE feedback loop between aggregate spending, employment, and income. More succinctly, we argue that the recent recession was the worst time to chatter about date-based interest-rate lift-offs and the best time to talk loud about stabilizing unemployment.

Robustness. Like the previously cited theoretical literature on forward guidance, our main analysis equates bounded rationality to a particular kind of bias: under-reaction of the beliefs of the behavior of others and, hence, under-estimation of policy’s GE effects. What if we consider the opposite bias, or perhaps random, non-systematic mistakes akin to “animal spirits”?

Although different assumptions about the distortion in beliefs can upset the logic of which form of forward guidance is attenuated or amplified, our result about their *relative* efficacy and the intuition that “optimal policy minimizes the role for distorted beliefs” remain intact. By the same token, a policymaker who suspects expectations are not "fully" rational but is not sure of the right model of mis-specification would still find concrete guidance from our analysis.

We further show that our main result is robust to introducing unobserved shocks to fundamentals, imperfect control of the policy or the outcome, or measurement error. These possibilities bring in new trade offs, of the kind first considered in Poole (1970), but do not disrupt our own logic.

Last but not least, our insights our robust to letting the policymaker communicate a sophisticated, state-contingent, policy plan as opposed to restricting him to the simpler, binary instruments-versus-targets choice allowed in our baseline analysis. This point deserves further discussion.

We favor the aforementioned restriction not only for expositional simplicity but also because we think that more sophisticated policy plans are hard to explain in practice and may even backfire. This consideration was present in actual FOMC's deliberations. Most characteristically, Minneapolis Fed President Narayana Korcherlakota argued in December 2012 that a communication of a fine-tuned, stage-contingent plan risked "letting the perfect be the enemy of the good."[5]

Absent this complication, the policymaker can of course do better by communicating a flexible relation between the policy instrument and the targeted outcome. Indeed, this option recovers the first best is a version of our model that has the policymaker be sufficiently knowledgable about the structure of the economy. But even when this the case, the optimal policy is driven by the same trade off as that emphasized in our baseline analysis. To minimize the bite of bounded rationality and attain either the first best (in special circumstances) or the applicable second best (more generally), the policymaker oughts to put more weight on target communication when the GE feedback is larger.

Related literature. Apart from the literature on forward guidance in monetary policy, which was discussed above, our paper's most direct contributions are to the literatures on policy regimes and policy communications that follow the leads of, respectively, Poole (1970) and Morris and Shin (2002).

Poole (1970) considers how the optimal choice among different policy regimes, such as fixing the interest rate or the growth rate of money, depends on the composition of shocks to fundamentals, such as preferences and technology (or "demand" and "supply"). The same logic underlies Weitzman (1974)'s classic on "prices vs quantities;" the literature on "tariffs vs quotas" that follows his lead; the modern literature on optimal Taylor rules; and a line of work that adds time-inconsistency considerations (Atkeson, Chari and Kehoe, 2007). Our paper highlights a novel issue: how different policy regimes can regulate the impact of any mistakes agents make in reasoning about equilibrium.

Consider next the literature spurred by Morris and Shin (2002), such as Amador and Weill (2010), Angeletos and Pavan (2007), Chahrour (2014), Cornand and Heinemann (2008), James and Lawler (2011), and Myatt and Wallace (2012). We share this literature's emphasis on higher-order beliefs but, as already alluded to, change the meaning of policy communication. In this literature, policy communication means revelation of information about an exogenous shock to the agents' payoffs, holding constant their strategic interaction. In our paper, it means regulation of that interaction and thereby of the bite

[5] Blinder (2018) offers a different but complementary argument: even when the Fed tried to spell out additional contingencies, markets largely ignored them and instead read a simpler from of target communication like that modeled in our paper.

of higher-order beliefs or bounded rationality. Furthermore, as explained later on, in our setting the revelation of the exogenous shock *per se* is both irrelevant and ineffective; what matters is only the communication of the policymaker's choice.

Angeletos and Pavan (2009) and Cornand and Heinemann (2015) lie in the middle ground between the above literature and our paper. Angeletos and Pavan (2009) allow a policymaker to regulate the agent's strategic interaction but maintain rational expectations and focus, instead, on how such regulation influences the use and the aggregation of information. Cornand and Heinemann (2015) allow bounded rationality but abstract from policy and focus, instead, on how bounded rationality influences the use and the social value of information. These paper therefore foreshadow two key ingredients of our analysis, policy and bounded rationality, but do not share our context or results.

Related are also Bergemann and Morris (2016), who study the robustness of a mechanism to the designer's uncertainty about the players' information, and Hansen and Sargent (2007), who study the robustness of policy to an adversarial Nature. Our exercise, instead, represents a form of robustness to bounded rationality in the private sector.

Finally, we connect not only to the US-centric retrospectives reviewed earlier but also to a more global discussion of new techniques for central banking. In a survey of 95 international central bank heads, Blinder et al. (2017) report that the instrument-vs-target question has "stuck" in the policy toolbox. We hope our paper formalizes an important trade-off that might prioritize one strategy over the other.

Layout. Section 2 introduces our framework and expands on the policy context. Section 3 studies our rational-expectations benchmark and lays down the foundations of the subsequent analysis. Section 4 contains our main specification, anchored higher-order beliefs. Section 5 translates our abstract results to the ZLB context. Section 6 explores the robustness of our results to alternative specifications of the belief friction and to a more complicated landscape of policy goals and policy options. Section 7 concludes.

2 Framework and Context

In this section we introduce the physical environment, the incentives of the private agents, the objective of the policymaker, and the timing of actions. This *excludes* belief formation, which we will return to later. We also discuss how our abstract model maps to policy settings.

2.1 General Structure

The economy is populated by a continuum of private agents, indexed by $i \in [0,1]$, and a policymaker. Each private agent chooses an action $k_i \in \mathbb{R}$. The policymaker controls a policy instrument $\tau \in \mathbb{R}$ and is interested in manipulating an aggregate outcome $Y \in \mathbb{R}$.

The aggregate outcome is related to the policy instrument and the behavior of the agents as follows:

$$Y = (1-\alpha)\tau + \alpha K \tag{1}$$

where $K \equiv \int k_i \, \mathrm{d}i$ is the average action of the private agents and $\alpha \in (0,1)$ is a fixed parameter. This parameter controls how much of the effect of the policy instrument τ on the outcome Y is direct, or mechanical, rather than channeled through the endogenous response of K.

The behavior of the private agents, in turn, is governed by the following best responses:

$$k_i = (1-\gamma)\mathbb{E}_i[\tau] + \gamma\mathbb{E}_i[Y] \tag{2}$$

where $\mathbb{E}_i$ denotes the subjective expectation of agent i and $\gamma \in (0,1)$ is a fixed parameter. Depending on assumptions made later on, the operator $\mathbb{E}_i$ may or may not be consistent with Rational Expectations Equilibrium (REE). The parameter γ controls how much private incentives depend on expectations of the aggregate outcome, which in turn depends on the behavior of others.

Key features and interpretation. Our framework stylizes three features likely shared by many applications. First, individual decisions depend on two kinds of expectations: the expectations of a policy instrument, such as a tax or the interest rate set by the central bank, and the expectations of an aggregate outcome, such as aggregate output. Second, the realized aggregate outcome depends on the realized aggregate behavior. And third, the policy instrument has a direct effect on the aggregate outcome even if we hold constant the decisions under consideration.

The first two assumptions capture the interdependence of economic decisions such as firm investment and consumer spending. In macroeconomics, this interdependence typically reflects general equilibrium (GE) interactions. Accordingly, the parameter γ, which plays a crucial role in the subsequent analysis, may be interpreted as a measure of the strength of the GE interaction. The third assumption and the parameter α, on the other hand, play a more mechanical function. Had $1-\alpha$ been zero, the policymaker could not possibly commit to a specific target for Y "no matter what" (i.e., regardless of K). Letting $\alpha < 1$ makes sure that such a commitment is viable.

Policy objective. The policymaker minimizes the expectation of the following loss function:

$$L(\tau, Y, \theta) \equiv (1-\chi)(\tau-\theta)^2 + \chi(Y-\theta)^2. \tag{3}$$

where $\chi \in (0,1)$ is a fixed scalar and θ is a zero-mean random variable that represents the policymaker's ideal or first-best combination of the instrument and the outcome.

The micro-foundations of this objective are left outside the analysis. The main insights regarding the bite of bounded rationality on implementability and the regulation of this bite by the form of forward guidance do not depend at all on the specification of the policymaker's objective. The adopted specification only sharpens the normative exercise by letting the policymaker attain her first best (zero loss) in the rational-expectations benchmark studied in the next section.

The realization of θ is observed by the policymaker but not by the private agents. Because we assume full commitment, this does not introduce incentive problems. And because θ does not enter conditions (1) and (2), the agents do not care to know θ *per se*; they only care to know what the policymaker plans to do and how this may affect the behavior of others. As anticipated in the Introduction, the sole purpose of letting θ be random and unobserved to the agents is therefore to motivate why the agents do not a priori know what the policymaker will do—they need "forward guidance."

Timing. There are three stages, or periods, which are described below:

0. The policymaker observes θ and, conditional on that, chooses whether to engage in "instrument communication," namely announce a value $\hat{\tau}$ for policy instrument, or "target communication," namely announce a target $\hat{Y}$ for the outcome.

1. Each agent i chooses k_i.

2. K is observed by the policymaker and (τ, Y) are determined as follows. In the case of instrument communication, $\tau = \hat{\tau}$ and Y is given by condition (1). In the case of target communication, $Y = \hat{Y}$ and τ is adjusted so that condition (1) holds with $Y = \hat{Y}$.

This structure embeds the assumption of that the policymaker always honors in stage 2 any promise made in stage 0. Different communications are therefore equated to different commitments: instrument communication means forward guidance in the form of a commitment to a value for τ and, similarly, target communication means forward guidance in the form of a commitment to a target for Y. However, the choice between these two strategies has nothing to do with time-inconsistency considerations, because commitment is full. As it will become clear in the sequel, this choice only has to do with the management of the expectations agents form in stage 2 about the behavior of others.

2.2 Micro-foundations and applications

Monetary policy and liquidity traps. Our primary application is forward guidance by a monetary authority during a liquidity trap. Several advanced-economy central banks attempted during and after the Great Recession (Table 1). The main distinction we draw is between forward guidance that deals *exclusively* with plans for policy instruments (e.g., dates at which interest rates will remain low) with forward guidance that ties future policy with clear objectives (e.g., an unemployment target). The US Fed and Bank of England notably experimented with the latter (rows 2 and 3).[6] In the targets category we also draw some inspiration from Mario Draghi's famous claim to do "whatever it takes" to preserve the Euro, non-quantitative guidance that perhaps *more* effectively demonstrates commitment ("And, believe me, it will be enough.").

In Appendix B.1, we show how to cast our distinction between instruments and targets within a micro-founded, albeit stylized, New Keynesian model. The instrument τ is the interest rate set after the economy exits a liquidity trap (i.e., after the zero lower bound has ceased to bind), the action K is aggregate spending in the middle of the trap, the outcome Y is the relevant income measure, the present discounted value of aggregate output both within and outside of the trap.

[6] Feroli et al. (2017), in their article reviewing (and critiquing) US monetary policy communication, provide a useful classification of the US Fed's communication policies both inside and outside the crisis. Section 2.3 of that paper, in particular, provides a qualitative classification of communication on its relative "data-intensity."

Central Bank	Date	Source	Type	Statement
US Federal Reserve	Aug 9, 2011	Policy statement by Committee	Instrument	[T]he Committee decided today to keep the target range for the federal funds rate at 0 to 1/4 percent. The Committee currently anticipates that economic conditions--including low rates of resource utilization and a subdued outlook for inflation over the medium run--are likely to warrant exceptionally low levels for the federal funds rate at least through mid-2013
US Federal Reserve	Dec 12, 2012	Policy statement by Committee	Target	[T]he Committee decided to keep the target range for the federal funds rate at 0 to 1/4 percent and currently anticipates that this exceptionally low range for the federal funds rate will be appropriate at least as long as the unemployment rate remains above 6-1/2 percent, inflation between one and two years ahead is projected to be no more than a half percentage point above the Committee's 2 percent longer-run goal, and longer-term inflation expectations continue to be well anchored.
Bank of England	Aug 7, 2013	Letter from Governor Mark Carney to the Chancellor of the Exchequer	Target	In practice, that means the [Monetary Policy Committee] intends not to raise Bank Rate above its current level of 0.5%, at least until the Labour Force Survey headline measure of unemployment has fallen to a threshold of 7%. While the unemployment rate remains above 7%, the MPC stands ready to undertake further asset purchases if additional stimulus is warranted.
Bank of Japan	Apr 4, 2013	Policy statement by Committee	Target	The Bank will achieve the price stability target of 2 percent in terms of the year-on-year rate of change in the consumer price index (CPI) at the earliest possible time, with a time horizon of about two years. In order to do so, it will enter a new phase of monetary easing both in terms of quantity and quality.
European Central Bank	Jul 4, 2013	Press conference by President Mario Draghi	Instrument	The Governing Council expects the key ECB interest rates to remain at present or lower levels for an extended period of time. This expectation is based on the overall subdued outlook for inflation extending into the medium term, given the broad-based weakness in the real economy and subdued monetary dynamics.
European Central Bank	Jul 26, 2012	Speech by President Mario Draghi at the Global Investment Conference	Target?	But there is another message I want to tell you. Within our mandate, the ECB is ready to do whatever it takes to preserve the euro. And believe me, it will be enough.

Table 1: Examples of forward guidance in recent monetary history.

The mechanics of instrument communication are familiar: lower interest rates tomorrow increase consumption today both through a PE effect (intertemporal substitution) and a GE effect (a feedback loop between aggregate spending and income)

Target communication, on the other hand, is like announcing a target for permanent income, or the path of employment, while leaving the interest rate path undetermined. The target is ultimately met by adjusting the "lift-off" interest rate, with more aggressive policy undertaken after the trap if output was low during it and less aggressive policy if output was high. It is simple to contemplate an extension with multiple periods after the liquidity trap, so a lower "lift-off" rate (or a higher τ) is like a longer time at zero nominal rates after economy has exited the liquidity trap.

Fiscal policy and anticipation effects. In a second application, we consider forward guidance about future taxationin a purely Neoclassical environment with aggregate demand externalities (Appendix B.2). This serves three goals. First, it illustrates how a substantially different, flexible-price mechanism could generate the basic structure of equations (1) and (2). Second, it captures within the same context both the case of *strategic complements* ($\gamma > 0$) and the case of *strategic substitutes* ($\gamma < 0$), the former arising from aggregate demand externalities and the latter from competition over a scarce resource (labor). And third, it builds a bridge to the empirical literature assessing the anticipatory responses to future fiscal policy

(Mertens and Ravn, 2010, 2012; Leeper, Walker and Yang, 2013).

Communicating with financial markets. Both of our examples emphasize the "real side" of the economy instead of financial markets, which are certainly more attentive to the fine details of policy communications (but could still be subject to bounded rationality). A quick fix is to relabel terms in each of the previous two models, treating forward-looking decisions (consumption or investment) as financial trades. Each force could then also translate into asset price movements, so the abstract K could measure either quantity or price. The GE feedback can still be a Keynesian cross in the real economy; a feedback loop between household wealth and aggregate demand as in Caballero and Simsek (2017); or perhaps a financial accelerator as in Bernanke, Gertler and Gilchrist (1999) and Kiyotaki and Moore (1997).

2.3 Why instruments vs. targets

In reality, it is natural to think about more complex communication than speaking about a single instrument or about a single target. For instance, in a survey of 95 international central bank heads, Blinder et al. (2017) report that 26.9% of respondents would favor conducting future forward guidance via a "data-based" approach with explicit contingency on targets (which is herein proxied by target communication), versus 13.5% for a "calendar-based" approach that focuses on the future path of interest rates (instrument communication). But 38.5% of the respondents favored other qualitative approaches that may mix the two.[7]

Even the US Fed's famous announcement of an "unemployment target" in December 2012 (row 2 of Table 1) was, as originally stated, a more sophisticated contingency on unemployment and inflation among other indicators. Our focus on a starker instument-vs-target choice is nevertheless consistent with a plausible real-word *interpretation* of that December 2012 policy shift. As Alan Blinder (2018) wrote in a retrospective analysis, much of the public seemed to hear the following much simpler edict:

> "The Fed would begin to raise rates as soon as the unemployment rate dipped below 6.5 percent. Period."

That is, even though this particular communication was not as sharp or resolute as Mario Draghi's "whatever it takes" speech, it may well be approximated for our purposes by what we call "target communication."

During the December 2012 FOMC meeting deliberations, San Francisco Fed President John C. Williams expressed a similar idea based on limited attention:

> We should recognize we are shining a very bright spotlight on the unemployment rate. ... [P]eople have limited capacity to absorb information, and are, therefore, selective in what information they pay attention to. When we stated a specific date for lift-off, the spotlight was cast on the calendar, and that's what everyone focused on, for better or for worse. Once we start talking in terms of an unemployment threshold, it will be the unemployment rate that takes center stage, commanding all of the attention of our audience.

[7] The remaining shares were for "None" (11.5%), "Other" (15.4%), and "Too Early to Judge" (21.2%).

Our coming model for belief formation will formalize this idea by letting the policymaker be effective at anchoring a *single* salient expectation, either that of the policy instrument or that of the target outcome. For simplicity, we take this premise as given instead of explicitly micro-founding it.

What about communicating more complex objects, like a quantitative monetary rule? In the running example of the December 2012 communication switch, several Fed members saw simpler forms of target communication as the only feasible approximation of such an "ideal" rule in an uncertain policy environment. Minneapolis Fed President Narayana Kocherlakota commented that, absent "the perfect description of a reaction function," attempting a more complex communication would be "letting the perfect be the enemy of the good."

The real-world complications that make sophisticated policy rules hard to communicate or counterproductive are outside the scope of our analysis. Nonetheless, in Section 6.3 we will return to a relaxed policy problem that explores "the perfect": clear and credible communication of a linear rule linking τ and Y. Absent the aforementioned complications, this additional option will, of course, allow the policymaker to achieve better outcomes. But it will do so only in a way that fully preserves the main intuitions.[8]

3 Rational Expectations and Beyond

We now return to the abstract setting. We first explain why the form of forward guidance is irrelevant in the representative-agent, rational-expectations benchmark. This sets the stage for our subsequent, structured departures from it. We also lay out the foundations of the subsequent analysis by showing how the form of forward guidance influences the nature of agents' strategic interaction.

3.1 The REE benchmark

Consider first a "textbook" policy paradigm. There is a representative agent, who knows the structure of the economy, observes the policy announcement, and forms rational expectations.[9] In this benchmark, $\mathbb{E}_i[\cdot] = \mathbb{E}[\cdot|\hat{X}]$ for all i, where $\mathbb{E}[\cdot|\hat{X}]$ is the rational expectation conditional on announcement $\hat{X}$, with $X \in \{\tau, Y\}$ depending on the mode of communication. As a result, $k_i = K$ for all i and condition (2) reduces to the following condition for optimal behavior:

$$K = (1-\gamma)\mathbb{E}[\tau|\hat{X}] + \alpha\mathbb{E}[Y|\hat{X}]. \tag{4}$$

We can thus define the sets of the combinations of the policy instrument, τ, and the outcome, Y, that can be implemented under each form of forward guidance as follows:

[8] Our analysis also rules out the possibility that the policymaker communicates a different single number instead of the intended value for τ or Y: the policymaker's ideal point, θ, or a target for the average action, K. In Appendix F we show that neither option is well-posed once we allow for bounded rationality. The former is not particularly "informative," given that agents' payoffs do not depend on θ, and thus on a purely technical level does not select a unique equilibrium. The second is essentially impossible, because the policymaker has no tools to honor its commitment.

[9] This is effectively the same as imposing, in a game, complete information and Nash equilibrium.

Definition 1. *A pair* (τ, Y) *is implementable under instrument [respectively, target] communication if there is an announcement* $\hat{\tau}$ *[respectively,* $\hat{Y}$*] and an action* K *for the representative agent such that conditions (1) and (4) are satisfied, expectations are rational, and* $\tau = \hat{\tau}$ *[respectively,* $Y = \hat{Y}$*].*

This definition embeds Rational Expectations Equilibrium (REE). In the subsequent sections, we will revisit implementability under different solution concepts. In the rest of this section, we formulate and solve the policymaker's problem in a manner that parallels the analysis in the subsequent sections.

Denote with $\mathscr{A}^*_\tau$ and $\mathscr{A}^*_Y$ the sets of (τ, Y) that are implementable under, respectively, instrument and target communication. The policymaker's problem uis:

$$\min_{\mathscr{A} \in \{\mathscr{A}^*_\tau, \mathscr{A}^*_Y\}, (\tau, Y) \in \mathscr{A}} \mathbb{E}[L(\tau, Y, \theta)] \tag{5}$$

The choice $\mathscr{A} \in \{\mathscr{A}^*_\tau, \mathscr{A}^*_Y\}$ captures the choice of the optimal mode of communication (instrument vs target). The choice $(\tau, Y) \in \mathscr{A}$ captures the optimal choice of the pair (τ, Y) taking as given the mode of communication. Both of these choices are conditional on θ.

We now proceed to show that $\mathscr{A}^*_\tau = \mathscr{A}^*_Y$. Using condition (1) to compute $\mathbb{E}[Y]$ and noting that $\mathbb{E}[K] = K$ (the representative agent knows his own action), we can restate condition (4) as

$$K = (1 - \alpha\gamma)\mathbb{E}[\tau|\hat{X}] + \alpha\gamma K$$

Since $\alpha\gamma \neq 1$, this implies that, in any REE,

$$K = \mathbb{E}[\tau|\hat{X}], \qquad Y = (1-\alpha)\tau + \alpha\mathbb{E}[\tau|\hat{X}] \qquad \text{and} \qquad \mathbb{E}[Y|\hat{X}] = \mathbb{E}[\tau|\hat{X}] = K$$

These properties hold regardless of the mode of communication. With instrument communication, we also have $\tau = \hat{\tau} = \mathbb{E}[\tau|\hat{X}]$. It follows that, for any $\hat{\tau}$, the REE is unique and satisfies $K = Y = \tau = \hat{\tau}$. With target communication, on the other hand, we have $Y = \hat{Y} = \mathbb{E}[Y|\hat{X}]$. It follows that, for any $\hat{Y}$, the REE is unique and satisfies $K = Y = \tau = \hat{Y}$. Combining these facts, we infer that, regardless of the mode of communication, a pair (τ, Y) is implementable if and only if $\tau = Y$.

Proposition 1. $\mathscr{A}^*_\tau = \mathscr{A}^*_Y = \mathscr{A}^* \equiv \{(\tau, Y) : \tau = Y\}$.

That $\mathscr{A}^*$ is a linear locus with slope 1 is a simplifying feature of our environment. The relevant point here is that the implementability constraint faced by the planner is invariant to the form of forward guidance,[10] which in turn implies the following.

Proposition 2. *The policymaker attains her first best* $(L = 0)$ *by announcing* $\hat{\tau} = \theta$*, as well as by announcing* $\hat{Y} = \theta$*. The optimal form of forward guidance is therefore indeterminate.*

In fact, the first best is attained even if the policymaker only announces the shock θ itself, as opposed to announcing a policy plan. For, once θ is known, every agent can reason, without the slightest grain of doubt and without any chance of error, that all other agents will play $K = \theta$ and that the policymaker will set $\tau = \theta$, in which case it is optimal for him to play $k_i = \theta$ as well.

[10]This invariance mirrors the equivalence of the "dual" and "primal" approaches in the Ramsey literature (Chari and Kehoe, 1999): in our setting, $\mathscr{A}^*_\tau$ corresponds to the primal problem, where the planner chooses instruments, and $\mathscr{A}^*_Y$ corresponds to the dual, where she chooses allocations.

3.2 Unpacking the assumptions

Any departure from rational expectations has to be done in a structured way, or else "anything goes." To be more clear about where we are heading, we first recast the rational-expectations benchmark as the combination of two assumptions: one regarding the agents' *own* rationality and awareness; and another regarding the beliefs about *others*.

Assumption 1. *Every agent is rational and attentive in the following sense: he is Bayesian (although possibly with a mis-specified prior), acts according to condition (2), understands that the outcome is determined by condition (2) and that the policymaker has full commitment and acts so as to minimize (3), and receives any message sent by the policymaker.*

Assumption 2. *The aforementioned facts are common knowledge.*

Proposition 3. *Provided that $\alpha < \frac{1}{2-\gamma}$, the REE benchmark studied in the previous section is equivalent to the joint of Assumptions 1 and 2.*

The basic idea is that, for any policy announcement made at stage 0, the joint of Assumptions 1 and 2 yield a unique rationalizable outcome in stages 1 and 2, which coincides with the REE outcome obtained in the previous section. The restriction $\alpha < \frac{1}{2-\gamma}$ is needed for the uniqueness of the rationalizable outcome, but not for the uniqueness of the REE and can be dispensed with for most of the applied lessons. We next discuss what Assumptions 1 and 2 mean and how they help structure the forms of "bounded rationality" considered in the rest of the paper.

Assumption 1 imposes that, for any i, agent i's subjective beliefs and behavior satisfy the following three restrictions:

$$\mathbb{E}_i[X] = \hat{X}, \qquad \mathbb{E}_i[Y] = (1-\alpha)\mathbb{E}_i[\tau] + \alpha\mathbb{E}_i[K], \qquad \text{and} \qquad k_i = \left(1-\gamma\right)\mathbb{E}_i[\tau] + \gamma\mathbb{E}_i[Y], \tag{6}$$

where $X \in \{\tau, Y\}$ depending on the mode of communication. The first restriction follows from the agent's attentiveness to policy communications and his knowledge of the policymaker's commitment; the second follows from his knowledge of condition (1); the third repeats condition (2).

Assumption 2, in turn, imposes that agents can reason, with full confidence and no mistake, that the above restrictions extend from their own behavior and beliefs to the behavior and the beliefs of others, to the beliefs of others about the behavior and the beliefs of others, and so on, ad infinitum. It is such *boundless* knowledge and rationality that our frictionless benchmark and the textbook policy paradigm alike impose—and that we instead seek to relax.

This explains the approach taken in the rest of the paper: we modify Assumption 2 while maintaining Assumption 1. This aims at isolating the role of any mistakes agents make when trying to predict or reason about the behavior of others and the GE consequences of any policy plan.

3.3 Forward guidance and strategic interaction

We close this section with an important observation that is hidden by the simplicity of the REE calculation. The communication choice determines which variable the agents are told directly and which they have to "reason about" or forecast. We recast this reasoning in a reduced-form game between agents conditional on each communication type. In the process we show how *deviations* in expectations could have opposite effects depending on what agents have to think about.

Consider first the case in which the policymaker announces, and commits on, a value $\hat{\tau}$ for the instrument. Recall that Assumption 1 yields the three restrictions given in condition (6). Under instrument communication, the first restriction becomes $\mathbb{E}_i[\tau] = \hat{\tau}$ and the remaining two restrictions reduce to

$$k_i = (1-\gamma)\hat{\tau} + \gamma\mathbb{E}_i[Y] \qquad \text{and} \qquad \mathbb{E}_i[Y] = (1-\alpha)\hat{\tau} + \alpha\mathbb{E}_i[K].$$

The first equation highlights that, under instrument communication, agents only need to predict Y. The second highlights that predicting Y is the same as predicting the behavior of others, or K. Combining them gives the following result.

Lemma 1. *Let $\delta_\tau \equiv \alpha\gamma$. When the policymaker announces and commits to a value $\hat{\tau}$ for the instrument, agents play a game of strategic complementarity in which best responses are given by*

$$k_i = (1-\delta_\tau)\hat{\tau} + \delta_\tau\mathbb{E}_i[K]. \tag{7}$$

The level of the best responses in this game is controlled by $\hat{\tau}$, the announced value of the policy instrument, while their slope is given by δ_τ. The latter encapsulates how much aggregate behavior depends on the forecasts agents form about one another's behavior relative to the policy instrument—or, equivalently, how much aggregate investment depends on the perceived GE effect of the policy relative to its PE effect.[11]

Consider now the case in which the policymaker announces a target $\hat{Y}$ for the outcome. In this case, $\mathbb{E}_i[Y] = \hat{Y}$ and the remaining two restrictions from condition (6) can be rewritten as

$$k_i = (1-\gamma)\mathbb{E}_i[\tau] + \gamma\hat{Y} \qquad \text{and} \qquad \mathbb{E}_i[\tau] = \tfrac{1}{1-\alpha}\hat{Y} - \tfrac{\alpha}{1-\alpha}\mathbb{E}_i[K].$$

The first equation highlights that, under target communication, agents need to predict the subsidy that will support the announced target. The second shows that, for given an announced target $\hat{Y}$, the expected subsidy is a *decreasing* function of the expected K : an agent who is pessimistic about aggregate investment expects the policymaker to use a higher subsidy in order to meet the given output target. Combining these two equations, we reach the following counterpart to Lemma 1.

Lemma 2. *Let $\delta_Y \equiv -\frac{\alpha}{1-\alpha}(1-\gamma)$. When the policymaker announces and commits to a target $\hat{Y}$ for the outcome, agents play a game of strategic substitutability in which best responses are given by*

$$k_i = (1-\delta_Y)\hat{Y} + \delta_Y\mathbb{E}_i[K]. \tag{8}$$

[11] The game obtained above is similar to the static beauty-contest games studied in, *inter alia*, Morris and Shin (2002), Woodford (2003), Angeletos and Pavan (2007, 2009), and Bergemann and Morris (2013), with $\hat{\tau}$ corresponding to the "fundamental," or the shifter of best responses, in these papers. There are, however, two subtle differences. First, whereas the fundamental in those papers is exogenous, here $\hat{\tau}$ is controlled by the policymaker. Second, whereas these papers let the fundamental be observed with noise, here $\hat{\tau}$ is perfectly observed.

This game is similar to that obtained in Lemma 1 in the following respect: in both cases, the policymaker's announcement controls the intercept of the best responses. The two games are nevertheless different in the following key respect: whereas the game obtained in Lemma 1 displayed strategic complementarity ($\delta_\tau > 0$), the one obtained here displays strategic substitutability ($\delta_Y < 0$). In the first scenario, an agent who expects the others to invest more has a higher incentive to invest, because higher K maps to higher Y and hence to higher returns for fixed τ. In the second scenario, the same agent has a *lower* incentive to invest, because a higher K means that a lower subsidy will be required in order to meet the announced target for Y.

We summarize this elementary, but important, point in the following corollary.

Corollary 1. *Switching from instrument communication to target communication changes the game played by the agents from one of strategic complementarity to one of strategic substitutability.*

In math, with $X \in \{\tau, Y\}$ indexing the mode of communication, the best responses obtained in Lemmas 1 and 2 are nested in the following form:

$$k_i = (1-\delta_X)\mathbb{E}_i[X] + \delta_X \mathbb{E}_i[K]. \tag{9}$$

for $\delta_\tau \in (0,1)$ and $\delta_Y < 0$. Given the restriction $\alpha < \frac{1}{2-\gamma}$, assumed from here on out, we have further that $\delta_X \in (-1,1)$ for both $X \in \{\tau, Y\}$.[12]

4 Optimal Forward Guidance with Anchored Beliefs

We now turn to the core of our contribution, which is to characterize the optimal strategy for managing expectations when the friction is anchored beliefs about others' responses to the announcement. This friction is introduced by replacing Assumption 2 with the following.

Assumption 3 (Lack of Common Knowledge of the Policy Message)**.** *Every agent believes that all other agents are rational but only a fraction $\lambda \in [0,1]$ of them is attentive to or aware of the policy message: every i believes that, for every $j \neq i$, $\mathbb{E}_j[X] = \mathbb{E}_i[X] = \hat{X}$ with probability λ and $\mathbb{E}_j[X] = 0$ with probability $1-\lambda$, where $X \in \{\tau, Y\}$ depending on the mode of communication. This fact and the value of λ are common knowledge.*

Relative to Assumption 2, Assumption 3 maintains common knowledge of rationality but drops common knowledge of the policy message. The former allows us to characterize behavior by iterating on best responses; the latter introduces the friction of interest.[13]

[12] This sharpens the analysis, but is not strictly need for the applied lessons. See the discussion in Appendix D.

[13] Under the restriction $\alpha < \frac{1}{2-\gamma}$, this is equivalent to changing the solution concept from REE to Perfect Bayesian Equilibrium with the following heterogeneous priors: each agent i receives a private signal s_i of the announcement; believes correctly that his signal is a drawn from a Dirac measure at $\hat{X}$; and believes incorrectly that, for any $j \neq i$, s_j is drawn from a Dirac measure at $\hat{X}$ with probability λ and from a Dirac measure at 0 with probability $1-\lambda$. A similar specification was used in Angeletos and La'O (2009) to add belief inertia in the New Keynesian model.

As noted in the Introduction, Assumption 3 is grounded on a literature that studies the role of higher-order uncertainty in common-prior, rational-expectations settings: in such settings, the inertia of higher-order beliefs to news is rationalized by noisy and heterogenous information, which itself could be the product of rational inattention. See, for example, Angeletos and Lian (2018) for an application to the ZLB context. But whereas that literature typically ties the friction in higher-order beliefs to a friction in first-order beliefs (noise or inattention), Assumption 3 idecouples the two frictions and maintains only the former.[14]

Assumption 3 also amounts to a "smooth" version of Level-k Thinking (Nagel, 1995; Garcıa-Schmidt and Woodford, 2019; Farhi and Werning, 2019). The exact mapping is spelled out in Appendix C. The punchline is that one can think of λ as a continuous index of the depth of reasoning. At the one extreme, $\lambda = 0$, agents are effectively level-0 thinkers, meaning they expect the others not to respond at all. At the other extreme, $\lambda = 1$, they are level-∞ thinkers, or "infinitely" rational. And for $\lambda \in (0,1)$, they are "boundedly rational".

Another possible re-interpretation of λ is in terms of the form of "cognitive discounting" introduced in Gabaix (2018): this formalization, too, amounts to anchoring higher-order beliefs, or the beliefs of the behavior of others. The available empirical evidence generally supports the existence of this kind of anchored beliefs, even if it is more agnostic on the underlying mechanism or micro-foundation.[15] This fact, and the precedent in the literature, explains why our baseline analysis focuses on this particular type of belief distortion. That said, in Section 6.1 we will show that our main result is robust to other belief distortions, including over-reactive or erratic higher-order beliefs.

4.1 Beliefs or reasoning

Assumption 3 is sufficient to prove that beliefs are more inertial than realized actions:

Lemma 3 (Anchored beliefs)**.** *For both modes of communication and for any value $\hat{X}$ of the policy message, $\bar{\mathbb{E}}[K] = \lambda K$.*

If the typical agent believes that only a fraction λ of the population is aware of the policy message like herself, she also expects the same fraction to respond like herself, and the remaining fraction to stay put. That is, $\mathbb{E}_i[K] = \lambda k_i$ for the typical agent and therefore also $\bar{\mathbb{E}}[K] = \lambda K$ on the aggregate.

A more detailed derivation helps reveal the underlying reasoning. Because we have maintained common knowledge of rationality, we can express an agent's reasoning about K by iterating on the best responses. This gives the expectations of K as a weighted average of her higher-order beliefs about X:

$$\mathbb{E}_i[K] = \mathbb{E}_i\left[(1-\delta_X)\sum_{h=1}^{\infty}(\delta_X)^{h-1}\bar{\mathbb{E}}^h[X]\right]. \tag{10}$$

[14] The importance of this decoupling is discussed in detail in Subsection D.3.

[15] For example, see Coibion and Gorodnichenko (2012) and Coibion et al. (2018) for evidence based on surveys of expectations, and Crawford, Costa-Gomes and Iriberri (2013), Nagel (1995) and Heinemann, Nagel and Ockenfels (2009) for experiments.

Because we have dropped common knowledge of the policy message, and in particular we have let the typical agent believe that only a fraction λ of the other agents is aware of the policy message, second-order beliefs satisfy

$$\mathbb{E}_i\left[\bar{\mathbb{E}}^1[X]\right] = \mathbb{E}_i\left[\mathbb{E}_j[X]\right] = \lambda\hat{X} + (1-\lambda)0 = \lambda\hat{X}.$$

By induction, for any $h \geq 1$,

$$\mathbb{E}_i\left[\bar{\mathbb{E}}^h[X]\right] = \lambda^h\hat{X}. \tag{11}$$

Relative to the frictionless benchmark, higher-order beliefs are therefore anchored to zero, and the more so the higher their order. It follows that $\mathbb{E}_i[K]$, which is a weighted average of the beliefs of order $h = 2$ and above, is also anchored to zero. And because k_i is itself a weighted average of $\hat{X}$ and $\mathbb{E}_i[K]$, k_i responds more strongly than $\mathbb{E}_i[K]$, or beliefs are more anchored than actions.[16]

4.2 Attenuation vs. amplification

Although the *nature* of the belief friction is qualitatively the same between the two forms of forward guidance, its *impact* on actual behavior is qualitatively different. Indeed, replacing $\mathbb{E}_i[K] = \lambda K$ in the best-response condition (9) and aggregating across agents, we reach the following result.

Lemma 4. *The realized aggregate investment following announcement $\hat{X}$ is given by*

$$K = \kappa_X\hat{X} \qquad \text{with} \qquad \kappa_X \equiv \frac{1-\delta_X}{1-\lambda\delta_X}, \tag{12}$$

where $X \in \{\tau, Y\}$ depending on the mode of communication. Furthermore, $\kappa_\tau = 1 = \kappa_Y$ for $\lambda = 1$; $\kappa_\tau < 1 < \kappa_Y$ for every $\lambda < 1$; and the distance of either κ_τ or κ_Y from 1 increases as λ falls.

Recall that the frictionless benchmark had $K = \hat{X}$, which corresponds to $\kappa_X = 1$. When $\delta_X > 0$, the ratio $\frac{1-\delta_X}{1-\lambda\delta_X}$ is strictly lower than 1 for every $\lambda < 1$ and is increasing in λ. When instead $\delta_X < 0$, this ratio is strictly higher than 1 for every $\lambda < 1$ and is decreasing in λ. Along with the fact that $\delta_\tau > 0 > \delta_Y$, this verifies the properties of κ_τ and κ_Y mentioned above. In simpler words:

Corollary 2. *Anchored beliefs attenuate the actual response of K under instrument communication, and amplify it under target communication. Furthermore, a larger friction (lower λ) translates to larger attenuation in the first case and to larger amplification in the second case.*

This result explains how the mode of communication regulates the impact of the introduced friction on actual outcomes. When agents play a game of strategic complementarity, anchoring the beliefs of the behavior of others causes each agent to respond less than in the frictionless benchmark. When instead agents play a game of strategic substitutability, the same friction causes each agent to respond more than in the frictionless benchmark. The result then follows directly from our earlier observation that the mode of communication changes the nature of the strategic interaction.

[16] In particular, using (11) into (10) yields $\mathbb{E}_i[K] = \lambda\frac{1-\delta_X}{1-\lambda\delta_X}\hat{X}$; using this to substitute $\hat{X}$ in best-response condition (9) and solving for $\mathbb{E}_i[K]$ gives $\mathbb{E}_i[K] = \lambda k_i$, as claimed.

4.3 Implementability

We now spell out the implications of the preceding observations for the combinations of τ and Y that are implementable under each mode of communication. This combines the previous Lemma about the implementable response of K to the announcement with the policymaker's freedom to announce $\hat{\tau}$ or $\hat{Y}$ as any number:

Proposition 4 (Implementation with anchored beliefs)**.** *Let $\mathscr{A}_\tau$ and $\mathscr{A}_Y$ denote the sets of the pairs (τ, Y) that are implementable under, respectively, instrument and target communication. Then,*

$$\mathscr{A}_\tau = \left\{(\tau, Y) : \tau = \mu_\tau(\lambda,\gamma) Y\right\} \qquad \textit{and} \qquad \mathscr{A}_Y = \left\{(\tau, Y) : \tau = \mu_Y(\lambda,\gamma) Y\right\},$$

where[17]

$$\mu_\tau(\lambda,\gamma) \equiv \left((1-\alpha) + \alpha \frac{1-\alpha\gamma}{1-\lambda\alpha\gamma}\right)^{-1} \quad \textit{and} \quad \mu_Y(\lambda,\gamma) \equiv \left(1 + \frac{\alpha^2(1-\lambda)(1-\gamma)}{1+\alpha(\lambda(1-\gamma)+\alpha\gamma-2)}\right)^{-1}.$$

The frictionless benchmark is nested by $\lambda = 1$ and results in $\mu_\tau = 1 = \mu_Y$. By contrast, for any $\lambda < 1$, we have $\mu_\tau \neq \mu_Y$. That is, the two implementable sets cease to coincide as soon as we move away from the frictionless benchmark.

The next proposition, which is proved in Appendix A, offers a sharper characterization of how μ_τ and μ_Y, the slopes of the two implementability constraints, compare to one another, as well as to the frictionless counterpart.

Proposition 5. *(i) $\mu_\tau(\lambda,\gamma) \geq 1$ with equality only when $\lambda = 1$ or $\gamma = 0$.*

(ii) $\mu_Y(\lambda,\gamma) \leq 1$ with equality only when $\lambda = 1$ or $\gamma = 1$

(iii) $\mu_\tau(\lambda,\gamma)$ increases in λ and $\mu_Y(\lambda,\gamma)$ decreases in λ.

The belief friction under consideration has opposite effects on the slope of the "budget lines" faced by the policymaker. With instrument communication, a higher friction (smaller λ) increases the slope, meaning that a higher variation in τ is needed to attain any given variation in Y. With target communication, the opposite is true. The distortion of the implementability constraint, as measured by the absolute value of $\mu_X(\lambda) - 1$, therefore increases in both cases, but the sign is different.

4.4 The role of GE feedback

Let us now turn attention to the role played by γ. Recall that γ proxies for the strength of the underlying GE feedback—the Keynesian income-spending multiplier in the application to monetary policy and the aggregate demand externality in the investment example. The next proposition, whose proof can be found in Appendix A, studies how this interacts with the belief friction in shaping the distortion of the implementability constraints.

[17] Throughout, we omit the dependence of μ_τ and μ_Y on α because we focus on the comparative statics in λ and γ. And in the main text, we often write μ_τ and μ_Y without their arguments in order to ease notation.

Proposition 6. *Fix any $\lambda \in (0,1)$ and $\alpha \in (0,1)$. As γ increases, both $\mu_\tau(\lambda,\gamma)$ and $\mu_Y(\lambda,\gamma)$ increase. Furthermore, $\mu_\tau(\lambda,1) > 1$ and $\mu_Y(\lambda,1) = \mu_Y^* = 1$, whereas $\mu_\tau(\lambda,0) = \mu_\tau^* = 1$ and $\mu_Y(\lambda,0) < 1$.*

As the GE effects gets stronger (γ increases), the distortion is *exacerbated* under instrument communication, in the sense that μ_τ gets further away from μ_τ^*, whereas it is *alleviated* under target communication, in the sense that μ_Y gets closer to μ_Y^*. The logic is best illustrated by considering the extremes in which $\gamma = 0$ and $\gamma = 1$.

Consider first the case in which the GE effect is absent, or $\gamma = 0$. Behavior is pinned down purely by the direct or PE effect of the policy: $k_i = \mathbb{E}_i\tau$ for all i. As a result, announcing and committing on a value $\hat{\tau}$ for the instrument guarantees that that $K = \hat{\tau}$, regardless of λ. Condition (1) then gives $Y = \hat{\tau}$, which means that $\mathscr{A}_\tau = \mathscr{A}_\tau^*$, for all $\lambda < 1$. That is, there is no distortion with instrument communication—but there is one with target communication. For when $\gamma = 0$, target communication transforms the game played among the agents from one with a null strategic interaction to one with a non-zero strategic substitutability (indeed, $\delta_\tau = 0$ but $\delta_Y < 0$ when $\gamma = 0$), thus also allowing the belief friction to influence the implementability constraint.

The converse is true when the GE effect is maximal, or $\gamma = 1$. Behavior is then pinned down exclusively by expectations of the outcome: $k_i = \mathbb{E}_i Y$ for all i. The distortion is then eliminated by, and only by, announcing and committing to a target for Y.

4.5 Optimal strategy

The previous discussion implies that, in the extreme cases of $\gamma \in \{0,1\}$, the first-best outcome remains implementable under one and only one form of forward guidance: instrument communication when $\gamma = 0$, target communication when $\gamma = 1$. Each strategy, in its most favorable case, sidesteps the friction entirely by eliminating agents' need to forecast, or reason about, others' actions.

What about the intermediate cases $\gamma \in (0,1)$? Neither strategy completely eliminates the need to reason about others' behavior. Still, the continuity and monotonicity properties of the implementable sets with respect to γ suggest that target communication is strictly preferred to instrument communication if and only if the GE effect is strong enough. The next theorem verifies this intuition.

Theorem 1 (Optimal Forward Guidance)**.** *For any $\lambda < 1$, there exists a threshold $\hat{\gamma} \in (0,1)$ such that: when $\gamma \in (0,\hat{\gamma})$, instrument communication is strictly optimal for all θ; and when $\gamma \in (\hat{\gamma},1)$, target communication is strictly optimal for all θ.*

A detailed proof is provided in Appendix A. Below we sketch the main argument. We also characterize the pairs (τ, Y) that get implemented by the optimal strategy for all θ.

Given any θ, the policymaker chooses a set $\mathscr{A} \in \{\mathscr{A}_\tau(\lambda), \mathscr{A}_Y(\lambda)\}$ and a pair $(\tau, Y) \in \mathscr{A}$ to minimize her loss:

$$\min_{\mathscr{A} \in \{\mathscr{A}_\tau(\lambda),\mathscr{A}_Y(\lambda)\},(\tau,Y)\in\mathscr{A}} L(\tau, Y, \theta)$$

Let $(\mathscr{A}^{\text{sb}}, \tau^{\text{sb}}, Y^{\text{sb}})$ be the (unique) triplet that attains the minimum.

Given the assumed specification of L and the characterization of the implementability sets in Proposition 4, we can restate the problem as the following choice of a *slope* between τ and Y. Letting $r \equiv \tau/\theta$, we reach the following even simpler representation after substituting in the implementability constraint:

$$\min_{\mu \in \{\mu_\tau(\lambda), \mu_Y(\lambda)\}, r \in \mathbb{R}} \left[(1-\chi)(r-1)^2 + \chi(r/\mu - 1)^2\right]$$

This makes clear that the optimal form of forward guidance is the same for all realizations of θ and lets r identify the optimal covariation of τ with θ. The policy problem reduces to choosing a value for $r \in \mathbb{R}$ and a value for $\mu \in \{\mu_\tau(\lambda), \mu_Y(\lambda)\}$. That is, if we let $(r^{\text{sb}}, \mu^{\text{sb}})$ be the solution to the above problem, the second-best values of the instrument and the outcome are given by, respectively, $\tau^{\text{sb}} = r^{\text{sb}}\theta$ and $Y^{\text{sb}} = (r^{\text{sb}}/\mu^{\text{sb}})\theta$.

Consider the "inner" problem of choosing r for given μ. It is simple to solve for r in closed form and arrive at the one-dimensional objective over the slope μ:[18]

$$\mathscr{L}(\mu) \equiv \min_r \left[(1-\chi)(r-1)^2 + \chi(r/\mu - 1)^2\right] = \frac{\chi(1-\chi)(1-\mu)^2}{\mu^2(1-\chi)+\chi}$$

We thus have that the optimal r satisfies $r(\mu) < 1$ for $\mu < 1$, $r(\mu) = 1$ for $\mu = 1$, and $r(\mu) > 1$ for $\mu > 1$; and that the resulting payoff is a U-shaped function of $\mu \in (0, \infty)$, with a minimum equal to 0 and attained at $\mu = 1$ (the frictionless case).

Let us now turn to the optimal choice of μ, which encodes the choice of the form of forward guidance. The magnitude of the policymaker's loss increases in the distance between μ and 1. The closer μ is to 1, the smaller would be the distortion from the frictionless benchmark even if we were to hold r fixed at 1. The fact that the policymaker can adjust r as a function of μ moderates the distortion but does not upset the property that the loss is smaller the closer μ is to 1.

Varying γ changes the feasible values of μ without affecting the loss incurred from any given μ. In particular, raising γ drives μ_τ further way from 1, brings μ_Y closer to 1, and leaves $\mathscr{L}(\mu)$ unchanged. It follows that $\mathscr{L}(\mu_\tau)$ is an increasing function of γ, whereas $\mathscr{L}(\mu_Y)$ is a decreasing function of it. Next, note that both $\mathscr{L}(\mu_\tau)$ and $\mathscr{L}(\mu_Y)$ are continuous in γ and recall from our earlier discussion that $\mathscr{L}(\mu_\tau) = 0 < \mathscr{L}(\mu_Y)$ when $\gamma = 0$ and $\mathscr{L}(\mu_\tau) > 0 = \mathscr{L}(\mu_Y)$ when $\gamma = 1$. It follows that there exists a threshold $\hat{\gamma}$ strictly between 0 and 1 such that $\mathscr{L}(\mu_\tau) < \mathscr{L}(\mu_Y)$ for $\gamma < \hat{\gamma}$, $\mathscr{L}(\mu_\tau) = \mathscr{L}(\mu_Y)$ for $\gamma = \hat{\gamma}$, and $\mathscr{L}(\mu_\tau) > \mathscr{L}(\mu_Y)$ for $\gamma > \hat{\gamma}$. In a nutshell, because a stronger GE feedback increases the distortion under instrument communication but reduces the distortion under target communication, target communication is optimal if and only if the GE effect is strong enough.

The next result completes the characterization of the optimal strategy by describing the pair (τ, Y) that is obtained for any given θ.

Proposition 7. *For any $\lambda \in [0,1)$ and any $\gamma \in [0,1]$, let $r^{sb} \equiv r(\mu^{sb})$ and $\varphi^{sb} \equiv r(\mu^{sb})/\mu^{sb}$, with $\mu^{sb} = \mu_\tau(\lambda,\gamma)$ if $\gamma < \hat{\gamma}$ and $\mu^{sb} = \mu_Y(\lambda,\gamma)$ if $\gamma > \hat{\gamma}$. Then,*

[18] The expression for r is

$$r(\mu) \equiv \arg\min_r \left[(1-\chi)(r-1)^2 + \chi(r/\mu - 1)^2\right] = \frac{\mu^2(1-\chi)+\mu\chi}{\mu^2(1-\chi)+\chi}$$

(i) If $\gamma < \hat{\gamma}$, the policymaker sets the value $\tau = r^{sb}\theta$ for the instrument and obtains the value $Y = \varphi^{sb}\theta$ for the outcome. If instead $\gamma > \hat{\gamma}$, she sets the target $Y = \varphi^{sb}\theta$ for the outcome and meets this target with the value $\tau = r^{sb}\theta$ for the instrument.

(ii) r^{sb} displays a downward discontinuity at $\gamma = \hat{\gamma}$, is continuous and strictly increasing in γ everywhere else, and satisfies $r^{sb} > 1$ for $\gamma \in (0, \hat{\gamma})$ and $r^{sb} < 1$ for $\gamma \in (\hat{\gamma}, 1)$

(iii) φ^{sb} displays an upward discontinuity at $\gamma = \hat{\gamma}$, is continuous and strictly decreasing in γ everywhere else, and satisfies $\varphi^{sb} < 1$ for $\gamma \in (0, \hat{\gamma})$ and $\varphi^{sb} > 1$ for $\gamma \in (\hat{\gamma}, 1)$

Part (i) follows directly from the preceding analysis and lets r^{sb} and φ^{sb} measure the optimal slope of, respectively, the instrument and the outcome with respect to the underlying fundamental. Parts (ii) and (iii) then follow the characterization of the functions $r(\cdot)$, $\mu_\tau(\cdot)$ and $\mu_Y(\cdot)$. The discontinuity of r^{sb} and φ^{sb} at $\gamma = \hat{\gamma}$ reflects the switch from one form of forward guidance to the other and the flipping of the distortion. When $\gamma < \hat{\gamma}$, the policymaker engages in instrument communication, the friction causes *attenuation*, and the optimal policy moderates the distortion by having τ move *more* than one to one with θ. When instead $\gamma > \hat{\gamma}$, the policymaker engages in target communication, the friction causes *amplification*, and the optimal policy has τ move *less* than one to one with θ.

4.6 Comparative statics and vanishing frictions

Because the model is highly tractable, we can characterize the dependence of the optimal form of forward guidance on all model parameters.

Proposition 8. *The threshold $\hat{\gamma}$, above which target communication is optimal, decreases with χ, increases with α, and decreases with λ.*

The effect of χ is obvious: raising the policymaker's concern about the "output gap" expands the range of γ for which target communication is optimal.

Consider next α. As α approaches 1, τ has a vanishingly small effect on Y for given K. The policymaker may therefore need to make very large adjustments in τ to hit a stated target for Y. This explains why target communication becomes less desirable as α increases.

Finally, consider λ. Raising the belief friction (lowering λ) intensifies the distortion under both modes of communication. As shown in the Appendix, however, the additional friction "bites harder" with target communication than under instrument communication. Conversely, a smaller friction favors target communication.

What happens as the friction vanishes?

We know that when $\lambda = 1$ the threshold $\hat{\gamma}$ is not well-defined because both communication methods are equivalent. Still, there is a well-defined limit in the case of a vanishingly small but positive behavioral distortion:

Corollary 3 (Optimal Forward Guidance with Small Frictions)**.** *Consider a sequence of threshold GE feedbacks, $(\hat{\gamma}_n)_{n=1}^{\infty}$, from applying Theorem 1 to a sequence of economies with constant α and varying behavioral parameter $(\lambda_n)_{n=1}^{\infty}$, where the previous is an increasing sequence with $\lim_{n\to\infty}\lambda_n = 1$. Then $\lim_{n\to\infty}\hat{\gamma}_n = \frac{1}{2-\alpha} \in \left(\frac{1}{2}, 1\right)$.*

This is theoretically like a "selection result" that makes the optimal communication choice continuous in λ. It holds too for sequences of λ that converge from above, capturing the possibility of small amounts of belief over-reaction, which will be discussed more carefully in Section 6.1.

4.7 Environments with strategic substitutability

We conclude this section with a comment on the role played by the restriction $\gamma > 0$. This restriction is consistent with the liquidity-trap application, where the GE effect of monetary policy *adds* to its PE effect. But it rules out environments in which the GE effects of taxes or other policies *offset* their PE effects. This includes situations in which agents compete for scarce resources and can be captured in the Neoclassical investment example studied in Appendix B by letting labor supply be sufficiently inelastic relative to the aggregate demand externality.

Had we allowed for this scenario, the games induced by both forms of forward guidance would display strategic substitutability, but the substitutability would be milder with instrument communication (i.e., $\delta_Y < \delta_\tau < 0$). The basic intuition about reducing the "bite" of higher-order considerations suggests that instrument communication should be necessarily optimal when $\gamma < 0$ and, hence, that our main result (Theorem 1) should remain intact. Appendix D verifies this claim this is true provided an additional, sensible assumption about the maximum possible distortion.

5 Application: Monetary Policy in a Liquidity Trap

In this section, we collect our model's relevant insights to the main motivating example: communication about monetary policy when interest rates are at their zero lower bound.

Clarity via instruments or targets. Our main result was about when the central bank should switch between instrument- and target-focused forward guidance, depending on the ferocity of GE feedback mechanisms. In the ZLB context, these mechanisms include the feedback between aggregate income and aggregate spending, or the Keynesian cross; the dynamic strategic complementarity in the firms' price-setting decisions; and the inflation-spending feedback, which is captured in the New Keynesian model by the interaction of the Dynamic IS curve and the New Keynesian Philips curve.[19]

In terms of our model, the combination of these mechanisms amount to a strong macroeconomic complementary, or a high γ. Furthermore, the results of Angeletos and Lian (2018) suggest that the effective γ increases with the the length of the liquidity trap, because this allows the feedback effects to

[19] Angeletos and Lian (2018) develop a game-theoretic representation of these mechanisms that roughly maps to our framework. The example in Appendix B obtains an *exact* mapping by making enough simplifying assumptions.

compound over more periods. The results of Farhi and Werning (2019), on the other hand, suggest that the effective γ also increases with the severity of liquidity constraints, because such constraints map to a steeper Keynesian cross.

Our analysis thus provides guidance on how the right notion of "clear communication" can change with the state of the economy. In particular, severe demand recessions associated with a long-lasting liquidity trap and/or a steep Keynesian cross may be the most opportune times to engage in target communication.

It is also worth clarifying the following point. Our analysis so far shares the prediction of Angeletos and Lian (2018), Farhi and Werning (2019) and Gabaix (2018) that bounded rationality arrests the power of instrument-based forward guidance. This particular prediction hinges on equating bounded rationality to under-reaction of the beliefs of the behavior of others. If instead such beliefs over-react, this particular prediction flips. That is, the lesson of the aforementioned paper hinges on a plausible but *specific* assumption about higher-order beliefs. By contrast, as shown in Section 6.1, the lesson obtained survives *even when* the opposite bias in beliefs is considered.

Belief frictions in practice. Our results and policy lessons are clearest in the case in which people are attentive to the policy announcement itself (undistorted first-order beliefs) but uncertain or wrong about the other equilibrium objects (distorted higher-order beliefs).

The FOMC's explicit date-based guidance in late 2011 and early 2012, seem to have induced such a pattern. On the one hand, the Fed was able to move expectations about interest rates to an extent that it could not during the previous, weaker regime of communication (see, e.g., Swanson and Williams, 2014); Williams (2016) writes succinctly that

> Compared to the gentle taps of the hammer of previous FOMC verbal guidance that appeared to have little effect on expectations, the time-based guidance was like a sledgehammer.

On the other hand, Andrade et al. (2019) show in professional forecast data that the same focusing of interest rate expectations coincided with an *increase* in the dispersion of forecasts of real variables including consumption. This is consistent with our theory's proposition, that instrument communication is successful in anchoring expectations of interest rates only at the expense of increasing the uncertainty about how the economy would respond to them.

The transition in December 2012 to data-based guidance coincided in turn with a sharp decrease in outcome forecast dispersion (see again Andrade et al., 2019). This is consistent with our theory's proposition that target communication helps anchor expectations of the target outcomes.

Although other explanations of these facts may be possible, their combination adds empirical support to the trade off we have identified in this paper. Under the lenses of our theory, and whether this was intensional or not, the Fed optimally sacrificed confidence about interests (and tranquility in the bond market) in order to anchor the "more useful" expectations about real outcomes.

There is also a useful parallel to Mario Draghi's famous proclamation to do "whatever it takes" to save the Eurozone. In our eyes, this is the sharpest possible example of how a switch from instrument communication to target communication proved immensely effective. We return to this parallel in the concluding section.

6 Robustness

To what extent do our results apply outside the *specific* model of bounded rationality assumed so far? And are they robust to more complex policy trade offs or richer policy options? In this section we demonstrate the robustness of our main insights to several such considerations.

6.1 General bias and animal spirits

So far the analysis has allowed for inertial or anchored beliefs. While this is the scenario studied in the aforementioned literature on forward guidance, two different scenarios are common in other strands of the literature. The first allows for the exact opposite bias in beliefs, namely belief over-reaction (Bordalo, Gennaioli and Shleifer, 2017; Bordalo et al., 2018). The second allows for entirely random variation in beliefs, or for "animal spirits" and "sentiments" without multiple equilibria (Akerlof and Shiller, 2009; Angeletos and La'O, 2013; Benhabib, Wang and Wen, 2015).

Motivated by these observations, we now consider the following, generalized specification of the friction in how agents form beliefs or reason about the behavior of others.

Assumption 4 (General distorted beliefs)**.** *Average beliefs satisfy* $\bar{\mathbb{E}}[K] = \lambda K + \sigma\varepsilon$ *for some* $\lambda > 0$ *and* $\sigma \geq 0$*, where* ε *is a unit-variance noise term unknown to the policymaker and independent of the policy announcement.*

Relative to the main analysis, the friction is now introduced directly in the beliefs about K as opposed to being derived from "first principles" (i.e., from lack of common knowledge of the announcement and/or the rationality of others). This short cut can easily be dispensed with. More importantly, note that our main specification is nested with $\lambda < 1$ and $\sigma = 0$. By contrast, $\lambda > 1$ captures belief over-reaction and $\sigma > 0$ captures animal spirits or sentiments.

The upshot for implementable sets is the following:

Proposition 9. *A pair* (τ, Y) *is implementable if and only if*

$$\tau = \mu_X(\lambda,\gamma)Y + \psi_X(\sigma,\gamma)\varepsilon$$

where $X \in \{\tau, Y\}$ *indexes the mode of communication,* $\mu_\tau(\lambda,\gamma)$ *and* $\mu_Y(\lambda,\gamma)$ *are defined in Proposition 4, and*

$$\psi_\tau(\sigma,\gamma) \equiv -\sigma\alpha\frac{\alpha\gamma}{1-\lambda\alpha\gamma+\alpha^2\gamma(\lambda-1)} \qquad \textit{and} \qquad \psi_Y(\sigma,\gamma) \equiv -\sigma\frac{\alpha^2(1-\gamma)}{(1-\alpha)((1-\alpha)+\lambda\alpha(1-\gamma))}$$

Two remarks are in order. First, compared to the case with anchored beliefs ($\lambda < 1$), the case with over-reactive beliefs ($\lambda > 1$) yields the opposite distortions on the implementability constraints: there is now amplification under instrument communication ($\mu_\tau > 1$) and attenuation under target communication ($\mu_Y < 1$). Nevertheless, the comparative statics of the two distortions with respect to the GE effect remain the same: as γ increases, the distortion under instrument communication gets larger and that under target communication gets smaller. Hence, our main policy result (Theorem 1), and the intuition about minimizing the distortion, also remain.

Second, the distortions induced by animal spirits ($\sigma > 0$) work similarly to the distortions induced by biased beliefs ($\lambda \neq 1$) insofar as one focuses on the interaction of the form of forward guidance and the GE effect.[20] The common mechanism is that a lower weight on beliefs in decisions dampens the effect of any belief mistakes on outcomes, regardless of whether these mistakes are perfectly correlated with the policy announcement ($\lambda \neq 1$ and $\sigma = 0$), entirely uncorrelated ($\lambda = 0$ and $\sigma > 0$), or imperfectly correlated ($\lambda \neq 1$ and $\sigma > 0$).

6.2 Imperfect control, additional shocks, and Poole (1970)

Much of the contemporary discussion of "instrument problems" follows the durable logic of Poole (1970): that the optimal instrument is the one best hedged against confounding shocks.[21] This is undoubtedly an important consideration, but is largely orthogonal to our own lessons.

Consider first the implications of imprecise implementation. The Fed's decision to retreat from an unemployment target in March 2014, as civilian unemployment hit 6.7% (and continued to fall), related quite a bit to the fact that unemployment fell for the "wrong reasons," a fall in labor force participation (Blinder et al., 2017). This resembles the extension studied in Appendix E.2 with measurement errors: some idealized measure of labor market tightness is proxied by an imperfect BEA statistic. The presence of such measurement error tilts the balance away from target communication other things equal, but does not change the conclusion that target communication is preferable if and only if the GE feedback is large enough.

Consider also the *ex ante* argument that target communication might hedge better against an uncertain length to the liquidity trap or other shocks the underlying economic fundamentals. To a first approximation, this is nested in the case studied in Appendix E.1 with unanticipated shocks to Y. By itself, this induces no dependence of the optimal policy on the GE feedback γ, and therefore leaves our main result unaffected.[22]

[20] This captures the scenario of purely extrinsic variation in higher-order beliefs ("sentiments") studied in Angeletos and La'O (2013) and Angeletos, Collard and Dellas (2018). These works have highlighted that the fluctuations sustained by such belief variation in unique-equilibrium models are tightly connected, both conceptually and empirically, to those triggered by sunspots in multiple-equilibrium models. Our own result builds a similar bridge in terms of policy recommendations: the optimality of target communication when γ is high enough can be seen as a unique-equilibrium extension of the logic that optimal communication "selects a good equilibrium" among a large set.

[21] See, for example, Atkeson, Chari and Kehoe (2007) for a recent applicatio. The literature on Taylor rules builds on the same basic logic, too.

[22] To be precise, such shocks induce no dependence of the optimal policy on γ insofar as they are common knowledge, which best represents Poole (1970) and the related literature. Otherwise, their presence can reinforce our message by adding another source of higher-order mistakes, whose bite the policymaker may wish to minimize.

Finally, let us re-interpret τ as an interest rate determined in financial markets on the basis of two objects: a "deeper" instrument under the policymaker direct control and a liquidity or risk-premium shock among financial-market participants. This maps to the case studied in Appendix E.2 with imperfect control of τ. Such imperfect control has the exact opposite effect than the cases discussed above: other things equal, it tilts the balance *towards* target communication. But again, the comparative static of the optimal policy with respect to γ remains intact.

In a nutshell, the policy considerations put forward here are both distinct from and robust to those captured by Poole (1970). The same applies to the literature on optimal Taylor rules, which is essentially a modern, New Keynesian re-incarnation of Poole (1970): this literature is concerned with the optimal policy response to different fundamentals and, unlike our approach, does *not* relate the optimal policy to either the strength of GE or bounded rationality. More on this next.

6.3 Sophisticated forward guidance and policy rules

In reality, central bankers are constantly talking about all kinds of things: instruments, targets, intermediate outcomes, and fundamentals. How our insights translate to the practice of choosing the exact wording of policy communications is of course beyond the scope of our paper. Our main policy recommendation may nevertheless be read as a gauge for when central bankers should *tilt* their focus from offering precise guidance about future interest rates to convincing the market that they will do "whatever it takes" to stabilize the economy.

This flexible interpretation is corroborated by an exercise that expands the forms of forward guidance the policymaker can engage in. It is worked out fully in Appendix G but summarized here.

The policymaker is now allowed to announce and commit to a flexible *relation* between the instrument τ and the outcome Y, given by

$$\tau = A - BY, \tag{13}$$

for some $(A, B) \in \mathbb{R}^2$. This form of forward guidance amounts to announcing the pair of numbers (A, B), instead of the single number $\hat{\tau}$ or $\hat{Y}$. The two simpler strategies considered in our baseline analysis are nested with $B = 0$ and $A = \hat{\tau}$ for instrument communication, and $B \to \infty$ and $A/B \to \hat{Y}$ for target communication. The extension allows the policymaker to choose and communicate an arbitrary pair (A, B), conditional on θ.

In this extension, the analogue of Assumption 1 imposes that each agent is rational and aware of the chosen pair (A, B). If we also impose the analogue of Assumption 2, namely common knowledge of that pair and of the agents' rationality, we once again recover the rational expectations benchmark typically considered in the literature (with indeterminate A and B).

If instead we allow the agents to make mistakes when trying to predict or reason about the responses of others, either of the type formalized before of or of *any* other type, the optimal pair (A, B) becomes determinate. Appendix G works this out carefully leading to the following main result:

Proposition 10. *Suppose that the policymaker can announce and commit on a policy rule as in (13) and let Assumptions 1 and 3 hold with $X = (A, B)$.*

When $\lambda = 1$ (rational expectations), the optimal rule is indeterminate: the first best is implemented with any (A, B) such that $B > -1$ and $A = (1 + B)\theta$.

When instead $\lambda < 1$ (anchored beliefs), the optimal rule is unique: the first best is implemented if and only if

$$B = \frac{\alpha\gamma}{1-\alpha\gamma} \qquad \textit{and} \qquad A = \frac{\theta}{1-\alpha\gamma}. \tag{14}$$

The linear policy rule can achieve first best by *entirely* bypassing the aforementioned friction. It smooths out our main insight in the following sense: as γ increases, the policy rule becomes steeper in (Y, τ) space and hence more precisely steers the economy to a particular level of Y. In fact, as $\gamma \to 1$, the optimal value for B explodes to ∞, recovering our extreme form of target communication as the unconstrained optimal choice. Similarly, $\gamma \to 0$ recovers instrument communication.

The following point is worth clarifying here. In the version of the model studied thus far, bounded rationality drives the economy away from the first best when the policymaker chooses between instrument and target communication. But once the policymaker is given the option to choose a more "sophisticated" policy rule as above, the first best is restored.

Does this mean that the trade off identified in our paper disappears once the policymaker has this option? No. As long as $\lambda \neq 1$, the trade off is present *regardless* of whether the aforementioned option is available or not. Indeed, the policy rule identified in condition (14) is optimal precisely because it minimizes the bite of bounded rationality.

Furthermore, the property that the optimal rule restores the first best is itself an artifact of strong assumptions about the policymaker's knowledge of the structure of the economy. But the rational that the optimal policy minimizes the bite of bounded rationality is not.

Proposition 11. *Suppose the policy maker is uncertain about α or γ.*

When $\lambda = 1$ (rational expectations), the first best is attainable and the optimal rule is indeterminate.

When instead $\lambda < 1$ (anchored beliefs), the first best is unattainable but the optimal rule is determined by the same considerations.

To recap, our main analysis abstracted from the more sophisticated forms of forward guidance allowed in this extension because these may be hard to explain and communicate, especially when the intended audience is the general public and the true environment is richer than our stylized model. Simpler forms of forward guidance, such as "we will keep the policy rate at zero for the next x years" or "we will do whatever it takes to bring unemployment down to z percent" may be more effective than complex rules for reasons left outside the analysis. To paraphrase Minneapolis Fed President Narayana Kocherlakota, we did not want to risk "letting the perfect be the enemy of the good."

Nevertheless, our main insight that the optimal policy minimizes the bite of bounded rationality is robust to the more sophisticated forms of forward guidance considered in the present extension. This extension also suggests a new perspective on policy rules more broadly. Consider, in particular, the literature on optimal Taylor rules for monetary policy. This literature has focused on how such rules can regulate the response of the economy to shocks in fundamentals such as preferences, technology, and monopoly markups when the policymaker cannot directly condition the policy instrument on such shocks. Our own result, instead, indicates how such rules can serve a new function: regulating the impact of bounded rationality, which affects the transmission of all these shocks or could generate new ones. The application of this insight to the type of richer, dynamic, macroeconomic models used in quantitative policy evaluation seems an interesting direction for future research.

6.4 Policy communication, information revelation, and commitment

The shock θ that enters the policymaker's preferences does not enter conditions (1) and (2). This restriction may be at odds with applications, in which the first best typically depends on fundamentals such as preferences and technology that directly affect agents' behavior for given policy. Put differently, our model equates θ to a pure externality.

This assumption was suitable for our purposes because it let us disentangle two mechanisms. The first, which is of interest to us, is the communication of different policy commitments and the associated regulation of the agents' strategic interaction. The second, which is the topic of the literature on the social value of information that follows Morris and Shin (2002), regards the revelation of information about fundamentals that affect the agents' behavior even in the absence of strategic interaction, or more generally holding *fixed* that interaction. A hybrid of the two may be interesting, but is beyond the scope of this paper.

The assumed policy objective also imposes that the first best is obtained in the frictionless benchmark, i.e., bounded rationality is the *only* distortion. This simplification is sufficient, but not strictly needed for our normative conclusions. The following analogy is useful. Consider the sticky-price model of Correia, Nicolini and Teles (2008). Even though the true first best is not attainable, the relevant "ideal point" for the Ramsey planner is one that minimizes the welfare bite of nominal rigidity because the latter does not substitute for missing tax instruments. We suspect that the same logic applies in our context, with "bounded rationality" in place of "nominal rigidity."[23]

Finally, our analysis has assumed that the policymaker has full commitment so as to separate our contribution from a literature that studies how different policy regimes influence the market's ability to detect policy deviations and, thereby, the severity of the time-inconsistency problem (Atkeson, Chari and Kehoe, 2007). That said, it is interesting to note the following. In our rational expectations benchmark, the assumption of commitment was not relevant because even in the absence of it the policymaker implements the same (τ, Y) pair. But once we depart from this benchmark, the ex post optimal policy strategy does not coincide with the ex ante one, because and only because of the mistakes agents make in predicting one another's responses to the policy. This illustrates how bounded rationality can itself be a source

[23] That said, it would be useful to extend the analysis to settings in which the opposite scenario holds. Our positive results regarding the effect of bounded rationality on implementability could continue to apply, but their normative implications would change if that distortion could be used to offset another distortion.

of time inconsistency—an idea that we leave open for future research.

7 Conclusion

What is the best way to manage expectations? Should a policymaker announce and commit to the intended value of the available policy instrument, such as the Federal Funds rate, or the target for the relevant economic outcome, such as employment?

We pose this question in a stylized model in which agents form mis-specified beliefs about each others' actions. Our main result is a sharp dependence of the optimal communication strategy on the GE feedback between aggregate outcomes and individual actions. Fixing outcomes instead of instruments is optimal if and only if this feedback is sufficiently high, as insituations with a prolonged liquidity trap, a steep Keynesian cross, or a large financial accelarator.

Why? Instrument communication pins down expectations of the policy instrument itself, but leaves agents to predict, or reason about, the determination of aggregate outcomes. Target communication does the opposite, sacrificing clarity about the policy instrument for more anchoring of the expectations of targetted outcome.. Which strategy is preferred depends on the relative cost of mistakes for each type of reasoning. High GE feedback, which make outcome expectations more essential for decisions (and associated mistakes more costly), tilt the balance toward target communication. Put more succinctly, the optimal form of forward guidance minimizes agents' need to "reason about the economy," precisely because such reasoning can produce distortions.

An empirical validation of our theory seems difficult given the scant examples of target-based guidance and the likely confounding factors. Nevertheless, the evidence provided in Andrade et al. (2019) supports the main ingredient of our theory, namely a trade-off between controlling market expectations of the policy instrument and controlling expectations of the targeted outcome. What is more, there is a useful parallel with Mario Draghi's famous proclamation to do "whatever it takes" to save the Eurozone.

Under the lens of our theory, this example of target communication was effective, not because it selected one equilibrium out of many, but instead because it eliminated the room for "mistakes" in markets' reasoning about the mapping from the available policy tools to the economy's ultimate fate. At first glance, the two rationales may appear to be very different, the one having to do with multiple equilibria and other not. But they are tightly connected in that they both regard the role of policy in minimizing the impact of higher-order beliefs, or the need to reason about the behavior of others.

It is also quite telling that, at least according to retrospective reports, Draghi himself did not have a precise policy in mind when delivering his now-famous speech. In a November 27, 2018, retrospective written for Bloomberg ("3 Words and $3 Trillion: The Inside Story of How Mario Draghi Saved the Euro"), authors Jana Randow and Alessandro Speciale write:

> After his pledge at Lancaster House to do whatever it takes, Draghi returns to Frankfurt and puts his staff to work turning half-formed plans into a viable program. Some heads of government and central bankers might take Draghi to task for not having a more fully formed

> strategy in the first place, but not Christian Noyer, the former governor of the Bank of France who was part of Draghi's inner circle. Draghi knew what he was doing, Noyer says: "He was relying on the capacity of the system to invent it. That's what I call genius intuition."

Appropriate communication itself, and manipulation of investors' expectations, may have quite literally made new policy strategies implementable. This dependence of what can be implemented on how it is communicated gets to the heart of our analysis.

References

Akerlof, George, and Robert Shiller. 2009. *Animal Spirits: How Human Psychology Drives the Economy, and Why it Matters for Global Capitalism.* Princeton, NJ:Princeton University Press.

Amador, Manuel, and Pierre-Olivier Weill. 2010. "Learning from Prices: Public Communication and Welfare." *Journal of Political Economy,* 118(5): 866–907.

Andrade, Philippe, Gaetano Gaballo, Eric Mengus, and BenoÃ®t Mojon. 2019. "Forward Guidance and Heterogeneous Beliefs." *American Economic Journal: Macroeconomics,* 11(3): 1–29.

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 Alessandro Pavan. 2009. "Policy with Dispersed Information." *Journal of the European Economic Association,* 7(1): 11–60.

Angeletos, George-Marios, and Chen Lian. 2018. "Forward Guidance without Common Knowledge." *American Economic Review,* 108(9): 2477–2512.

Angeletos, George-Marios, and Jennifer La'O. 2009. "Incomplete Information, Higher-order Beliefs and Price Inertia." *Journal of Monetary Economics,* 56: 19–37.

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

Angeletos, George-Marios, Fabrice Collard, and Harris Dellas. 2018. "Quantifying Confidence." *Econometrica,* 86(5): 1689–1726.

Atkeson, Andrew, Varadarajan V. Chari, and Patrick J. Kehoe. 2007. "On the optimal choice of a monetary policy instrument." National Bureau of Economic Research.

Benhabib, Jess, Pengfei Wang, and Yi Wen. 2015. "Sentiments and Aggregate Demand Fluctuations." *Econometrica,* 83(2): 549–585.

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

Bergemann, Dirk, and Stephen Morris. 2016. "Belief-Free Rationalizability and Informational Robustness." *Cowles Foundation Discussion Paper No. 2066.*

Bernanke, Ben S, Mark Gertler, and Simon Gilchrist. 1999. "The financial accelerator in a quantitative business cycle framework." *Handbook of Macroeconomics,* 1: 1341–1393.

Blinder, Alan, Michael Ehrmann, Jakob de Haan, and David-Jan Jansen. 2017. "Necessity as the mother of invention: monetary policy after the crisis." *Economic Policy,* 32(92): 707–755.

Blinder, Alan S. 2018. "Through a Crystal Ball Darkly: The Future of Monetary Policy Communication." *American Economic Review, Papers and Proceedings*, 108: 567–71.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2017. "Diagnostic Expectations and Credit Cycles." *The Journal of Finance*, 73(1): 199–227.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer. 2018. "Over-reaction in Macroeconomic Expectations." *NBER Working Paper No. 24932*.

Caballero, Ricardo, and Alp Simsek. 2017. "A Risk-centric Model of Demand Recessions and Macroprudential Policy." MIT miméo.

Campbell, Jeffrey R., Jonas D. M. Fisher, Alejandro Justiniano, and Leonardo Melosi. 2012. "Macroeconomic Effects of Federal Reserve Forward Guidance." *Brookings Papers on Economic Activity*, 43(1 (Spring)): 1–80.

Chahrour, Ryan. 2014. "Public Communication and Information Acquisition." *American Economic Journal: Macroeconomics*, 6(3): 73–101.

Chari, Varadarajan V., and Patrick J. Kehoe. 1999. "Optimal fiscal and monetary policy." *Handbook of Macroeconomics*, 1: 1671–1745.

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, Yuriy Gorodnichenko, Saten Kumar, and Jane Ryngaert. 2018. "Do You Know That I Know That You Know...? Higher-Order Beliefs in Survey Data." *NBER Working Paper No. 24987*.

Cornand, Camille, and Frank Heinemann. 2008. "Optimal Degree of Public Information Dissemination." *The Economic Journal*, 118(528): 718–742.

Cornand, Camille, and Frank Heinemann. 2015. "Limited Higher Order Beliefs and the Welfare Effects of Public Information." *The Economic Journal*, 118(528): 718–742.

Correia, Isabel, Juan Pablo Nicolini, and Pedro Teles. 2008. "Optimal Fiscal and Monetary Policy: Equivalence Results." *Journal of Political Economy*, 116(1): 141–170.

Crawford, Vincent P, Miguel A Costa-Gomes, and Nagore Iriberri. 2013. "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications." *Journal of Economic Literature*, 51(1): 5–62.

Farhi, Emmanuel, and Ivan Werning. 2019. "Monetary Policy, Bounded Rationality, and Incomplete Markets." *Harvard/MIT mimeo*.

Feroli, Michael, David Greenlaw, Peter Hooper, Frederic S. Mishkin, and Amir Sufi. 2017. "Language after liftoff: Fed communication away from the zero lower bound." *Research in Economics*, 71(3): 452 – 490. Special issue on Macroeconomics.

Gabaix, Xavier. 2014. "A Sparsity-Based Model of Bounded Rationality." *Quarterly Journal of Economics*, 129(4): 1661–1710.

Gabaix, Xavier. 2018. "A Behavioral New Keynesian Model." *Harvard mimeo.*

Garcıa-Schmidt, Mariana, and Michael Woodford. 2019. "Are Low Interest Rates Deflationary? A Paradox of Perfect-Foresight Analysis." *American Economic Review*, 109(1): 86–120.

Hansen, Lars Peter, and Thomas J. Sargent. 2007. *Robustness.* Princeton, NJ:Princeton University Press.

Heinemann, Frank, Rosemarie Nagel, and Peter Ockenfels. 2009. "Measuring strategic uncertainty in coordination games." *The Review of Economic Studies*, 76(1): 181–221.

James, Jonathan G., and Phillip Lawler. 2011. "Optimal Policy Intervention and the Social Value of Public Information." *American Economic Review*, 101(4): 1561–74.

Kiyotaki, Nobuhiro, and John Moore. 1997. "Credit Cycles." *The Journal of Political Economy*, 105(2): 211–248.

Leeper, Eric M., Todd B. Walker, and Shu-Chun Susan Yang. 2013. "Fiscal Foresight and Information Flows." *Econometrica*, 81(3): 1115–1145.

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

Lucas, Robert Jr., and Nancy L. Stokey. 1983. "Optimal Fiscal and Monetary policy in an Economy without Capital." *Journal of Monetary Economics*, 12(1): 55–93.

Mankiw, N Gregory, and Ricardo Reis. 2002. "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve." *Quarterly Journal of Economics*, 1295–1328.

Mertens, Karel, and Morten O. Ravn. 2010. "Measuring the Impact of Fiscal Policy in the Face of Anticipation: A Structural VAR Approach*." *The Economic Journal*, 120(544): 393–413.

Mertens, Karel, and Morten O. Ravn. 2012. "Empirical Evidence on the Aggregate Effects of Anticipated and Unanticipated US Tax Policy Shocks." *American Economic Journal: Economic Policy*, 4(2): 145–81.

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

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

Myatt, David P, and Chris Wallace. 2012. “Endogenous Information Acquisition in Coordination Games.” *Review of Economic Studies*, 79(1): 340–374.

Nagel, Rosemarie. 1995. “Unraveling in Guessing Games: An Experimental Study.” *The American Economic Review*, 85(5): 1313–1326.

Nakamura, Emi, and Jon Steinsson. 2018. “High-Frequency Identification of Monetary Non-Neutrality: The Information Effect.” *The Quarterly Journal of Economics*, 133(3): 1283–1330.

Poole, William. 1970. “Optimal Choice of Monetary Policy Instruments in a Simple Stochastic Macro Model.” *The Quarterly Journal of Economics*, 84(2): 197–216.

Powell, Jerome H. 2016. “Discussion of the paper "Language after Liftoff: Fed Communication Away from the Zero Lower Bound.” Remarks at the 2016 US Monetary Policy Forum.

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

Stahl, Dale. 1993. “Evolution of Smart Players.” *Games and Economic Behavior*, 5(4): 604–617.

Swanson, Eric T., and John C. Williams. 2014. “Measuring the Effect of the Zero Lower Bound on Medium- and Longer-Term Interest Rates.” *American Economic Review*, 104(10): 3154–85.

Weitzman, Martin. 1974. “Prices vs. Quantities.” *Review of Economic Studies*, 41(4): 477–491.

Williams, John C. 2016. “Discussion of "Language after Liftoff: Fed Communication Away from the Zero Lower Bound.” Remarks at the 2016 US Monetary Policy Forum.

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

A Proofs

Proof of Proposition 5

The relationship between action K and announcement $\hat{X}$, as derived in the main text, is the following:

$$K = \frac{1-\delta_X}{1-\lambda\delta_X}\hat{X}$$

Instrument communication. As shown in Proposition 5,

$$\mu_\tau(\lambda,\gamma) = \left((1-\alpha) + \alpha\frac{1-\delta_\tau}{1-\lambda\delta_\tau}\right)^{-1} \tag{15}$$

Clearly, for $\delta_\tau \equiv \alpha\gamma \in (0,1)$, as implied by $\gamma \in [0,1]$ and $\alpha \in (0,1)$, $(1-\delta_\tau)/(1-\lambda\delta_\tau) \in [0,1]$ and $\mu_\tau^{-1} \in [0,1]$ and $\mu_\tau \geq 1$.

Further, $\partial\mu_\tau^{-1}/\partial\lambda > 0$ given $\delta_\tau \in (0,1)$ and $\partial\mu_\tau/\partial\lambda = -(\mu_\tau)^{-2}\partial\mu_\tau^{-1}/\partial\lambda < 0$.

When $\delta_\tau < 0$, we can have $\mu_\tau < 1$. A sufficient condition for this is $\gamma < 0$, or negative GE feedback.

Target communication. Let b denote the responsiveness of the action to the announcement, $\partial K/\partial\hat{Y}$. In general, the slope of the implementability constraint is

$$\mu_Y(\lambda,\gamma) = \frac{1-\alpha b}{1-\alpha} = \frac{1-\lambda\delta_Y - \alpha(1-\delta_Y)}{(1-\alpha)(1-\lambda\delta_Y)} \tag{16}$$

Given that $\delta_Y \leq 0$, we know that $b \geq 1$ and hence $\mu_Y \leq 1$.

To check the derivative with respect to λ, note that

$$\frac{\partial b}{\partial\delta_Y} = -\frac{\delta_Y(\delta_Y-1)}{(1-\lambda\delta_Y)^2} > 0$$

and $\partial\delta_Y/\partial\gamma = \alpha/(1-\alpha) > 0$ and $\partial\mu_Y/\partial b = -\alpha/(1-\alpha) < 0$. Thus, by the chain rule, $\partial\mu_Y/\partial\gamma < 0$.

Further results

Lemma 5 (Sign of μ_Y)**.** *$\mu_Y > 0$ if and only if $\lambda \geq \alpha$ or $\gamma > \frac{1+\alpha(\lambda-2)}{\alpha(\lambda-\alpha)}$.*

Proof. Note that $\mu_Y \in [0,1]$ when $b \in [1, 1/\alpha]$ and $\mu_Y < 0$ when $b > 1/\alpha$. This reduces to to

$$\gamma\alpha(\lambda-\alpha) < 1-\alpha(2-\lambda)$$

Let's consider three cases of this. First, assume that $\lambda > \alpha$. Some algebraic manipulation yields the condition

$$\gamma < 1 + \frac{(1-\alpha)^2}{\alpha(\lambda-\alpha)}$$

which is obviously true for any $\gamma < 1$. Thus no more restrictions are required.

Next, consider $\lambda = \alpha$. The condition becomes

$$\alpha(2-\alpha) < 1$$

which is always true for $\alpha = \lambda \in (0,1)$.

Finally, consider $\lambda < \alpha$. In this case, the condition is

$$\gamma > \frac{1+\alpha(\lambda-2)}{\alpha(\lambda-\alpha)}$$

Note that the right-hand-side is less than 0 if $\lambda > 2 - \frac{1}{\alpha}$. Hence we used this as a sufficient condition for $\mu_Y > 0$ for all $\gamma \geq 0$. □

Lemma 6. *Assume that $\mu_Y > 0$ and $\alpha\gamma < 1$. Then $\mu_\tau > \mu_Y$.*

Proof. As long as $\mu_Y > 0$, we can show that $\mu_\tau > \mu_Y$. Written out in terms of parameters, this condition is:

$$\frac{1-\lambda\alpha\gamma}{(1-\alpha)(1-\lambda\alpha\gamma)+\alpha(1-\alpha\gamma)} \geq \frac{1+\frac{\lambda\alpha(1-\gamma\}}{1-\alpha}-\alpha\frac{1-\alpha\gamma}{1-\alpha}}{1-\alpha+\lambda\alpha(1-\gamma)}$$

Given that $\mu_Y > 0$, the left denominator is positive. The other three terms are necessarily positive. Thus an equivalent statement, after cross-multiplying, is the following:

$$(1-\lambda\alpha\gamma)(1-\alpha+\lambda\alpha(1-\gamma)) \geq \left((1-\lambda\alpha\gamma)+\frac{\alpha(1-\alpha\gamma)}{1-\alpha}\right)(1-\alpha+\lambda\alpha(1-\gamma)-\alpha(1-\alpha\gamma)$$

Subtracting like terms from each side, and dividing by $\alpha > 0$, yields the following condition:

$$(1-\lambda)(1-\alpha\gamma) \geq 0$$

Hence $\lambda < 1$ and $\alpha\gamma < 1$ are a sufficient condition for $\mu_\tau > \mu_Y$, and either $\lambda = 1$ or $\alpha\gamma = 1$ are a sufficient condition for $\mu_\tau = \mu_Y$. □

Proof of Proposition 6

Limit cases. At $\gamma = 1$, the slope given instrument communication is

$$\mu_\tau(\lambda, 1) = \left((1-\alpha)+\alpha\frac{1-0}{1-\lambda\cdot 0}\right)^{-1} = \frac{1}{1-\alpha} > 1.$$

Meanwhile, the slope with target communication is

$$\mu_Y(\lambda, 1) = 1$$

At the other extreme $\gamma = 0$, the slope given target communication is

$$\mu_Y(\lambda, 0) = \frac{1-\alpha\frac{1-\lambda}{1-\alpha}}{1-\alpha(1-\alpha)}$$

This is less than one if and only if $1-\alpha < (1-\lambda)/(1-\alpha) < \alpha^{-1}$ or $(1-\alpha)^2 < 1-\lambda < (1-\alpha)\alpha$. This is implied by the arguments of Proposition 5.

With instrument communication at $\gamma = 0$, the slope is $\mu_\tau(\lambda, 0) = ((1-\alpha)+\alpha\cdot 1)^{-1} = 1$.

Derivative of μ_τ with respect to γ. For fixed λ, we can calculate first a derivative of the inverse slope with respect to the interaction parameter

$$\frac{\partial\mu_\tau^{-1}(\lambda,\gamma)}{\partial\delta_\tau} = -\frac{\alpha(1-\lambda)}{(1-\lambda\gamma)^2}$$

which is unambiguously negative for $\lambda < 1$. The interaction parameter $\delta_\tau := \alpha\gamma$ increases with γ. Thus, by the chain rule, $\partial\mu_\tau/\partial\delta_\tau = -(\mu_\tau)^{-2}(\partial\mu_\tau^{-1}/\partial\delta_\tau)(\partial\delta_\tau/\partial\gamma) > 0$.

Derivative of μ_Y with respect to γ. For fixed λ, the partial derivative with respect to δ_Y is

$$\frac{\partial \mu_Y}{\partial \delta_Y} = \frac{\alpha(1-\lambda)}{(1-\alpha)(1-\lambda\delta_Y)^2} > 0$$

The interaction parameter $\delta_Y \equiv (\gamma-1)\alpha/(1-\alpha)$ increases with γ. Hence $\partial\mu_Y/\partial\gamma > 0$. Note that this argument made no reference to the fact that $\mu_Y \geq 0$.

Proof of Theorem 1

Let $r \equiv \tau/\theta$. The problem is, up to scale,

$$\min_{\mu \in \{\mu_\tau(\lambda), \mu_Y(\lambda)\}, r \in \mathbb{R}} (1-\chi)(r-1)^2 + \chi(r/\mu - 1)^2$$

We can concentrate out the parameter r with the following first-order condition

$$r^*(\mu) := \frac{\mu^2(1-\chi) + \mu\chi}{\mu^2(1-\chi)+\chi} \tag{17}$$

In this quadratic problem, the first-order condition is sufficient. We can further deduce that, given $\chi \in (0,1)$, $r^*/\mu > 1$ for $\mu \in [0,1]$, $r^*/\mu < 1$ for $\mu > 1$, and $r^*/\mu = 1$ for $\mu = 1$. Further, $r > 0$ as long as $\mu > 0$.

Let $\mathscr{L}(\mu)$ denote the loss function evaluated at this optimal r^*. Note that, from the envelope theorem, $\partial\mathscr{L}/\mu = -2\cdot\chi\cdot r^*\cdot(r^*/\mu - 1)/\mu^2$. Combined with the previous expression for r^*, this suggests that $\partial\mathscr{L}/\mu = 0$ when $\mu = 1$, $\partial\mathscr{L}/\partial\mu > 0$ when $\mu > 1$, and $\partial\mathscr{L}/\partial\mu < 0$ when $\mu \in [0,1]$.

Finally, let $\mathscr{L}_\tau$ and $\mathscr{L}_Y$ denote the value of the loss function evaluated at $r^*(\mu)$ and, respectively, μ_τ and μ_Y. For fixed λ and α, we let $\mathscr{L}_\tau(\gamma)$ and $\mathscr{L}_Y(\gamma)$ denote these losses as function of γ. Note that, by the chain rule, $\partial\mathscr{L}_\tau/\partial\gamma = \partial\mathscr{L}/\partial\mu\cdot\partial\mu_\tau/\partial\gamma$ and $\partial\mathscr{L}_Y/\partial\gamma = \partial\mathscr{L}/\partial\mu\cdot\partial\mu_Y/\partial\gamma$. We will argue that these functions cross exactly once at some $\hat{\gamma}$, the critical threshold of GE feedback.

From here, we branch off the analysis for different domains of the parameters.

Simplest case. Consider the first parameter case covered in Lemma 5.

Note that $\mathscr{L}_\tau(0) = \mathscr{L}_Y(1) = 0$ and both functions are strictly positive elsewhere, by normalization. Since these functions are continuous, there exists (at least one) crossing point $\hat{\gamma} \in [0,1]$ such that $\mathscr{L}_\tau(\hat{\gamma}) = \mathscr{L}_Y(\hat{\gamma})$.

In particular, $\mathscr{L}_\tau(\gamma)$ is strictly increasing and $\mathscr{L}_Y(\gamma)$ is strictly decreasing on the domain $\gamma \in (0,1)$. By the previous argument, to show $\partial\mathscr{L}_\tau/\partial\gamma > 0$ and $\partial\mathscr{L}_Y/\partial\gamma < 0$, it suffices to show that $\partial\mu_\tau/\partial\gamma > 0$, $\partial\mu_Y/\partial\gamma > 0$, and $\mu_\tau > 1 > \mu_Y$. All three are established in Proposition 5.

Possibility of $\mu_Y < 0$. Now let us assume $\lambda < 2 - 1/\alpha$. There now exists a threshold

$$\underline{\gamma} \equiv \frac{1+\alpha(\lambda-2)}{\alpha(\lambda-\alpha)} \in [0,1)$$

such that, for $\gamma < \underline{\gamma}$, $\mu_Y < 0$. For $\gamma \in [\underline{\gamma}, 1]$, we can apply the same logic as previously. It remains to show that instrument communication is optimal for $\gamma \in [0, \underline{\gamma})$.

First, note that $\partial \mathscr{L}_Y / \partial \gamma \leq 0$ as long as $r^*(\mu_Y) \geq 0$. The latter is true as long as $\mu_Y \geq -\chi/(1-\chi)$, which also implicitly defines a threshold $\check{\gamma}$ since μ_Y increases strictly in γ. Clearly the previous argument works for $\gamma \in [\check{\gamma}, 1]$, and it remains only to check $\gamma \in [0, \check{\gamma})$.

On this domain, $\partial \mathscr{L} / \partial \mu > 0$ since $r^*(\mu_Y) < 0$. But we also know that $\lim_{\mu \to -\infty} \mathscr{L}(\mu) = \chi$. This can be verified by direct calculation, or intuited by noticing that $\lim_{\mu \to -\infty} r^*(\mu) = 1$. Since μ_Y strictly increases in γ, it follows that $\mathscr{L}_Y(\gamma) > \chi$ for $\gamma \in (-\infty, \check{\gamma}]$. Meanwhile, a similar argument for $\mu > 1$ (with $\lim_{\mu \to \infty} \mathscr{L}(\mu) = \chi$ and $\partial \mathscr{L}/\partial \mu > 0$) suggests that $\mathscr{L}_\tau(\gamma) < \chi$ for $\gamma \geq 0$. This shows that $\mathscr{L}_Y(\gamma) > \chi > \mathscr{L}_\tau(\gamma)$ on this domain and thus instrument communication is strictly preferred.

It is worth pointing out that the limiting arguments for μ are "loose," since both μ_τ and μ_Y have finite limits:

$$\lim_{\gamma \to -\infty} \mu_\tau = \mu_{\tau,-\infty} \equiv \frac{\lambda}{\lambda + (1-\lambda)\alpha} \in (0,1) \tag{18}$$

$$\lim_{\gamma \to -\infty} \mu_Y = \mu_{Y,-\infty} \equiv \frac{\lambda(1-\alpha/\lambda)}{\lambda(1-\alpha)} \tag{19}$$

Proof of Proposition 8

The critical GE feedback threshold satisfies $\mathscr{L}_\tau(\hat{\gamma}) = \mathscr{L}_Y(\hat{\gamma})$. Plugging directly into the loss function produces a quadratic equation for the threshold. Of the two roots, the following one is in the correct domain $\gamma \in [0,1]$:

$$\hat{\gamma} = \left(1 - \alpha(1-\chi\alpha)(1-\lambda) + \left(\alpha(\alpha - 2\lambda - 2\alpha(1-\lambda)\chi + (1-\alpha(1-\lambda)(1-\alpha\chi))^2\right)^{\frac{1}{2}}\right)^{-1}$$

With this expression, we can do analytical comparative statics.

Policy parameter α. The partial derivative $\partial\hat{\gamma}/\partial\alpha$, up to a strictly positive constant C, is

$$\begin{aligned}\frac{\partial \hat{\gamma}}{\partial \alpha} \cdot C = (1-2\alpha\chi)\left(1 - \frac{\alpha(1-\lambda)(1-\alpha\chi)}{\sqrt{(\alpha(1-\lambda)(1-\alpha\chi))^2 + (1-\alpha)^2}}\right) \\ + \frac{1-\alpha}{\sqrt{(\alpha(1-\lambda)(1-\alpha\chi))^2 + (1-\alpha)^2}}\end{aligned}$$

First, consider the case of $2\alpha\chi < 1$. It remains to show that the term in parenthesis is positive. A sufficient condition for this is

$$1 - 2\alpha(1-\lambda)(1-\alpha\chi) - \alpha(2\alpha(1-\lambda)\chi + 2\lambda - \alpha) > 0$$

Canceling out terms, the above reduces to $(1-\alpha)^2 > 0$, which is trivially true for all $\alpha \in (0,1)$. Thus $\hat{\gamma}$ decreases with λ.

Next, consider the case $2\alpha\chi > 1$. We can re-write the expression as

$$\begin{aligned}\frac{\partial \hat{\gamma}}{\partial \alpha} \cdot C = (1-\alpha\chi)^2\left(1 - \frac{\alpha(1-\lambda)(1-\alpha\chi)}{\sqrt{(\alpha(1-\lambda)(1-\alpha\chi))^2 + (1-\alpha)^2}}\right) \\ + \frac{1-\alpha+(\alpha\chi)^2}{\sqrt{(\alpha(1-\lambda)(1-\alpha\chi))^2 + (1-\alpha)^2}} - (\alpha\chi)^2\end{aligned}$$

Note that the large denominator is bounded by $\sqrt{\alpha^2 + (1-\alpha)^2}$ and also bounded by one. Thus we can show that all terms are positive, and $\partial\hat{\gamma}/\partial\alpha > 0$.

Attentive fraction λ. Up to a (different) positive constant, the relevant partial derivative is

$$\frac{\partial\hat{\gamma}}{\partial\lambda} \cdot C = \frac{\alpha(1-\lambda)(1-\alpha\chi)}{\sqrt{(\alpha(1-\lambda)(1-\alpha\chi))^2 + (1-\alpha)^2}} - 1$$

By the intermediate step of the previous argument, this is negative and thus γ decreases with λ.

Output gap parameter χ. The relevant partial derivative (up to a constant) is equal to the previous one:

$$\frac{\partial\hat{\gamma}}{\partial\chi} \cdot C = \frac{\partial\hat{\gamma}}{\partial\lambda}$$

Hence we know it is negative, and $\hat{\gamma}$ decreases with χ.

B Micro-foundations

In this appendix we spell out the details of two micro-foundations that can be nested in our framework.

B.1 A New Keynesian economy

Here we describe our example of a stylized New Keynesian economy during a liquidity trap. We first set up the economy and then show how to map it to our abstract framework. As noted in the main text, this nesting depends on strong, simplifying assumptions. The goal is only to facilitate an appealing interpretation of our insights. A careful adaptation of our analysis to the full New Keynesian model is beyond the scope of this paper.

Set-up. Consider a simplified version of the textbook New Keynesian model, with perfectly rigid prices and no capital. There are countably infinite periods, indexed by $t \in \{0, 1, 2, \ldots\}$. As in the abstract model, period 0 exists only to index the time of forward guidance. Periods 1 and 2 will be most relevant for our analysis: $t = 1$ corresponds to the liquidity trap, when the zero lower bound is binding; and $t = 2$ to the phase right after the liquidity trap, when the central bank may keep the interest rate below the natural rate in an attempt to stimulate spending during the trap. The "infinite future" thereafter plays no essential role, it only define the phase in which the economy reverts to steady state and nothing interesting happens.

There is a unit measure of consumers, each of which consumes $C_{i,t}$ of the good and has the following utility function:

$$\mathcal{U}_{i,t} = \mathbb{E}_i \left[\sum_{t=1}^{\infty} \beta_t \log C_{i,t} \right]$$

for $\beta_t = \exp\left(-\sum_{j=1}^{t} \rho_j\right)$. Each consumer also faces a standard flow budget constraint in terms of her asset level $A_{i,t}$, income $Y_{i,t}$, and *real* interest rate R_t:

$$C_{i,t} + R_t^{-1} A_{i,t} = A_{i,t-1} + Y_{i,t}$$

The assets are in zero net supply. Income is commonly shared among all agents, so $Y_{i,t} \equiv Y_t$.

A monetary authority controls the real interest rate R_t. Output is completely demand determined, or $\int_i C_{i,t}\, di = Y_t$.

For all $t \geq 2$, the subjective discount rate is $\rho_t = \bar{\rho} > 0$ and the gross natural rate of interest is $\bar{R} = \exp(\bar{\rho}) > 1$. At $t = 1$, the discount rate is negative ($\rho_1 = \underline{\rho} < 0$) and the corresponding gross natural rate is less than 1. The zero lower bound becomes binding, or $R_1 = 1$, and the monetary authority cannot restore the flexible-price (and efficient) level of output. It can, however, set the interest rate in the period after *after* exiting the liquidity trap at a level below the natural rate, namely $R_2 \in [1, \bar{R})$. By offering forward guidance at $t = 0$ about what it will do at $t = 2$, the monetary authority may thus influence consumer spending and output *during* the liquidity trap.

The authority may announce either the post-trap interest rate, R_2, or a target for output (to be defined clearly later). Consumers, however, may have mis-specified beliefs about each other's attentiveness to the announcement. We assume that this affects beliefs at $t = 1$ but not at $t = 2$, at which point the interest rate and level of output become common knowledge.

Key equilibrium conditions. Let all lowercase variables now be in log deviations from the steady state in which $R = \bar{R}$.

The consumption of agent i at time t can be expressed as the following function of current and future interest rates, income, and discount rate shocks

$$c_{i,t} = (1-\beta)\sum_{j=0}^{\infty}\beta^j \mathbb{E}[y_{t+j}] + \beta\sum_{j=0}^{\infty}\beta^j \mathbb{E}[-(r_t - (\rho_t - \bar{\rho}))] \tag{20}$$

where $\beta = \exp(-\rho)$ is the steady-state discount factor. This expression is obtained by substituting the lifetime budget constraint into the consumer's Euler equation for inter-temporal decisions. It can also be interpreted, absent any micro-foundation, as a reduced-form "permanent income consumption function": agents consume fraction $1-\beta$ of the present discounted value of their income, with further adjustment based on the interest rate and patience shock

Let us first derive consumption and income at $t = 2$. We assume that, at this point, all agents have the same (rational) expectations. Furthermore, in our construction of a liquidity trap, we have assumed that the economy returns to a steady state of $c_t = y_t \equiv 0$ and $\rho_t = \bar{\rho}$ for $t \geq 2$. Condition ((20)) now gives consumption as a function of contemporaneous income and the next period interest rate:

$$c_{i,2} = (1-\beta)\mathbb{E}_i[y_2] + \beta\mathbb{E}_i[-r_2]$$

Imposing market clearing and rational expectations gives $c_2 = y_2 = -r_2$. Let us assume that these equilibrium relations are known to all agents in period 0.

Now we can solve for consumption in period 0. The same consumption function, given that the interest rate r_1 equals $-\rho$ in deviation from the steady state, reduces to the following:

$$c_{i,1} = (1-\beta)\mathbb{E}_i[y_1 + \beta y_2] + \beta^2\mathbb{E}_i[-r_2] + \beta(\bar{\rho} - \underline{\rho}) \tag{21}$$

Mapping to the abstract model. Let

$$Y \equiv \frac{y_1 + \beta y_2}{1+\beta}$$

be a measure of output during and right after the liquidity trap, $K \equiv c_1$ be consumer spending during the trap, and $\tau \equiv -r_2$ be the negative of the interest rate right after the trap. Because $y_2 = -r$, we can re-write the definition of Y in the following form:

$$Y = \tfrac{\beta}{1-\beta}\tau + \tfrac{1}{1-\beta}K$$

which matches condition (1) in our abstract framework for $\alpha = \frac{1}{1-\beta}$. The direct effect of policy occurs at $t = 2$. Condition (21), on the other hand, can be written, up to a constant, as

$$K_I = \beta^2 \mathbb{E}_i[\tau] + (1-\beta^2)\mathbb{E}_i[Y]$$

which matches condition (2) in our abstract framework for $\gamma = 1-\beta^2$. The GE complementarity is highest when agents are relatively impatient. With richer micro-foundations, this may correspond to longer horizons (Angeletos and Lian, 2018) or tighter liquidity constraints (Farhi and Werning, 2019).

Unlike what was the case in our neoclassical investment example, the present example has the "unpleasant" property that one deep parameter controls both of the reduced-form parameters γ and α. In particular, as β gets smaller, the GE feedback gets stronger (γ increases), which favors target communication; but the central bank's ability to honor output commitments also gets weaker ($1-\alpha$ falls), which favors instrument communication. Which force dominates for the comparative static is a quantitative question, which our stylized model is not fitted to address. Our insights about the size and direction of belief distortions, though, remain true. In particular, the central bank always obtains an *amplified* response to forward guidance if it announces an output target rather than an interest rate target.

B.2 A neoclassical economy with aggregate demand externalities

Set-up. The second micro-foundation differs in its approach (Neoclassical), application (fiscal policy), and key decision (investment). Appendix B works out all the details. Here, we sketch the main points. There are three periods, $t \in \{0,1,2\}$; a continuum of firms or entrepreneurs, $i \in [0,1]$, who choose investment at $t = 1$; and a policymaker, who can subsidize production at $t = 2$. The first period, $t = 0$, identifies only the time of policy announcement.

At $t = 1$, the entrepreneur has one unit of a good to consume or transform into an investment good. The latter can be sold to a final goods firm at $t = 2$ for price p_i. Their budget is therefore given by $c_{i,1} + x_i = 1$ at $t = 1$ and by $c_{i,2} = p_i x_i$ at $t = 2$, where $c_{i,t}$ denotes consumption in period t. Their lifetime utility is linear, $u_i = c_{i,1} + c_{i,2}$.

The final-good firm operates at $t = 2$. Its output is $Q = X^{\eta} N^{1-\eta}$ and its revenue $(1-r)Q - wN - \int p_i x_i \mathrm{d}i$, where r is the rate of taxation, $X \equiv (\int x_i^{1-\rho} \mathrm{d}i)^{1/1-\rho}$ is a CES aggregator of the differentiated capital goods, N is the labor input supplied by the worker, and $\rho \in [0,1]$ and $\eta \in [0,1]$ parametrize, respectively, the inverse elasticity of substitution of the differentiated inputs and the income share of capital. Finally, the worker lives, works, and consumes only in period $t = 2$ and has utility $v = wN - \frac{1}{1+\phi}N^{1+\phi}$, where w is the real wage and $\phi > 0$ parameterizes the Frisch elasticity.

Solution. It is easiest to solve this model backward in time.

In period 2, the final goods producer's demand for intermediates is the following:

$$p_i = \eta(1-r)QX^{\rho-1}x_i^{-\rho}$$

where X is the CES aggregator of the individual x_i. This implies that the revenue for the entrepreneur has the following form:

$$p_i \cdot x_i = \eta(1-r)Y\left(\frac{x_i}{X}\right)^{1-\rho} = \eta(1-r)X^{\eta+\rho-1}N^{1-\eta}x_i^{1-\rho}$$

Profits scale more with aggregate investment X when ρ is high (high complementarity and high demand externality).

Labor supply has the following form:

$$w = (1+\phi)N^{\phi}$$

Labor demand is set by the final-goods firm:

$$w = (1-\eta)(1-r)\frac{Q}{N}$$

which decreases in the tax rate (or increases in the subsidy).

In period 1, the entrepreneur invests until the marginal return on capital is one:

$$1 = \mathbb{E}_i\left[\frac{\partial(x_i \cdot p_i)}{x_i}\right]$$

The first-order condition re-arranges to

$$x_i^{\rho} = \eta(1-\rho)\mathbb{E}_i\left[(1-r)X^{\eta+\rho-1}N^{1-\eta}\right] \tag{22}$$

Investment solves this fixed-point equation.

REE benchmark. Assume rational expectations with no uncertainty. In equilibrium, the agent will conjecture that $x_{-i} = x_i \equiv X$. Since everything is now known, we can pull X out of the expectation and solve to get

$$X_i = X = (\eta(1-\rho))^{\frac{1}{1-\eta}}(1-r)^{\frac{1}{1-\eta}}N$$

It is immediate that output is linear in labor:

$$Q = X^{\eta}N^{1-\eta} = (\eta(1-\rho))^{\frac{\eta}{1-\eta}}(1-r)^{\frac{\eta}{1-\eta}}N$$

Setting labor supply to labor demand gives

$$N = \left(\frac{1-\eta}{1-\phi}\right)^{\frac{1}{1+\phi}}(1-r)^{\frac{1}{1+\phi}}Q^{\frac{1}{1+\phi}}$$

and plugging that back into the equation for output gives

$$Q = \left(\frac{1-\eta}{1-\phi}\right)^{\frac{1}{\phi}} (\eta(1-\rho))^{\frac{\eta}{1-\eta}\frac{1-\phi}{\phi}} (1-r)^{\frac{\eta}{1-\eta}\frac{1-\phi}{\phi}+\frac{1}{\phi}}$$

From this point, we can also solve for output as a function of investment X. Crucially, none of the exponents (i.e., elasticities) depend on the value of ρ: only the constants (levels) do.

Log-linear approximation. Now consider a more general model in which agents do not form rational expectations, because of either limited information or various behavioral biases. The fixed-point equation 22 can no longer be solved without expectations. To make progress, we will take log-linear approximations around $r = 0$. Let $(\bar{Q}, \bar{N}, \bar{X})$ denote output, labor, and investment evaluated at this point. Let $Y = \log Q - \log \bar{Q}$ and and $n = \log N - \log \bar{N}$ be log deviations of the first two quantities. Further, define $k_i = \frac{1+\eta\phi}{1+\phi}(\log x_i - \log \bar{X})$ and $\tau = \frac{1+\eta\phi}{\phi(1-\eta)}\log(1-r)$ be convenient monotonic transformations of investment and the tax, respectively, and $K = \int_i k_i \,\mathrm{d}i$ be the aggregate (log deviation) rescaled investment.

Aggregate production is log-linear:

$$Y = \frac{\eta(1+\phi)}{1+\eta\phi}K + (1-\eta)n$$

Equilibrium labor is

$$n = \frac{1}{1+\phi}Y + \frac{\phi(1-\eta)}{(1+\eta\phi)(1+\phi)}\tau$$

Combining these two expressions yields the following expression for output as a function of investment and policy:

$$Y = (1-\alpha)\tau + \alpha K \tag{23}$$

with

$$\alpha \equiv \frac{\eta(1+\phi)^2}{(\eta+\phi)(1+\eta\phi)} \tag{24}$$

The direct effect of policy, with weight $1-\alpha$, comes entirely through the expansion of labor demand. Unsurprisingly, this effect is strongest when the capital share of output η is relatively small

Let us now turn to the investment decision (22). To a log-linear approximation, it is

$$\log x_i - \log \bar{X} = \left(1 - \frac{1-\eta}{\rho}\right)\mathbb{E}_i[\log X - \log \bar{X}] + \frac{1-\eta}{\rho}\mathbb{E}_i[n] + \frac{1}{\rho}\mathbb{E}_i[\log(1-r)]$$

After substituting in equilibrium labor, rescaling investment and taxes, and approximating aggregate investment, we get

$$k_i = \left(1-\gamma\right)\mathbb{E}_i[\tau] + \gamma \mathbb{E}[Y] \tag{25}$$

for feedback parameter

$$\gamma \equiv \frac{(1+\eta\phi)(\rho(\eta+\phi) - \phi(1-\eta))}{\eta\rho(1+\phi)^2} \tag{26}$$

For all $\phi > 0$, $\rho \in (0,1)$, and $\eta \in (0,1)$, this parameter is in the relevant domain $(-\infty, 1]$. A higher aggregate demand externality always corresponds to a larger feedback:

$$\frac{\partial \gamma}{\partial \rho} = \frac{(1-\eta)(1+\eta\phi)\phi}{\eta\rho^2(1+\phi)^2} > 0$$

The feedback parameter is positive if and only if

$$\rho > \frac{\phi(1-\eta)}{\phi+\eta}$$

The right-hand-side is always strictly less than 1. The game more likely has reduced-form complementarity when the capital share is relatively high or the disutility of labor is relatively low. If wages are perfectly sticky, or $\phi = 0$, then the right-hand side is zero and there is always a (net) aggregate demand externality.

C Level-k Thinking

The key mechanism in the previous section is agents' under-forecasting of others' responses to the policy message: as demonstrated in Lemma 3, $\bar{\mathbb{E}}[K]$ moves less than K in response to variation in $\hat{X}$. One could recast this as the consequence of agents' bounded ability to calculate others' responses or to comprehend the GE effects of the policy.

A simple formalization of such cognitive or computational bounds is Level-k Thinking. This concept represents a relaxation of the part of Assumption 2 that imposes common knowledge of rationality: agents play rationally themselves, but question the rationality of others. In particular, this concept is defined recursively by letting the level-0 agent make an exogenously specified choice (this is the completely irrational agent), the level-1 agent play optimally given the belief that others are level-0 (this agent is rational but believes that others are irrational), the level-2 agent play optimally given the belief that others are level-1, and so on, up to some finite order k. Level-k Thinking therefore imposes a pecking order, with every agent believing that others are less sophisticated than herself in the sense that they base their beliefs on fewer iterations of the best responses than she does.

To see the implications of this concept in our context, assume all agents think to the same order $k \geq 1$ and let the "base case" (level-0 behavior) correspond to $K = 0$. Because level-k agents believe that all other agents are of cognitive order $k-1$, the expectation of K is now given by

$$\bar{\mathbb{E}}[K] = (1-\delta_X)\sum_{h=0}^{k-1}(\delta_X)^h \hat{X} = (1-(\delta_X)^k)\hat{X} \tag{27}$$

For *even* k and $\delta_X \in (-1,1)$, this always implies a dampened response of beliefs to the fundamental. Outcomes $K = ((1-\delta_X)+\delta_X(1-(\delta_X)^k))\hat{X}$ have dampened response to $\hat{X}$ for $\delta_X > 0$ and amplified response for $\delta_X < 0$. These distortions remain monotone in the extent of strategic interaction in either direction, $|\delta_X|$. Intuitively, higher $|\delta_X|$ puts higher weight on agents' faulty reasoning. As such our core results readily extend to this case.

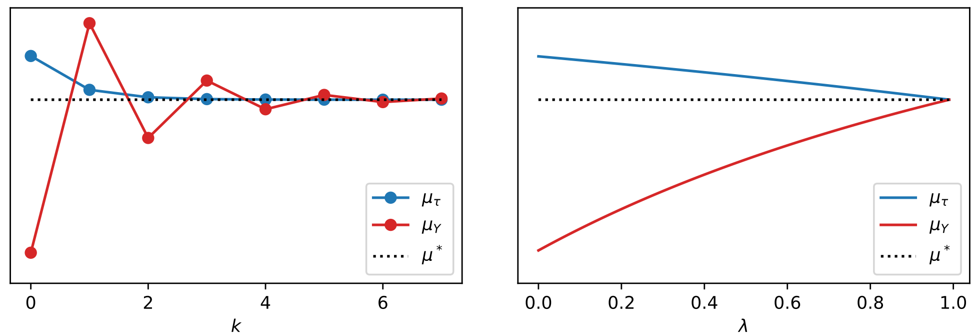

Figure 1: The implementability coefficients μ_τ and μ_Y under Level-k Thinking (left) and anchored beliefs (right).

The equivalence, however, breaks down for any even number k because Level-k Thinking displays a peculiar, "oscillatory" behavior in games of strategic substitutability. In our context, this problem emerges with target communication, precisely because this induces a game of strategic substitutability.

Let us explain. For any given announcement, an agent wants to invest more when he expects others to investment less. Because the level-0 agent is assumed to be completely unresponsive, a level-1 agent expects K to move *less* than in the frictionless benchmark and thus moves *more* himself. A level-2 agent then expects K to move *more* than in the frictionless benchmark and therefore chooses to move *less* himself. That is, whereas $k = 0$ amplifies the actual response of investment relative to rational expectations, $k = 1$ attenuates it. The left panel of Figure 1 shows that this oscillatory pattern continues for higher k, and that this oscillation with target communication is the only qualitative difference between the present specification and that studied in Section 4.

We are not aware of any experimental evidence of this oscillatory pattern. We suspect that it is an unintended "bug" of a solution concept that was originally developed and tested in the experimental literature primarily for games of complements and may not be applicable to games of substitutes without appropriate modification. Seen from this perspective, the formalization adopted in the previous section captures the essence of Level-k Thinking while bypassing its "pathological" feature.

The same goal can be achieved with a "smooth" version of Level-k Thinking along the lines of Garcıa-Schmidt and Woodford (2019). The concept of "cognitive discounting" introduced in Gabaix (2018) works in a similar manner, too, because it directly postulates that the subjective expectations of endogenous variables such as K move less than the rational expectations of it. Last but not least, incomplete information as in Angeletos and Lian (2018) generates the same friction in higher-order beliefs but also adds another friction in the form of inattention or irresponsive first-order beliefs. The role of this additional friction is studied in Subsection D.3.

D Other parameter cases

D.1 Negative GE feedback ($\gamma < 0$)

The entire analysis has presumed a positive GE feedback ($\gamma > 0$). We now briefly discuss the case with a negative GE feedback ($\gamma < 0$). In this case, K depends negative on expectations of Y. This may capture situations in which agents compete for finite resources, with higher output corresponding to higher prices and hence lower consumption or investment (see the micro-foundation of Section B.2 for an example). Both modes of communication now induce a game of strategic substitutes. In particular, the game of substitutes is more "severe" under target communication, or $\delta_Y < \delta_\tau < 0$.

How does translate to the optimal communication policy? Consider first the anchored beliefs model. If we make parameter assumptions to rule out the case $\mu_Y < 0$, which involves policy moving in the opposite direction of output, it is easy to show in the anchored beliefs model that instrument communication is strictly preferred to target communication for any $\gamma < 0$. To achieve the same result more generally, we need further assumptions on the loss function. The following Theorem elaborates on the technical details:

Theorem 2. *For any $\lambda < 1$, there exists some threshold $\dot{\gamma} < 0$ such that instrument communication is strictly preferred for $\gamma \in [\dot{\gamma}, 0]$. Further, if $\mu_Y > 0$ (as per the conditions of Lemma 5) or $\chi < 1/2$, $\dot{\gamma} = -\infty$.*

Proof. First, maintain Lemma 5 and its assumptions. Note that the second case ("more general") of the proof of the previous section does not use $\gamma > 0$. Hence the result is proved for $\dot{\gamma} = -\infty$ in this case.

Now relax those assumptions. Our best bound on the loss with target communication, for $\mu_Y < 0$, is $\min\{\mathscr{L}(\mu_{y,-\infty}), 1-\chi\}$, or the minimum loss between the $\gamma \to -\infty$ limit and the $\mu = 0$ extreme. $\mathscr{L}_\tau(\gamma)$ decreases smoothly on $\gamma \in (-\infty, 0]$ and is bounded above by $\mathscr{L}(0) = 1-\chi$. If $\mathscr{L}(\mu_{y,-\infty}) > 1-\chi$, it follows that $\underline{\gamma} = -\infty$ again. Since $\mathscr{L}(\mu_{y,-\infty}) > \lim_{\mu \to -\infty} \mathscr{L}(\mu) = \chi$, it follows that sufficient condition is $\chi > 1-\chi$ or $\chi > 1/2$.

Otherwise there must exist some $\dot{\gamma} < 0$ above which $\mathscr{L}_\tau(\gamma) < \chi$ and below which $\mathscr{L}_y(\gamma) > \chi$. We know for sure that instrument communication is optimal for $\gamma > \dot{\gamma}$ and target communication is optimal for $\gamma \in (-\infty, \dot{\gamma})$. □

In the model with erratic beliefs, we can similarly rank the size of the "wedges" in the implementability constraint

Proposition 12. *For any values of $\alpha \in (0,1)$, $\sigma \in [0,1)$, and $\gamma \le 0$, $\psi_Y > \psi_\tau > 0$.*

Proof. It is obvious from the expressions why the values are positive. To see their relative size, note that $\psi_\tau = -\alpha g(\delta_\tau)$ and $\psi_Y = -\alpha g(\delta_Y)/(1-\alpha)$ for $g(\delta) \equiv \delta(1-\sigma)/(1-\sigma\delta)$. Note that $g(\delta)$ is non-positive and increasing for $\delta < 0$, and $\delta_Y < \delta_\tau \le 0$ for $\gamma \le 0$. Thus $\delta_\tau = -\alpha g(\delta_\tau) < -\alpha g(\delta_Y) < -\alpha g(\delta_Y)/(1-\alpha) = \psi_Y$. □

For optimal policy, however, the policymaker's relative preference for *where* this wedge goes (in the instrument or outcome gap) will always matter. More specifically, in contrast to the anchored beliefs model, there is no tool to shift the distortion between gaps (setting r). Thus, even though $\psi_Y^2 > \psi_\tau^2$ unambiguously for all $\gamma < 0$, there exists a large enough weight on the output gap (χ) such that target communication is still preferred. Of course if the weights are equal or lower on the output gap ($\chi \leq 1/2$), instrument communication will be strictly preferred.

D.2 Extreme substitutability ($\delta_X < -1$)

Most of our analysis restricts $\alpha < \frac{1}{2-\gamma}$ so as to guarantee that $-1 < \delta_X < 1$ for both modes of communication. This allows the characterization of beliefs and behavior by repeated iteration of the best responses. In particular, in Section 3 it guarantees that the joint of Assumptions 1 and 2 replicates the REE benchmark; in Appendix C, it guarantees that the Level-k outcome converges to the REE outcome as agents become "infinitely rational" ($k \to \infty$); and in Sections 4 and 6.1, it guarantees that Assumptions 3 and 4 yield the corresponding PBE outcomes.

When the aforementioned restriction is violated, our main lessons continue to apply as long as one focuses directly on the relevant REE and PBE outcomes. For instance, take the case studied in Section 4 and recast it in terms of heterogeneous priors. Except for the degenerate case in which $\alpha = \frac{1}{2-\gamma}$, or $\delta_Y = -1$, there exists a unique linear PBE and it is such that all the results of that section apply regardless of whether $\alpha > \frac{1}{2-\gamma}$ or $\alpha < \frac{1}{2-\gamma}$. What is lost is only the "global stability" of this outcome, in the sense that the fixed point is no more obtainable as the limit of iterated best responses.

D.3 First-order vs higher-order mistakes

Our analysis has allowed imperfect reasoning about the behavior of other agents or policy's GE effects ("higher-order mistakes") but ruled out any friction in agents' own awareness of the policy or policy's PE effects ("first-order mistakes"). The latter kind of mistakes can be obtained—either in isolation from or in combination with the former kind—due to rational inattention (Sims, 2003), sparsity Gabaix (2014), or noisy/sticky information (Lucas, 1972; Mankiw and Reis, 2002). How would these frictions affect our results?

Let us assume that first-order beliefs are possibly attenuated, or satisfy $\bar{\mathbb{E}}[X] = qX$ for some $q \in [0,1]$. Maintain that higher-order beliefs have the familiar structure from Section 4, or $\bar{\mathbb{E}}^h[X] = \lambda^{h-1}\bar{\mathbb{E}}[X]$ for some $\lambda \in (0,1]$ and all $h \geq 2$. Following similar arguments as in our main analysis, it can be shown that the slopes of the implementability constraints under instrument and target communication are now given by, respectively,

$$\mu_\tau = \frac{1}{1-\alpha+\frac{1-\alpha\gamma}{1-\alpha\gamma\lambda}\alpha q} \qquad \text{and} \qquad \mu_Y = \frac{1-\alpha+\alpha(1-\gamma)\lambda-\alpha(1-\alpha\gamma)q}{(1-\alpha)\left(1-\alpha+\alpha(1-\gamma)\lambda\right)}. \tag{28}$$

Instrument communication necessarily produces attenuation, or $\mu_\tau > 1$, because both frictions ($q < 1$ and $\lambda < 1$) work in the same direction. By contrast, the case for target communication is ambiguous ($\mu_Y \lesseqgtr 1$), because the amplification induced by anchored higher-order beliefs ($\lambda < 1$) opposes the attenuation induced by inattention ($q < 1$). Which effect dominates depends on the belief parameters (q, λ) *and* the GE feedback γ, because the last interacts with anchored beliefs as explained in our main analysis.[24]

Of particular interest is the case $q = \lambda$, which is isomorphic to a rational expectations model with a Gaussian prior and Gaussian private signals.[25] As it turns out, this special case induces *equal* attenuation under both strategies, or $\mu_\tau = \mu_Y > 1$, and therefore replicates the irrelevance result of the frictionless benchmark.

A generalization of Theorem 1 is instead obtained if and only if $\lambda < q$. This can be proven by verifying that the expressions in (28) remain monotone in γ and always satisfy $\mu_\tau > \mu_Y$.

Counterexample and intuition. When $q < \lambda$ and $q \neq 1$, we can get different optimal policy results. Note that $q \neq 1$ is important; for the same model with $q = 1$, nested in the previous subsection, our result goes through. This underscores the fact that completely arbitrary deviations from rational expectations do not preserve our main intuitions.

What is going on economically? Whenever $q \neq 1$, there are two distinct frictions in the economy. The first relates to *first-order mistakes* which uniformly attenuate (or amplify) the response of the economy to communication. The second relates to *higher-order mistakes* which matter more when "equilibrium reasoning" is more intense. The intuition about "telling the public what it needs to know" translates purely into optimal policy when we shut down the first category of mistakes (e.g., by setting $q = 1$). The case $\lambda < q < 1$, somewhat conveniently, "stacks" the two distortions without changing the comparative static; other cases allow them to fight one another, and the policymaker may *prefer* distorted reasoning that cancels out with a distorted hearing of the announcement.

The important question, then, is which of these cases is economically more relevant. We prioritize $\lambda < q$, and its extreme cousin $\lambda < q = 1$, because we think that imperfect computation of equilibrium is the main "primitive friction" that anchors beliefs, especially in very unfamiliar economic circumstances like a liquidity trap. Appendix C works out a formal connection between this case and a model of level-k thinking, with the caveat that the pure level-k model (without any effort to "smooth" predictions) makes erratic predictions for games of strategic substitutes.

In the case $q < \lambda < 1$, agents uniformly under-react to the announcement but are perfectly coordinated with *each other* in doing so. This could be relevant in regimes with low credibility for the policymaker (e.g., because of a history of "over-promising" via forward guidance) but a highly sophisticated and coordinated public. In such cases, there is a well-defined instruments-versus-targets trade-off, and the *single* friction we highlight about the distortion from imperfect computation of equilibrium remains. But, because there are two frictions that *interactively* induce dependence on the GE feedback γ, there is no simple threshold comparative static that generalizes Theorem 1.

[24] Indeed, attenuation is obtained with target communication (i..e., $\mu_Y > 1$) if and only if $q < \tilde{q}(\lambda, \gamma) \equiv \frac{1-\alpha(1-(1-\gamma)\lambda)}{1-\alpha\gamma}$. The threshold $\tilde{q}$ is increasing in both λ and γ, always exceeds λ, and reaches 1 when either $\lambda = 1$ or $\gamma = 1$.

[25] To see this, let $X \sim \mathcal{N}(0, \sigma_X^2)$, let each agent i observe a private signal $s_i = X + \varepsilon_i$, where $\varepsilon_i \sim \mathcal{N}(0, \sigma_s^2)$, and let these facts as well the agents' rationality be common knowledge. Then, the previously described structure of first- and higher-order beliefs holds with $q = \lambda = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_s^2}$.

E Adding More Shocks

Our baseline model included exogenous shocks to the preferences of the policymaker but excluded such shocks from conditions (1) and (2). This is without loss of generality if the other shocks are common knowledge and observed by the policymaker. These assumptions are extreme, but common in the Ramsey policy paradigm. In our context, they guarantee that implementability results remain true provided that that the quantities (τ, Y) are re-defined to be "partialed out" from the extra shocks.

A more plausible scenario, perhaps, is that other shocks are unobserved and the policymaker cannot condition on them. This introduces into our analysis similar considerations as those in Poole (1970). The latter focused on how two different policies—fixing the interest rate or fixing the money supply—differed in their robustness to external shocks. Primitive shocks (to supply and demand) had different effects on the policy objective (output gap) depending on the slope of the model equations and the policy choice. Poole could do comparative statics of optimal policy in these slopes as well as the relative variance of the shocks.

Such "Poole considerations" can be inserted into our framework and will naturally affect the choice between fixing τ and fixing Y. However, such consideration matter even in the REE benchmark and, roughly speaking, are "separable" from the mechanism we have identified in our paper. We make this point clearer with a few examples in the sequel.

E.1 Shocks to output

Consider now a model in which output contains a random component:

$$Y = (1-\alpha)\tau + \alpha K + u,$$

where u is drawn from a Normal distribution with mean 0 and variance σ_u^2, is orthogonal to θ, and is unobserved by both the policymaker and the private agents. In this case, announcing and committing to a value for Y stabilizes output at the expense of letting the tax distortion fluctuate with u. Conversely, announcing and committing to a value for τ stabilizes the tax distortion at the expense of letting output fluctuate with u. It follows that, even in the frictionless benchmark ($\lambda = 1$), the policymaker is no more indifferent between the two. In particular, target communication is preferable if and only if the welfare cost of the fluctuations in Y exceeds that of the fluctuations in τ, which is in turn is the case whenever χ is high enough.

The above scenario has maintained the assumption that the ideal level of output is $Y^{\text{fb}} = \theta$. What if instead we let $Y^{\text{fb}} = \theta + u$? This could correspond to a micro-founded business-cycle model in which technology shocks that have symmetric effects on equilibrium and first-best allocations. Under this scenario, it becomes desirable to let output fluctuate with u, which in turn implies that, in the frictionless benchmark, instrument communication always dominates target communication. A non-trivial trade off between the two could then be recovered by adding unobserved shocks to the tax distortion. The optimal strategy is then determined by the relative variance of the two unobserved shocks and the relative importance of the resulting fluctuations, along the lines of Poole (1970).

While these possibilities are interesting on their own right, they are orthogonal to the message of our paper. Indeed, the shock considered above does not affect the strategic interaction of the private agents under either mode of communication: Lemmas 1 and 2 remain intact. By the same token, when $\lambda = 1$, the sets of the implementable (τ, Y) pairs remain invariant to γ, even though they now depend on the realization of u. It then also follows that, as long as $\lambda = 1$, the optimal mode of communication does not depend on γ. But as soon as $\lambda < 1$, the implementability sets and the optimal mode of communication start depending on γ, for exactly the same reasons as those explained before: a higher γ increases the bite of strategic uncertainty under instrument communication and decreases it under target communication, thus also tilting the balance in favor of the latter as soon as one departs from the frictionless benchmark.

E.2 Measurement errors and trembles

The same logic as above applies if we introduce measurement errors in the policymaker's observation of τ and Y, or equivalently trembles in her control of these objects. To see this, consider a variant of our framework that lets the policymaker control either $\tilde{\tau}$ or $\tilde{Y}$, where

$$\tilde{\tau} = \tau + u_\tau, \qquad \tilde{Y} = Y + u_Y,$$

and the u's are independent Gaussian shocks, orthogonal to θ, and unpredictable by both the policymaker and the private agents. Instrument communication now amounts to announcing and committing to a value for $\tilde{\tau}$, whereas target communication amounts to announcing and committing to a value for $\tilde{Y}$.

By combining the above with condition (1), we infer that, under both communication modes, the following restriction has to hold:

$$\tilde{Y} = (1-\alpha)\tilde{\tau} + \alpha K + \tilde{u},$$

where

$$\tilde{u} \equiv -(1-\alpha)u_\tau + u_Y.$$

At the same time, because the u's are unpredictable, the best response of the agents can be restated as

$$k_i = (1-\gamma)\mathbb{E}_i[\tilde{\tau}] + \gamma \mathbb{E}_i[\tilde{Y}].$$

This maps directly to the version with unobserved shocks just discussed above if we simply reinterpret $\tilde{\tau}$, $\tilde{Y}$, and $\tilde{u}$ as, respectively, the actual tax rate, the actual level of output, and the unobserved output shock.

To sum up, the presence of unobserved shocks and measurement error can tilt the optimal strategy of the policymaker one way or another in manners already studied in the literature that has followed the lead of Poole (1970). This, however, does not interfere with the essence of our paper's main message regarding the choice of a communication strategy as a means for regulating the impact of strategic uncertainty and the bite of the considered forms of bounded rationality.

F Communicating other objects

Our initial focus on communicating τ or Y seemed natural for applications. But, for completeness, we should also check whether it would be wiser either to communicate directly the realized value of θ, or to commit to a target for the aggregate action K.

F.1 Communicating the value of θ

Consider the first scenario. In this scenario, the policymaker is picking, and committing on, a mapping from θ to τ or Y, but does not tell this mapping to the agents. Instead, she only tells them what θ is. In other words, the policymaker tells the agents what he would like to achieve, but not the way she is going after it.

As already noted, such communication implements the first best under rational expectations. Because REE imposes a unique mapping from θ to both τ and Y, and the agents know that mapping, there is no need for the policymaker to communicate it. Away from that benchmark, however, many such mappings can be part of an equilibrium and, as a result, communicating merely θ does not necessarily pin down the agents' beliefs about either the policy or the outcome. In particular, there exists an equilibrium that replicates instrument communication, as well as an equilibrium that replicates target communication.

F.2 Communicating a target for K

Consider next the scenario in which the policymaker communicates a target for K. This option may be impractical if K stands for a complex set of decisions that is hard to measure. But even abstracting from such measurement issues, this option may not be viable—or at least it is not well-posed in our model.

Consider in particular the specification studied in Section 4 and let the policymaker announce and commit to a value $\hat{K}$ for aggregate investment. Assume that first-order beliefs about investment are correct ($\bar{\mathbb{E}}[K] = \hat{K}$) and higher-order beliefs are anchored toward zero ($\bar{\mathbb{E}}^h[K] = \lambda^{h-1}\hat{K}$). For the announcement to be fulfilled in equilibrium, it must be the case that

$$\hat{K} = (1-\delta_X)\bar{\mathbb{E}}[X] + \delta_X \bar{\mathbb{E}}[K] = (1-\delta_X)\bar{\mathbb{E}}[X] + \delta_X \hat{K}$$

for either fundamental $X \in \{\tau, Y\}$. The only first-order beliefs compatible with this announcement, then, are $\bar{\mathbb{E}}[\tau] = \bar{\mathbb{E}}[Y] = \bar{\mathbb{E}}[K] = \hat{K}$: on average (and, in fact, uniformly), agents believe that equilibrium will be $\tau = Y = K$. This is an ideal scenario for the policymaker.

It turns out, however, that a rational agent who doubts the attentiveness of others will doubt that other agents play the announcement, or that $K = \hat{K}$. If a given agent i thinks that agent j plays $k_j = \hat{K}$, she is implicitly taking a stand on agent j's beliefs about τ and Y. Specifically, agent i believes that agent j is following her best response (here, written with $X = \tau$), namely

$$\mathbb{E}_i[k_j] = (1-\delta_\tau)\mathbb{E}_i\mathbb{E}_j[\tau] + \delta_\tau \mathbb{E}_i\mathbb{E}_j[K]$$

We have assumed that $\mathbb{E}_i[k_j] = \hat{K}$ and $\mathbb{E}_i\mathbb{E}_j[K] = \lambda\hat{K}$. This produces the following restriction on second-order beliefs about τ:

$$\mathbb{E}_i\mathbb{E}_j[\tau] = \frac{1-\lambda\delta_\tau}{1-\delta_\tau}\hat{K}.$$

This has a simple interpretation: to rationalize aggregate investment being $\hat{K}$ despite the fact that fraction $(1-\lambda)$ of agents were inattentive to the announcement, agent i thinks that a typical other agent has *over*-forecasted the policy instrument τ.

At the same time, agent i knows that, like himself, all attentive agents expect τ to coincide with $\hat{K}$. And since agent i believes that the fraction of attentive agents is λ, the following restriction of second-order beliefs also has to hold:

$$\mathbb{E}_i\mathbb{E}_j[\tau] = \lambda\hat{K}.$$

When $\lambda = 1$ (rational expectations), the above two restrictions are jointly satisfied for any $\hat{K}$. When instead $\lambda < 1$, this is true only for $\hat{K} = 0$. This proves the claim made in the text that, as long as $\lambda < 1$, there is no equilibrium in which is infeasible to announce and commit to any $\hat{K}$ other than 0 (the default point).

In a nutshell, the problem with communicating K is that the policymaker has no direct control over it. From this perspective, output communication worked precisely because the policymaker had some plausible commitment. Agents could rationalize $Y = \hat{Y}$ regardless of their beliefs about K because there always existed some level of τ that implemented $\hat{Y}$. We alluded to the failure of this mechanism as $\alpha \to 1$, and the direct effect of policy vanished, in our baseline model (Section 4.6).

We could bypass this issue, of course, by giving the policymaker an instrument z that directly affects investment decisions; this amounts to replacing the best response with $k_i = (1-\alpha)\mathbb{E}_i[\tau] + \alpha\mathbb{E}_i[Y] + z$. But this could bypass the issue of interest: instead of trying to influence K by manipulating the expectations of τ and Y, the policymaker could just use z to directly control K regardless of these expectations. It is the absence of such an instrument that justifies the focus on "managing expectations" as a relevant policy tool. In the context of the liquidity trap, for example, the absence of z reflects a binding ZLB on the current interest rate (along with the usual unavailability or ineffectiveness of consumption taxes or other fiscal-policy substitutes).

G Linear policy rules

Throughout this paper, we have not directly addressed the issue of credible commitment. The previous discussion highlights that our analysis may have subtle interactions with commitment problems. Indeed, agents' (higher-order) beliefs about commitment problems may be crucial. We leave the formal investigation of this topic to future work.

The choice between instrument and target communication remains a choice of "extremes." One could imagine a more sophisticated strategy in which the policy maker announces and commits to a policy rule of the following type:

$$\tau = A - BY \tag{29}$$

where (A, B) are free parameters. In the context of monetary policy, of course, this expression is a familiar Taylor rule.

Instrument communication can then be nested with $B = 0$ and $A = \hat{\tau}$, for arbitrary $\hat{\tau}$; and target communication can be though as the limit in which $B \to \infty$ and $A/B \to \hat{Y}$, for arbitrary $\hat{Y}$. Away from these two extremes, the policymaker's strategy is indexed by the pair (A, B) and policy communication amounts to the announcement of this pair, as opposed to a fixed value for either τ or Y.

For reasons outside our model, such feedback rules may be hard for the agents to comprehend and may therefore be less effective than the two extremes considered so far. We suspect that, in many real-world situations, there is a gain in conveying a sharp policy message of the form "we will keep interest rates at zero for 8 quarters" or "we will do whatever it takes to bring unemployment down to 4%," as opposed to communicating a complicated feedback rule. This explains why we a priori found it more interesting to focus on the two extremes.

Having said that, it is useful to explore how such policy rules work within our model. The key insights survive and, in fact, their scope expands: once one deviates from rational expectations, such policy rules play a function not previously identified in the literature and akin to that identified in the preceding analysis.

Consider first the rational expectations benchmark (as in Section 3). In this benchmark, the additional flexibility afforded by this class of policy rules is entirely useless, because the first best was already attained by the two extremes. Furthermore, our earlier irrelevance result directly extends: not only for the first best, but also for any other point in $\mathscr{A}^*$, there exist a continuum of values for (A, B) that implement it as part of an REE. The only subtlety worth mentioning is that such an REE may fail to be the unique equilibrium if $B < -1$. The logic is similar to the one underlying the Taylor principle.

To understand these properties, solve (29) and (1) jointly for τ and Y and substitute the solution into (2) to obtain the following game representation:

$$k_i = \zeta(A, B; \alpha, \gamma) + \delta(B; \alpha, \gamma)\mathbb{E}_i[K] \tag{30}$$

where

$$\zeta(A, B; \alpha, \gamma) \equiv \frac{(1-\alpha\gamma)A}{1+(1-\alpha)B} \qquad \text{and} \qquad \delta(B; \alpha, \gamma) \equiv \frac{\alpha(\gamma - B(1-\gamma))}{1+(1-\alpha)B}.$$

It is then evident that B controls the slope of the best responses and A their intercept. When $B < -1$, the policy induces a game of strategic complementarity in which the slope exceeds 1, opening the door to multiple equilibria. When instead $B \in (-1, \frac{\gamma}{1-\gamma})$, the slope is positive but less than one. And when $B > \frac{\gamma}{1-\gamma}$, the slope becomes negative, which means that the policy rule induces a game of strategic complementarity. Finally, it is clear that, for any value of K, there exist a continuum of (A, B) that induces this K as the fixed point of (30).

Consider now the case with anchored beliefs (as in Section 4). The extra flexibility afforded by the policy rules now becomes relevant: by varying A and B, the planner can induce a wide range of outcomes beyond those contained in $\mathscr{A}_\tau$ and $\mathscr{A}_Y$. What is more, there actually exist a subclass of policy rule that replicates $\mathscr{A}^*$, namely the set of outcomes that are attained under rational expectations. This subclass is given by setting B such that $\delta(B;\alpha,\gamma)=0$, or equivalently $B=\frac{\gamma}{1-\gamma}$, and letting A vary in $\mathbb{R}$. Intuitively, setting B so that $\delta(B;\alpha,\gamma)=0$ completely eliminates the need for the agents to forecast, or calculate, the behavior of others, which in turn guarantees that the distortion on the set of implementable vanishes regardless of λ. By varying A, the policymaker can then span the set $\mathscr{A}^*$. And by picking A so that $\zeta(A,B;\alpha,\gamma)=\theta$, she can implement the first best.[26]

We summarize these lessons in the following result.

Proposition 13. *Suppose that the policymaker can announce and commit on a policy rule as in (29) and let Assumptions 1 and 3 hold with* $X=(A,B)$.

When $\lambda=1$ *(rational expectations), the first best is implemented with any* (A,B) *such that* $B>-1$ *and* $A=(1+B)\theta$.

When instead $\lambda<1$ *(anchored beliefs), the first best is implemented if and only*

$$B=\frac{\gamma}{1-\gamma}\qquad \textit{and}\qquad A=\frac{\theta}{1-\gamma}.$$

At first glance, this result may appear to dilute our take-home message: a more sophisticated strategy than the ones studied in the main body of our paper completely eliminates the problem. However, this property is fragile in the following sense. When the policymaker is uncertain about the structure of the economy, in particular about the values of γ, the values of B and A obtained above are also uncertain. The first best is therefore unattainable when $\lambda<1$, even though it remains attainable under rational expectations.

Most importantly, our take-home message survives in the following two keys senses. First, the optimal strategy is indeterminate under rational expectations ($\lambda=1$), whereas it is determinate with anchored beliefs ($\lambda<1$). And second, for any $\lambda<1$, a stronger GE effects calls for a policy rule that has a steeper slope with respect to Y and, in this sense, looks closer to target communication. In fact, in the limit as $\gamma\to 1$, the optimal policy rule has $B\to-\infty$ and $B/A\to\theta$, which is the same as the target communication with $\hat{Y}=\theta$.

We thus interpret Proposition 13 as a complement to our main analysis, not a sign that the choice between instrument and target communication was too narrowly framed. Proposition 13 also offers a new perspective on Taylor rules. The pertinent literature has focused on two functions: how the slope of the Taylor rule can induce a unique equilibrium; and how it must be designed if the policymaker cannot directly condition the intercept of the Taylor rule on the underlying fundamentals. The first issue maps to our discussion above about setting $B>-1$ as is know as the Taylor principle. The second issue is a modern variant of Poole (1970). Our own result brings up a completely different function: the role of such rules in regulating the distortionary effects of bounded rationality.

This function extends to common-prior settings that maintain rational expectations but allow for higher-order uncertainty. This is because policy rules that regulate the agents' strategic interaction also regulate the impact that any "belief wedge" (any gap between first- and higher-order beliefs) has on actual

[26] Clearly, this logic extends to the variants with Level-k Thinking and erratic beliefs.

outcomes regardless of whether this wedge represents a departure from rational expectations or a rich enough informational friction. We view this point as another facet of the insights developed in the earlier sections of our paper.