Managing Expectations:
Instruments vs Targets[*]

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

July 4, 2019

Abstract

Should a policymaker manage expectations by offering forward guidance in terms of the likely value of a future policy instrument or a target for an equilibrium outcome such as aggregate output? We study how the optimal approach depends on plausible bounds on agents' depth of knowledge and rationality. Agents make mistakes in predicting, or reasoning about, the behavior of others and the GE effects of policy. The optimal communication strategy minimizes the bite of such mistakes on implementability and welfare. This goal is achieved by offering guidance in terms of an outcome target rather than a policy value 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 steep Keynesian cross or a prolonged liquidity trap.

Monetary policy is 98 percent talk and only two percent action.

Bernanke, 2015

1 Introduction

Forward guidance, the art of managing expectations, is rarely comprehensive. For example, even if the central bank can shape expectations about future interest rates, it remains up to the market to predict the consequences for GDP or unemployment. Under what circumstances, we ask, is it better to do the opposite, promising to do "whatever it takes" to achieve a target for the outcome of interest and leaving the market to ponder what policy will support this target?

We study how the answer to this question depends on the possibility that agents make mistakes in reasoning about the behavior of others and the equilibrium effects of the policy. Such mistakes are assumed away in the textbook policy paradigm via strong assumptions about agents' depth of knowledge and rationality. We allow such mistakes to exist and shed light on how the policymaker can mitigate them via the choice and communication of an appropriate policy plan.

We work with an abstract framework, bypassing the micro-foundations of specific applications and focusing on the key concepts. Our main result is a sharp dependence of the optimal strategy on "GE considerations," or the feedback between aggregate outcomes and individual behavior. Offering guidance in the form of a target for the outcome instead of a value for the instrument is optimal if and only if this feedback is sufficiently high, as in situations with a strong aggregate demand externality, a steep Keynesian cross, or a prolonged liquidity trap.

Framework. The following example, nested in our abstract framework, helps fix ideas. The economy is in a liquidity trap and output (the targeted outcome) is demand-determined, forming an analogue of the textbook Keynesian cross (GE feedback). Consumer demand depends on expected income and interest rates (the policy instrument). A policymaker can commit either to keeping interest rates low or taking any necessary action to boost future output. An alternative, purely Neoclassical example recasts the consumers as capital investors, the instrument as a subsidy, and the GE feedback as an aggregate demand externality of the Dixit-Stiglitz variety.

Unlike in Morris and Shin (2002) and a large follow-up literature,[1] policy communications in our setting contain no relevant information about the exogenous payoffs of the agents. They only state the policymaker's chosen plan of action. Communication is nevertheless necessary because agents do not *a priori* know what the policymaker plans to do—they need forward guidance.

The policymaker chooses between two forms of such guidance. In the first, she announces, and commits to, a value for the policy instrument (the interest rate or subsidy). In the second, she does the same with a target for the relevant outcome (aggregate output). We refer to the former strategy as *instrument communication* and to the latter one as *target communication*.

---

[1] See, *inter alia*, Amador and Weill (2010), Angeletos and Pavan (2007), Chahrour (2014), Cornand and Heinemann (2008, 2015), James and Lawler (2011), Morris and Shin (2007), Myatt and Wallace (2012), and Svensson (2006).

This approach equates each form of communication with a commitment to a policy plan. But whereas the literature has been concerned primarily with robustness to fundamental uncertainty (Poole, 1970) and commitment problems (Atkeson, Chari and Kehoe, 2007), our analysis shifts the focus to how agents reason about the behavior of others and the GE effects of policy.

A rational expectations benchmark and beyond.   Like the textbook policy paradigm, our benchmark assumes a representative, rational-expectations agent and predicts that the form of forward guidance is irrelevant: the implementable combinations of policy and outcome are invariant to the policymaker's choice between the aforementioned two strategies. This irrelevance depends, not only on the assumption that the typical agent is *herself* rational and aware of the policy communication, but also on the assumption that such rationality and awareness is common knowledge ("I know that you know...").

We focus on relaxing of the second, stronger assumption. 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; the question is whether and how the policymaker can work around it.

Main results.   Our main, preferred specification allows agents to doubt the awareness or the attentiveness of others. This amounts to removing common knowledge of the announced policy plan while preserving every agent's own knowledge of it. The upshot is that agents form anchored beliefs about the actions of others, similarly to the literature that studies the role of higher-order uncertainty (e.g., Abreu and Brunnermeier, 2003; Morris and Shin, 1998, 2002; Woodford, 2003). A related specification that has agents face difficulty in strategic reasoning, based on Level-k Thinking (Nagel, 1995; Stahl, 1993), delivers essentially the same results

Our take-home lesson is that, in the presence of the aforementioned friction, offering guidance in terms of targets rather than instruments is preferable when and only when the GE feedback is sufficiently strong. This lesson builds on two intermediate results.

The first regards the implementability constraint faced by the policymaker, namely the equilibrium relation between the instrument and the outcome. This relation is invariant to the form of forward guidance in our rational-expectations benchmark but not away from it.

With instrument communication, the agents play a game of strategic *complementarity* and the belief friction produces *attenuation*: when an agent expects the others to invest or spend less in response to the announcement, she responds less herself. As a result, a larger change in interest rates or subsidies is needed in order to induce the same change in output.

With target communication, everything flips. Conditional on an announced GDP target, a household that expects higher aggregate spending also expects a higher interest rate, which reduces the incentive to consume; similarly, a firm that expects a higher aggregate investment also expects a lower required subsidy to that target, which reduces the incentive to invest. Agents now play a game of strategic *substitutability*, in which the same belief friction produces *amplification*. As a result, the implementability constraint is "flattened."

Our second result relates to the interaction between the form of forward guidance and the un-

derlying 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 or subsidies; when it is strong, they care more about aggregate demand.

These observations, together, imply that instrument communication minimizes the bite of the distortion if and only if GE feedback is sufficiently small. Since we focus on bounded rationality as the only source of distortion, this immediately implies the optimal policy result.

ZLB application.   To illustrate how our results might translate into concrete policy recommendations, we return to our example of forward guidance about future monetary policy during a liquidity trap. A recent literature has studied this issue focusing only on what this paper has called instrument communication.[2] Our result that anchored beliefs attenuate the effectiveness of instrument communication is essentially the same as the results obtained in that literature.

We add two new insights in this context. First, we show how the policymaker can flip this friction by engaging in forward guidance about targets. Second, we provide specific advice about when to do so. In conventional models a liquidity trap switches on powerful positive feedback loops between income, spending, and inflation. In the language of our abstract model, this corresponds to a high GE feedback, which tilts the balance toward target communication. Succinctly, a ZLB recession is the worst time to chatter about interest rates and the best time to talk about stabilizing unemployment.

Robustness.   Although our main analysis imposes that beliefs under-react, our policy lesson extends to situations in which beliefs over-react or are subject to arbitrary "noise." In particular, while the exact distortion of the implementability constraints under the two modes of communication depends on the exact specification of the belief friction, the intuition that "optimal policy minimizes the role for distorted beliefs" does not. By the same token, a policymaker who suspects beliefs are not "fully" rational but is not sure of the particular mode of mis-specification would still find concrete guidance from our analysis.

Would our simple policy lesson be overturned in a more complex model? We verify that our trade-off is orthogonal to adding additional unobserved shocks, as in Poole (1970)'s classic IS/LM model or, presumably, something that resembles a modern monetary DSGE model. We show also, relevantly for the latter context, that our consideration can guide the design of optimal policy rules.

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).

---

[2]Angeletos and Lian (2018) and Wiederholt (2016) attribute the friction to lack of common knowledge; García-Schmidt and Woodford (2019) and Farhi and Werning (2019) to Level-k Thinking; Gabaix (2018) to "cognitive discounting." Iovino and Sergeyev (2017) offer a related application to quantitative easing.

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 of higher-order beliefs or bounded rationality, via commitment to a policy plan. 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.[3]

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. Related is also Bergemann and Morris (2016), who study the robustness of a mechanism to the designer's uncertainty about the players' information. Our exercise, instead, represents a form of robustness to the players' bounded rationality.

The relaxation of rational expectations separates our paper more broadly from a large literature in macroeconomics and finance that studies the role of incomplete information and higher-order uncertainty *without* such a relaxation. As explained in Subsection 7.3, this allows us to decouple the friction in higher-order beliefs (which, for our purposes, is synonymous to the imperfection in the agents' reasoning about the GE effects of policy) from inattention or any other friction in first-order beliefs. Such decoupling is not only consistent with our paper's motivation, but also the key to its results. At the same time, the emphasis on higher-order beliefs and the specific policy insights thus delivered distinguish our contribution from a long tradition that studies other aspects of relaxing rational expectations in macroeconomics.[4]

Layout. Section 2 introduces our abstract framework, two micro-foundations for which are given in Section 3. Section 4 studies our rational-expectations benchmark and lays down the foundations of

---

[3]The same basic points also distinguish our paper from the literature on Bayesian persuasion and information design (Bergemann and Morris, 2013, 2018; Kamenica and Gentzkow, 2011; Inostroza and Pavan, 2018).

[4]E.g., Sargent (1993); Evans and Honkapohja (2001); Hansen and Sargent (2007); Woodford (2013)

the subsequent analysis. Section 5 contains our main specification, anchored higher-order beliefs. Section 6 translates our abstract results to the ZLB context. Section 7 extends the analysis to other belief frictions. Section 8 discusses our simple trade-off fits into a more complicated landscape of distortions and policy options. Section 9 concludes.

2 Framework

In this section we introduce the physical environment, the incentives of the private agents, the objective of the policymaker, and the timing of actions. We postpone, however, the specification of how agents form expectations (i.e., how they reason about the behavior of others) until later.

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 $1 - \alpha$

5

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.

The next section will outline two complementary micro-foundations for the simple model, one New Keynesian (related to monetary policy in a liquidity trap) and one Neoclassical (related to investment with aggregate demand externalities). Each will demonstrate a more precise mapping from the "reduced form parameters" (γ, α) to more familiar "deep parameters."

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

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

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.

3   Parenthesis: Two Micro-foundations

In this section we show how our abstract framework can nest a New Keynesian model in which τ represents interest rate set after the economy exits a liquidity trap, and a Neoclassical model in which τ represents a tax instrument. A reader more curious about the main results, which involve only the abstract framework, might skip this section and return to it in the context of policy applications.

3.1   Monetary policy in a liquidity trap

Consider a New Keynesian economy with perfectly rigid prices. There are countably infinite periods, $t \in \{1, 2, \ldots\}$, and a unit measure of consumers, $i \in \{1, 2, \ldots\}$.

Consumer i has the following utility function:

$$\mathcal{U}_i = \mathbb{E}_i \left[\sum_{t=1}^{\infty} \left\{ \exp\left(-\sum_{j=1}^{t} \rho_j \right) \log C_{i,t} \right\} \right]$$

where $C_{i,t}$ denotes his consumption in period t and ρ_t denotes the subjective discount rate between t and $t+1$. His budget is given by

$$C_{i,t} + \exp(-r_t) A_{i,t} = A_{i,t-1} + Y_{i,t}$$

where $Y_{i,t}$ denotes his income, $A_{i,t}$ his assets, and r_t the real interest rate. Aggregate assets are in zero net supply, income is the same for all consumers ($Y_{i,t} = Y_t$), and output is demand-determined ($Y_t = \int_i C_{i,t}\, di$). Finally, prices are completely rigid, inflation is fixed at zero, and the central bank directly controls r_t, subject to the ZLB constraint, or $r_t \geq 1$.

At $t = 1$, the natural rate is negative and the ZLB is binding: $\rho_1 = \rho < 0$ and $r_t = 1$. For $t \geq 2$, the natural rate is positive: $\rho_t = \bar{\rho} > 0$. For $t \geq 3$, the central bank sets the natural rate, $r_t = \bar{\rho}$, and this is assumed to induce $Y_t = 1$. At $t = 2$, the central bank may set $r_2 \neq \bar{\rho}$ and may thereby induce $Y_2 \neq 1$. Finally, the central bank may offer forward guidance at $t = 1$ about what it plans to do at $t = 2$.

The essence of the problem is captured by allowing consumers to lack common knowledge of the policy communications and/or hold mis-specified beliefs about the behavior of others when making their period-1 choices. At the same time, the dynamic complexity of the present setting is minimized by making the following simplifying assumption: for $t \geq 2$ (but not $t = 1$), consumers know both r_t and y_t when making their period-t choices and have rational expectations about the future thereafter. In other words, the friction operates only between periods 0 and 1 ("in the short run").

Individual optimality and the budget constraint yield the following consumption function:

$$c_{i,t} = (1 - \beta) \sum_{j=0}^{\infty} \beta^j \mathbb{E}[y_{t+j}] + \beta \sum_{j=0}^{\infty} \beta^j \mathbb{E}[-(r_{t+j} - (\rho_{t+j} - \bar{\rho}))] \tag{4}$$

where $c_{i,t} \equiv \log C_{i,t}$, $y_t \equiv \log Y_t$, $\beta \equiv \exp(-\bar{\rho})$, and $\mathbb{E}_{i,t}$ is the consumer's expectation in period t. The above holds for arbitrary subjective beliefs. The simplifying assumption that agents have full

information and rational expectations at $t \geq 2$ guarantees that $c_{i,t} = c_t = y_t = 0$ for $t \geq 3$ and

$$c_{i,2} = c_2 = y_2 = -r_2.$$

That is, outcomes are pinned down as in the standard New Keynesian model at $t = 2$ and thereafter—but not necessarily before. Using these facts and evaluating condition (4) at $t = 1$, we conclude that period-1 consumption satisfies

$$c_{i,1} = (1 - \beta)\mathbb{E}_{i,1}[y_1 + \beta y_2] + \beta^2 \mathbb{E}_{i,1}[-r_2] + \beta(\bar{\rho} - \varrho).$$

To simplify, ignore henceforth the term $\beta(\bar{\rho} - \varrho)$ in the last equation; for our purposes, this just an innocuous constant. It is then straightforward to show that the behavior of the economy at $t \in \{1, 2\}$ can be summarized in the following two equations:

$$k_i = \mathbb{E}_i[(1 - \gamma)\tau + \gamma Y] \ \forall i \qquad \text{and} \qquad Y = (1 - \alpha)\tau + \alpha K, \tag{5}$$

where we have dropped the time index from $\mathbb{E}_{i,1}$ and have adopted the following transformations:

$$k_i \equiv c_{1,i} \quad K \equiv \int k_{i,1} = c_1, \quad \tau \equiv -r_2, \quad Y \equiv \frac{y_1 + \beta y_2}{1 + \beta}, \quad \alpha \equiv \frac{1}{1 + \beta}, \quad \text{and} \quad \gamma \equiv 1 - \beta^2.$$

This completes the nesting of the present setting to our abstract framework and allows the interpretation of τ as the interest rate set after the economy has exited the ZLB. Finally, because a smaller β corresponds to a larger marginal propensity to consume out of current income, which in turn maps to a higher γ, we have that, for our purposes, a higher γ may represent a steeper Keynesian cross.

3.2 Investment with externalities

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 $(1 - r)Q - wN - \int p_i x_i di$, where r is the rate of taxation, $X \equiv (\int x_i^{1-\rho} di)^{1/1-\rho}$ 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.

We log-linearize around a full-information, non-stochastic steady state and introduce the following transformation of variables:

$$k_i \equiv \tfrac{1+\eta\phi}{1+\phi}(\log x_i - \log \bar{x}), \quad \tau \equiv \tfrac{1+\eta\phi}{\phi(1-\eta)}\log(1-r), \quad \text{and} \quad Y \equiv \log Q - \log \bar{Q}.$$

where $\bar{x}$ and $\bar{Q}$ are constants (the "steady-state" quantities corresponding to $r = 0$). Optimality for the final-good firm and the worker, plus market clearing, imply that output can be written as

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

with $\alpha \equiv \eta\frac{(1+\phi)^2}{(\eta+\phi)(1+\eta\phi)} \in [0,1]$. The first term captures the effect of the subsidy on labor supply and thereby on output. The second captures the role of capital in production. Optimality for the typical entrepreneur, on the other hand, implies that investment can be written as

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

with $\gamma \equiv \frac{(1+\eta\phi)(\eta\rho+\phi(\eta+\rho-1))}{\eta\rho(1+\phi)^2} < 1$ as long as $\rho < 2$. The first term captures the direct or PE effect of the subsidy on investment. The second term combines two GE effects. On the one hand, because of the aggregate-demand externality, higher aggregate output raises the individual return to investment for given wages and given τ; this contributes towards $\gamma > 0$. On the other hand, higher aggregate output boosts aggregate labor demand, raises wages, and lowers the return on capital; this contributes towards $\gamma < 0$. The restriction $\gamma > 0$, which is assumed in the main analysis but is relaxed in Appendix F, therefore amounts to assuming that the aggregate-demand externality is sufficiently strong.[5]

4 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.

4.1 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.[6] 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:

$$K = (1 - \gamma)\mathbb{E}[\tau|\hat{X}] + \alpha\mathbb{E}[Y|\hat{X}]. \tag{6}$$

---

[5] The following is also useful to note. Changing the deep parameter ρ, which controls the aggregate-demand externality, varies the reduced-form parameter γ while keeping constant α. This micro-foundation therefore justifies the interpretation of the comparative static in γ for fixed α in our abstract framework as variation in the "underlying GE feedback."

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

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:

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

$$\min_{\mathcal{A}\in\{\mathcal{A}_\tau^*, \mathcal{A}_Y^*\}, (\tau, Y)\in\mathcal{A}} \mathbb{E}[L(\tau, Y, \theta)] \tag{7}$$

The choice $\mathcal{A} \in \{\mathcal{A}_\tau^*, \mathcal{A}_Y^*\}$ captures the choice of the optimal mode of communication (instrument vs target). The choice $(\tau, Y) \in \mathcal{A}$ 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 $\mathcal{A}_\tau^* = \mathcal{A}_Y^*$. 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 (6) 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. $\mathcal{A}_\tau^* = \mathcal{A}_Y^* = \mathcal{A}^* \equiv \{(\tau, Y) : \tau = Y\}$.

That $\mathcal{A}^*$ 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,[7] 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.*

---

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

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.

4.2 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 joint 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 $\alpha < \frac{1}{2-\gamma}$, 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:

$$\mathbb{E}_i[X] = \hat{X}, \qquad \mathbb{E}_i[Y] = (1-\alpha)\mathbb{E}_i[\tau] + \alpha\mathbb{E}_i[K], \qquad \text{and} \qquad k_i = (1-\gamma)\,\mathbb{E}_i[\tau] + \gamma\mathbb{E}_i[Y], \quad (8)$$

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.

4.3 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 (8). 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*

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

Note that 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 subsidy relative to its PE effect.[8]

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 (8) can be rewritten as

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

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 $\hat{Y}$,

---

[8]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.

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*

$$k_i = (1 - \delta_Y)\hat{Y} + \delta_Y \mathbb{E}_i[K]. \tag{10}$$

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:

$$k_i = (1 - \delta_X)\mathbb{E}_i[X] + \delta_X \mathbb{E}_i[K]. \tag{11}$$

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 $X \in \{\tau, Y\}$.[9]

5   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.*

---

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

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.[10]

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), our approach decouples the two frictions by relaxing the common-prior assumption.[11]

What is more, Level-k Thinking produces similar results by recasting the parameter λ as increasing function of the depth of thinking. There is only one subtle difference, which, at least in our view, represents a "bug" of Level-k Thinking that the present specification avoids.[12]

We thus invite the reader to interpret the results presented in this section as the product of introducing plausible bounds on *either* the depth of knowledge or the depth of rationality. The available evidence is inconclusive on which interpretation is *most* relevant, but it generally supports the existence of the kind of anchored beliefs we are after in this section.[13]

5.1 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, $\bar{\mathbb{E}}[K] = \lambda K$.*

The basic idea behind this lemma is quite simple. 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 kind of reasoning, or higher-order beliefs, that underlies this property. Because we have maintained common knowledge of rationality, we can express an

---

[10]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.

[11]The importance of this decoupling is discussed in detail in Subsection 7.3.

[12]These points are made clear in Subsection 7.1.

[13]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.

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:

$$\mathbb{E}_i[K] = \mathbb{E}_i \left[(1 - \delta_X) \sum_{h=1}^{\infty} (\delta_X)^{h-1} \bar{\mathbb{E}}^h[X] \right]. \tag{12}$$

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 $h \geq 1$,

$$\mathbb{E}_i \left[\bar{\mathbb{E}}^h[X] \right] = \lambda^h \hat{X}. \tag{13}$$

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.[14]

5.2 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 (11) and aggregating across agents, we reach the following result.

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

$$K = \kappa_X \hat{X} \qquad \text{with} \qquad \kappa_X \equiv \frac{1 - \delta_X}{1 - \lambda \delta_X}, \tag{14}$$

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.*

---

[14] In particular, using (13) into (12) 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 (11) and solving for $\mathbb{E}_i[K]$ gives $\mathbb{E}_i[K] = \lambda k_i$, as claimed.

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.

5.3 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.

With instrument communication, the value τ of the instrument is pegged at $\hat{\tau}$. Condition (14) then becomes $K = \frac{1-\delta_\tau}{1-\lambda\delta_\tau}\hat{\tau}$ and condition (1) gives the outcome as $Y = \left((1-\alpha) + \alpha\left(\frac{1-\delta_\tau}{1-\lambda\delta_\tau}\right)\right)\hat{\tau}$. With target communication, instead, the outcome is itself pegged at $Y = \hat{Y}$. Condition (14) then becomes $K = \frac{1-\delta_y}{1-\lambda\delta_y}\hat{Y}$ and condition (1) gives the value of the instrument needed to hit the target $\hat{Y}$ as $\tau = \left(\frac{1}{1-\alpha} - \left(\frac{\alpha}{1-\alpha}\right)\left(\frac{1-\delta_y}{1-\lambda\delta_y}\right)\right)\hat{Y}$. Combining these findings, using the definitions of δ_τ and δ_Y, and noting that the policymaker is free to choose any $\hat{\tau}$ in the first case and any $\hat{Y}$ in the second case, we reach the following result.

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

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

where[15]

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

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.

---

[15]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.

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.

5.4 Role of the 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 interact with the belief friction in shaping the distortion of the implementability constraints.

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 $\mathcal{A}_\tau = \mathcal{A}_\tau^*$, 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.

5.5 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. With instrument communication, the agents do so to predict the

outcome; with target communication, they do so to predict the value of the instrument that will be required to honor the target. The policymaker can no longer sidestep the friction.

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 $\mathcal{A} \in \{\mathcal{A}_\tau(\lambda), \mathcal{A}_Y(\lambda)\}$ and a pair $(\tau, Y) \in \mathcal{A}$ to minimize her loss:

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

Let $(\mathcal{A}^{\text{sb}}, \tau^{\text{sb}}, Y^{\text{sb}})$ be the (unique) triplet that attains the minimum. Then, $\mathcal{A}^{\text{sb}}$ identifies the optimal mode of communication; $(\tau^{\text{sb}}, Y^{\text{sb}})$ identifies the second-best combination of the instrument and the outcome; and the communicated message is given either by $\hat{\tau} = \tau^{\text{sb}}$ or by $\hat{Y} = Y^{\text{sb}}$, depending on whether $\mathcal{A}^{\text{sb}} = \mathcal{A}_\tau$ or $\mathcal{A}^{\text{sb}} = \mathcal{A}_y$.

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:

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

$$\text{s.t.} \quad \tau = \mu Y$$

Solving the constraint for Y as τ/μ, substituting this in the objective, and letting $r \equiv \tau/\theta$, we reach the following even simpler representation:

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

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. The optimal r is given by

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

and the resulting payoff is

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

18

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).

How do we explain this shape? Recall that $\mu = 1$ is not feasible away from the frictionless benchmark. Instead, the policymaker has to choose either $\mu = \mu_\tau > 1$ (with instrument communication) or $\mu = \mu_Y < 1$ (with target communication). The policymaker can moderate the incurred loss by adjusting r, the responsiveness of τ to θ, away from $r = 1$, the frictionless value. Conditional on instrument communication, it is indeed optimal to choose $r > 1$, that is, to let the subsidy vary more strongly with the fundamental than in the frictionless benchmark. This offsets the attenuated response of Y to τ, which in turn helps reduces the wedge between Y and Y^{fb}; but since this comes at the cost of a large wedge between τ and τ^{fb}, the policymaker chooses an $r > 1$ that only partly offsets the distortion. A similar logic applies with target communication, except that now the effects flip: the policymaker chooses $r < 1$ in order to moderate the amplification effect.

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 $\mathcal{L}(\mu)$ unchanged. It follows that $\mathcal{L}(\mu_\tau)$ is an increasing function of γ, whereas $\mathcal{L}(\mu_Y)$ is a decreasing function of it. Next, note that both $\mathcal{L}(\mu_\tau)$ and $\mathcal{L}(\mu_Y)$ are continuous in γ and recall from our earlier discussion that $\mathcal{L}(\mu_\tau) = 0 < \mathcal{L}(\mu_Y)$ when $\gamma = 0$ and $\mathcal{L}(\mu_\tau) > 0 = \mathcal{L}(\mu_Y)$ when $\gamma = 1$. It follows that there exists a threshold $\hat{\gamma}$ strictly between 0 and 1 such that $\mathcal{L}(\mu_\tau) < \mathcal{L}(\mu_Y)$ for $\gamma < \hat{\gamma}$, $\mathcal{L}(\mu_\tau) = \mathcal{L}(\mu_Y)$ for $\gamma = \hat{\gamma}$, and $\mathcal{L}(\mu_\tau) > \mathcal{L}(\mu_Y)$ 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 $\gamma > \hat{\gamma}$.*

(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 r^{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.

5.6 Comparative statics

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. And as the friction vanishes, the threshold $\hat{\gamma}$ has a well-defined limit given by $\lim_{\lambda \uparrow 1} \hat{\gamma} = \frac{1}{2-\alpha} \in \left(\frac{1}{2}, 1\right)$. Thus, whereas *exact* rational expectations (nested as $\lambda = 1$) leaves the optimal form of forward guidance indeterminate, *near* rational expectations (i.e., λ arbitrarily close to, but strictly lower than, 1) gives a non-trivial answer to the question of interest.

5.7 The sign of γ

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 discussed in the next section, 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 fixed resources and can be captured in the neoclassical, investment example introduced in Section 2 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 strategic consideration

suggests that instrument communication should be necessarily optimal when $\gamma < 0$ and, hence, that our main result (Theorem 1) should remain intact. Appendix F verifies this claim this is true provided an additional, sensible assumption about the maximum possible distortion.

6   Application to Monetary Policy

We now return to the first example introduced in Section 2, that regarding monetary policy in a liquidity trap.

The Great Recession, and many Central Banks' experiences trying to manage expectations at the zero lower bound, has intensified practitioners' interest in effective policy communication (Blinder, 2018). The same issues also have also motivated a large theoretical literature on these policies' effectiveness away from from the conventional, full-rationality, full-information benchmark. An important conclusion, echoed in our own work, is that anchored beliefs can severely limit the power of forward guidance for monetary policy during a liquidity trap, under the restriction that forward guidance takes the form of a commitment to an interest-rate target.[16]

Our analysis qualifies this lesson by showing that the central bank may be able to bypass, or even flip, the friction by engaging in the opposite form of forward guidance, committing to do "whatever it takes" to meet an aggressive target for GDP or unemployment. This recommendation reminds ECB chairman Mario Dragi's famous proclamation but does not rest on the idea of ruling out a bad equilibrium out of many. Instead, it rests on the idea of minimizing the bite of "mistakes" in how agents form beliefs or reason about the economy along a unique equilibrium,

Furthermore, our analysis sheds light on the question of *when* the central bank should switch from one form of forward guidance to the other, depending on the ferocity of GE feedback mechanisms. 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.[17]

When the ZLB binds, the combination of these mechanisms amount to a strong macroeconomic complementary, or a high γ. 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 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.

Combining these insights with our own results suggests the following. Consider a situation in which the liquidity trap is expected to be sufficiently long and/or the Keynesian cross is sufficiently steep. It

---

[16]Angeletos and Lian (2018) and Wiederholt (2016) model the friction is modeled as anchored higher-order beliefs; Farhi and Werning (2019) and Garcıa-Schmidt and Woodford (2019) as Level-k Thinking.

[17]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.

is precisely then that "instrument communication" is severely constrained. But it is also then that the policymaker can, and probably should, bypass the friction by engaging on "target communication." We leave a more careful consideration of this issue for future work.

7   Alternative Frictions

To what extent do our results apply outside the *specific* model of bounded rationality assumed so far? In this section we demonstrate the robustness of our main insights to several alternatives. All in all, we show that, although the exact distortion of the implementability constraints under the two modes of communication depends on the exact departure made from the frictionless, rational-expectations benchmark, the main lesson about the optimal form of forward guidance does not.

7.1   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

$$\bar{\mathbb{E}}[K] = (1 - \delta_X) \sum_{h=0}^{k-1} (\delta_X)^h \hat{X} = (1 - (\delta_X)^k)\hat{X} \tag{15}$$

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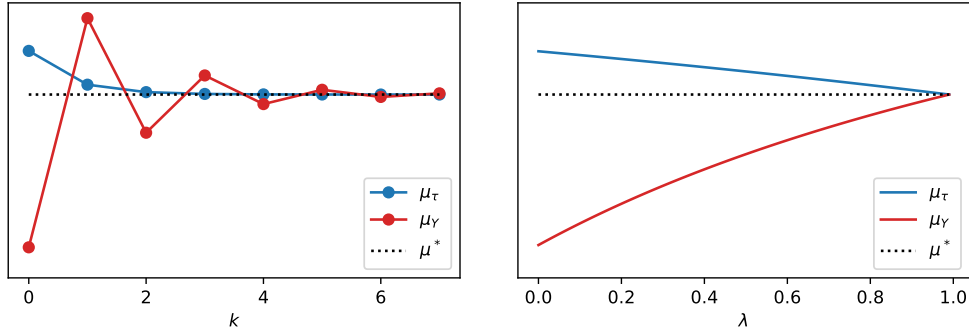
Figure 1: The implementability coefficients μ_τ and μ_Y under Level-k Thinking (left) and anchored beliefs (right).

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 5.

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 García-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 7.3.

7.2   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

23

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—anchored beliefs—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*

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

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 ($\mu_Y < 1$). Intuitively, the entire story flips. 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. 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 $\sigma > 0$).

7.3 First-order vs higher-order mistakes

Our analysis has allowed imperfect reasoning about the behavior of other agents (or the GE effects of policy) but ruled out any friction in agents' own awareness of the policy (or of its PE effects). In the language of the model, we considered "higher-order mistakes" but not "first-order mistakes." The latter kind of mistakes can obtain—either in isolation from or in combination with the former kind—due to rational inattention (Sims, 2003), sparsity Gabaix (2014), noisy or sticky information (Lucas, 1972; Mankiw and Reis, 2002), etc. We now study the role of this conceptually distinct friction.

To start, consider *only* first-order mistakes. In particular, restrict higher-order beliefs to coincide with first-order beliefs ($\bar{\mathbb{E}}^h[X] = \bar{\mathbb{E}}[X]$ for all $h \geq 2$) but allow the latter to satisfy

$$\bar{\mathbb{E}}[X] = q\hat{X},$$

for some $q \in (0, 1)$. This helps capture inattention to the policy communication while abstracting from the friction we have been after in the rest of the paper. Indeed, it is straightforward to verify that $K = \bar{\mathbb{E}}[K]$, which means precisely that there is no distortion in how the typical agent reasons about the behavior of others or the GE effect of policy. Instead, there is only a distortion in the perceived PE effect, which is now given by q times the frictionless counterpart.

Formally, there is a representative agent whose equilibrium behavior is given the solution to the following fixed point:

$$K = (1 - \delta_X)qX + \delta_X K.$$

It is then immediate that $q < 1$ attenuates the response of K to the policy announcement under *both* modes of communication, by the same degree, and in a manner that is invariant to GE considerations. Plain inattention therefore does not deliver any of our main insights.

Instead, a *different* insight emerges: with only inattention, instrument communication is necessarily optimal. This is because the implementability constraints, represented again as $\tau = \mu_\tau Y$ for instrument communication and $\tau = \mu_Y Y$ for target communication, now feature $\mu_Y > \mu_\tau > 1$, or a globally smaller distortion for instrument communication.[18] The wedge between the two options, though, does not depend on γ, underscoring that the mechanism through which inattention works is orthogonal to that identified in our main analysis.

What if we flexibly combine the two kinds of friction? To this end, continue to assume that first-order beliefs satisfy $\bar{\mathbb{E}}[X] = qX$ but now let higher-order beliefs also feature the distortion accommodated in our main analysis:

$$\bar{\mathbb{E}}^h[X] = \lambda^{h-1}\bar{\mathbb{E}}[X], \tag{16}$$

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,

$$\mu_\tau = \frac{1}{1 - \alpha + \frac{1-\alpha\gamma}{1-\alpha\gamma\lambda}\alpha q} \quad \text{and} \quad \mu_Y = \frac{1 - \alpha + \alpha(1-\gamma)\lambda - \alpha(1-\alpha\gamma)q}{(1-\alpha)\left(1 - \alpha + \alpha(1-\gamma)\lambda\right)}. \tag{17}$$

---

[18]This can be verified by evaluating condition (17) under the restriction $\lambda = 1$.

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 ($\mu_Y \lessgtr 1$), 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.[19]

Of particular interest is the case $q = \lambda$, which is isomorphic to a rational expectations model with a Gaussian prior and Gaussian private signals.[20] 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 $\lambda < q$.

The case $\lambda < q$, while ruled out by the "canonical" Gaussian example, is *not* necessarily inconsistent with rational expectations. For example, it can be obtained, at least under certain states of nature, in a rational-expectations setting in which agents are uncertain about the *precision* of others' private signals.[21] We nevertheless prefer the interpretation in terms of bounded rationality, and in particular our extreme version of $\lambda < q = 1$, because it isolates agents' imperfect reasoning about each other's behavior. As already explained, the case $\lambda < q = 1$ also captures Level-k Thinking.

Another special case of interest is $\lambda = q = 1$ and $\sigma > 0$. 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 multiple-equilibrium logic behind Mario Draghi's proclamation to do "whatever it takes." The common thread is the presence of a strong GE feedback, or a large concern about the behavior of others.

8 Stretching the Model

Our framework was intentionally designed to focus on the issue of greatest interest to us: the accommodation of plausible bounds on the agents knowledge or rationality. In this section we discuss how our insights extend to more general contexts, with more complicated policy considerations or richer policy options. The general theme is that, relative to the conventional policy paradigm, our approach *uniquely* links policy choice to strategic interaction, or the GE feedback.

---

[19]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 $\gamma = 1$.

[20]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 $q = \lambda = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_s^2}$.

[21]Such an example is contained in Angeletos and La'O (2009), although it is used for different purposes there.

8.1 Additional shocks and relation to Poole (1970)

The case with noisy expectations nested in Section 7.2 recalls Poole (1970)'s analysis in its focus on dampening the effects of "unwanted" shocks. However, this resemblance is somewhat superficial. In Poole's classic, and in the modern literature on optimal monetary policy alike, the irrelevance of different policy regimes is broken because, and only because, the policymaker does not observe underlying shocks to payoff-relevant fundamentals (e.g., preferences, technology, or monopoly power), or is otherwise unable to condition the policy instrument directly on them. Here, instead, the key friction is how agents form beliefs or reason about the behavior of others—not what the policymaker knows or what she can condition her choices on.

This explains the following two key differences between our contribution and the state of the art. First, our result is robust to letting the policymaker condition her action on the "unwanted" shock. And second, whereas our analysis emphasizes the dependence of the optimal policy choice on the parameter γ, or on the decomposition of the equilibrium between PE and GE effects, a variant that embeds Poole's considerations does not share this insight.

With regard to the first point, it is indeed straightforward to check that our results with noisy beliefs is robust to the policymaker's observing a (possibly perfect) signal of the noise ε. Our result is driven by how the form of forward guidance helps regulate the impact of the "unwanted" mistake in beliefs, regardless of the policymaker's observation of it. The same point is evident from the irrelevance of the policymaker's conditioning, in various examples, on belief inertia λ, degree of thinking k, or size of shocks σ.

As for the second point, Appendix C works out variations of our model that introduce measurement error in the policymakers' observation of the outcome, trembles in her control of the policy instrument, or other payoff-relevant shocks that affect the relation between τ and Y without introducing errors in the agents' beliefs or reasoning about one another's behavior. Such shocks may tilt the balance toward either policy choice, via a logic similar to Poole (1970). However, such shocks do *not* deliver a dependence of the optimal strategy on the parameter γ. This dependence instead emerges in our setting, with or without such shocks, because the form of forward guidance and the parameter γ interact in shaping how behavior depends on the agents' reasoning about the behavior of others and hence also on any mistakes, random or not, in such reasoning.

In a nutshell, the policy considerations put forward here are distinct from those captured by Poole (1970). The same point applies to Weitzman (1974) and the literature on "quotas vs tariffs" that follows his lead. An application of our insights to that literature may indeed shed light on how the answer to that classic question depends on the interaction of bounded rationality and strategic considerations.

8.2 Sophisticated forward guidance and policy rules

Although our analysis presumes a binary choice between "talking about interest rates" and "taking about unemployment," the reality is that central bankers typically talk about both all the time. 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.

Such a flexible interpretation is corroborated by an exercise that expands the forms of forward guidance the policymaker can engage to. Suppose, in particular, that the policymaker can announce and commit to a flexible *relation* between the instrument τ and the outcome Y, given by

$$\tau = A - BY, \tag{18}$$

for some $(A, B) \in \mathbb{R}^2$. The two simpler strategies considered so far 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. In this benchmark, the optimal pair (A, B) is indeterminate. 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: there is a unique such pair that minimizes, indeed eliminates, the bite of higher-order beliefs. Furthermore, the following "smooth" version of our take-home lesson applies: a larger γ, or stronger GE feedback, maps to a larger optimal value for B, which can be interpreted as a tilt towards "taking about Y" relative to "talking about τ."[22]

The more sophisticated forms of forward guidance allowed in this extension may be hard to explain and communicate, especially when the intended audience is the general public and the true environment is richer than the simple model consider here. 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 thus be more effective than complex rules for reasons left outside the analysis.

Nevertheless, the aforementioned 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 entirely new function: regulating the impact of bounded rationality. The application of this insight to the class of richer, dynamic models considered in that literature seems an interesting direction for future research.

---

[22] 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.

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

9   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?

---

[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.

We pose this question in a stylized model in which agents form mis-specified beliefs, either anchored to a reference point or subject to erratic impulses. 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 in a model of high aggregate demand externalities or a steep Keynesian cross.

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, leaving agents to predict what policy will support the announced outcome. Which strategy is preferred depends on the relative cost of mistakes for each type of reasoning. High GE feedbacks, which make outcome expectations more essential for decisions (and associated mistakes more costly), tilt the balance toward directly communicating those outcomes.

Put more succinctly, the optimal form of forward guidance minimizes agents' need to "reason about the economy" precisely because this reasoning produces distortions.

Along the way, we uncovered additional insights, such as how Taylor rules can play a new role in regulating the bite of mis-specified beliefs, or bounded rationality. But our analysis remained too stylized to give fully satisfying answers. We also took for granted the desirability of minimizing the distance of the equilibrium outcomes from their rational-expectations counterparts. But one could imagine situations with one distortion offsetting another—for instance, anchored beliefs offsetting financial amplification. Last but not least, we abstracted from the possibility of multiple instruments and/or multiple policy goals. Each of these issues merits a more complete investigation.

References

Abreu, Dilip, and Markus K Brunnermeier. 2003. "Bubbles and crashes." *Econometrica*, 71(1): 173–204.

Akerlof, Georges, 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.

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*.

Bergemann, Dirk, and Stephen Morris. 2018. "Information Design: A Unified Perspective." *Cowles Foundation Discussion Paper No. 2075R3*.

Bernanke, Ben. 2015. "Inaugurating a New Blog." *https://www.brookings.edu/blog/ben-bernanke/2015/03/30/inaugurating-a-new-blog/*, Accessed: 2018-12-10.

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*.

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.

Evans, Georges, and Seppo Honkapohja. 2001. *Learning and Expectations in Macroeconomics.* Princeton, NJ:Princeton University Press.

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

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*.

García-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.

Inostroza, Nicolas, and Alessandro Pavan. 2018. "Persuasion in Global Games with Application to Stress Testing." *Northwestern University mimeo*.

Iovino, Luigi, and Dmitriy Sergeyev. 2017. "Quantitative Easing without Rational Expectations." *Work in progress*.

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

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian Persuasion." *American Economic Review*, 101(6): 2590–2615.

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

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.

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

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

Morris, Stephen, and Hyun Song Shin. 2007. "Optimal Communication." *Journal of the European Economic Association*, 5(2-3): 594–602.

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.

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

Sargent, Thomas J. 1993. *Bounded rationality in macroeconomics*. Oxford University Press.

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

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

Svensson, Lars E. O. 2006. "Social Value of Public Information: Comment: Morris and Shin (2002) Is Actually Pro-Transparency, Not Con." *American Economic Review*, 96(1): 448–452.

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

Wiederholt, Mirko. 2016. "Empirical Properties of Inflation Expectations and the Zero Lower Bound." *Goethe University Frankfurt mimeo*.

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*.

Woodford, Michael. 2013. "Macroeconomic Analysis Without the Rational Expectations Hypothesis." *Annual Review of Economics*, 5: 303–346.

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,

$$\mu_\tau(\lambda, \gamma) = \left((1 - \alpha) + \alpha \frac{1 - \delta_\tau}{1 - \lambda \delta_\tau} \right)^{-1} \tag{19}$$

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

$$\mu_Y(\lambda, \gamma) = \frac{1 - \alpha b}{1 - \alpha} = \frac{1 - \lambda \delta_y - \alpha(1 - \delta_y)}{(1 - \alpha)(1 - \lambda \delta_y)} \tag{20}$$

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 $\gamma \geq 0$. $\square$

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:

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

35

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 $\mu_\tau = \mu_Y$. $\qquad\square$

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 $\mu_Y \geq 0$.

36

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

$$r^*(\mu) := \frac{\mu^2(1 - \chi) + \mu\chi}{\mu^2(1 - \chi) + \chi} \tag{21}$$

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

Finally, let $\mathcal{L}_\tau$ and $\mathcal{L}_Y$ 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 $\mathcal{L}_\tau(\gamma)$ and $\mathcal{L}_Y(\gamma)$ denote these losses as function of γ. Note that, by the chain rule, $\partial \mathcal{L}_\tau/\partial \gamma = \partial \mathcal{L}/\partial \mu \cdot \partial \mu_\tau/\partial \gamma$ and $\partial \mathcal{L}_Y/\partial \gamma = \partial \mathcal{L}/\partial \mu \cdot \partial \mu_Y/\partial \gamma$. 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 $\mathcal{L}_\tau(0) = \mathcal{L}_Y(1) = 0$ 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 $\mathcal{L}_\tau(\hat{\gamma}) = \mathcal{L}_Y(\hat{\gamma})$.

In particular, $\mathcal{L}_\tau(\gamma)$ is strictly increasing and $\mathcal{L}_Y(\gamma)$ is strictly decreasing on the domain $\gamma \in (0, 1)$. By the previous argument, to show $\partial \mathcal{L}_\tau/\partial \gamma > 0$ and $\partial \mathcal{L}_Y/\partial \gamma < 0$, 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 $\partial \mathcal{L}_Y/\partial \gamma \leq 0$ 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, $\partial \mathcal{L}/\partial \mu > 0$ since $r^*(\mu_Y) < 0$. But we also know that $\lim_{\mu \to -\infty} \mathcal{L}(\mu) = \chi$. 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 $\mathcal{L}_Y(\gamma) > \chi$ for $\gamma \in (-\infty, \check{\gamma}]$. Meanwhile, a similar argument for

37

$\mu > 1$ (with $\lim_{\mu \to \infty} \mathcal{L}(\mu) = \chi$ and $\partial \mathcal{L}/\partial \mu > 0$) suggests that $\mathcal{L}_\tau(\gamma) < \chi$ for $\gamma \geq 0$. This shows that $\mathcal{L}_Y(\gamma) > \chi > \mathcal{L}_\tau(\gamma)$ 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:

$$\lim_{\gamma \to -\infty} \mu_\tau = \mu_{\tau,-\infty} \equiv \frac{\lambda}{\lambda + (1 - \lambda)\alpha} \in (0, 1) \tag{22}$$

$$\lim_{\gamma \to -\infty} \mu_Y = \mu_{Y,-\infty} \equiv \frac{\lambda(1 - \alpha/\lambda)}{\lambda(1 - \alpha)} \tag{23}$$

Proof of Proposition 8

The critical GE feedback threshold satisfies $\mathcal{L}_\tau(\hat{\gamma}) = \mathcal{L}_Y(\hat{\gamma})$. 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 $\gamma \in [0, 1]$:

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

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

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

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

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

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

$$c_{i,t} = (1 - \beta) \sum_{j=0}^{\infty} \beta^j \mathbb{E}[y_{t+j}] + \beta \sum_{j=0}^{\infty} \beta^j \mathbb{E}[-(r_t - (\rho_t - \bar{\rho}))] \tag{24}$$

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 ((24)) 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:

$$c_{i,1} = (1 - \beta)\mathbb{E}_i[y_1 + \beta y_2] + \beta^2 \mathbb{E}_i[-r_2] + \beta(\bar{\rho} - \rho) \tag{25}$$

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 ((25)), 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

The set up was described in Section 2, p.3.2. Here, we solve the model and explain how it is nested in our abstract framework.

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

$$x_i^\rho = \eta(1-\rho)\mathbb{E}_i \left[(1-r)X^{\eta+\rho-1}N^{1-\eta}\right] \tag{26}$$

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 26 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 $K = \int_i k_i \, di$ 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:

$$Y = (1 - \alpha)\tau + \alpha K \tag{27}$$

with

$$\alpha \equiv \frac{\eta(1+\phi)^2}{(\eta+\phi)(1+\eta\phi)} \tag{28}$$

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 (26). 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

$$k_i = (1 - \gamma)\, \mathbb{E}_i[\tau] + \gamma\mathbb{E}[Y] \tag{29}$$

for feedback parameter

$$\gamma \equiv \frac{(1+\eta\phi)(\rho(\eta+\phi) - \phi(1-\eta))}{\eta\rho(1+\phi)^2} \tag{30}$$

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   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.

C.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.

C.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.

D   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.

D.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.

D.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 5 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 5.6).

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).

E   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:

$$\tau = A - BY \tag{31}$$

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 4). 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 $\mathcal{A}^*$, 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 (31) and (1) jointly for τ and Y and substitute the solution into (2) to obtain the following game representation:

$$k_i = \zeta(A, B; \alpha, \gamma) + \delta(B; \alpha, \gamma)\mathbb{E}_i[K] \tag{32}$$

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 (32).

Consider now the case with anchored beliefs (as in Section 5). 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 $\mathcal{A}_\tau$ and $\mathcal{A}_Y$. What is more, there actually exist a subclass of policy rule that replicates $\mathcal{A}^*$, 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 $\mathcal{A}^*$. And by picking A so that $\zeta(A, B; \alpha, \gamma) = \theta$, she can implement the first best.[24]

We summarize these lessons in the following result.

Proposition 10. *Suppose that the policymaker can announce and commit on a policy rule as in (31) 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

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

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,

---

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

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 10 as a complement to our main analysis, not a sign that the choice between instrument and target communication was too narrowly framed. Proposition 10 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 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.

F Other parameter cases

F.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 3.2 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

50

case.

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

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

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

Proposition 11. *For any values of $\alpha \in (0, 1)$, $\sigma \in [0, 1)$, and $\gamma \leq 0$, $\psi_Y > \psi_\tau > 0$.*

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 $\delta_Y < \delta_\tau \leq 0$ for $\gamma \leq 0$. Thus $\delta_\tau = -\alpha g(\delta_\tau) < -\alpha g(\delta_Y) < -\alpha g(\delta_y)/(1 - \alpha) = \psi_Y$. $\qquad \square$

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.

F.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 4 it guarantees that the joint of Assumptions 1 and 2 replicates the REE benchmark; in Section 7.1, it guarantees that the Level-k outcome converges to the REE outcome as agents become "infinitely rational" ($k \to \infty$); and in Sections 5 and 7.2, 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 5 and recast it in terms of heterogenous 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.