NBER WORKING PAPER SERIES

MANAGING EXPECTATIONS WITHOUT RATIONAL EXPECTATIONS

George-Marios Angeletos
Karthik A. Sastry

Working Paper 25404
http://www.nber.org/papers/w25404

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2018

Angeletos acknowledges the financial support of the National Science Foundation (Award 1757198). We have no other financial support, interest or conflict to disclose. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Managing Expectations without Rational Expectations
George-Marios Angeletos and Karthik A. Sastry
NBER Working Paper No. 25404
December 2018
JEL No. D82,D84,E03,E52,E58

ABSTRACT

Should a policymaker offer forward guidance by committing to a path for the policy instrument or a target for an equilibrium outcome? 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 policy minimizes the bite of such mistakes on implementability and welfare. This goal is achieved by fixing and communicating an outcome target if and only if the GE feedback is strong enough. Our results suggest that central banks should stop talking about interest rates and start talking about unemployment when faced with a steep Keynesian cross or a prolonged liquidity trap.

George-Marios Angeletos
Department of Economics, E52-530
MIT
50 Memorial Drive
Cambridge, MA 02142
and NBER
angelet@mit.edu

Karthik A. Sastry
Department of Economics, MIT
77 Massachusetts Avenue E52-300
Cambridge, MA 02139
ksastry@mit.edu

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

Ben Bernanke, 2015

1 Introduction

Forward guidance, the art of managing expectations, is rarely comprehensive. A central bank may, for example, successfully move the yield curve by announcing the intended path for future short-term rates. But it remains up to the market to predict how this will affect equilibrium outcomes such as GDP or unemployment. Under what circumstances, we ask, would it be better to engage in the opposite type of forward guidance, namely committing to a sharp target for the relevant equilibrium outcome and leaving the market to forecast the policy that will support this target?

We study how the answer to this question depends on certain kinds of bounded rationality or belief biases. We work with a stylized model which abstracts from familiar considerations such as lack of commitment or imperfect control of the targeted outcome and instead focuses on how agents form expectations, or reason, about the behavior of others and the effects of the policy.[1] Our main result shows a sharp dependence of the optimal communication strategy on the feedback between aggregate outcomes and individual actions ("GE considerations"). Fixing outcomes instead of instruments is optimal if and only if this feedback is sufficiently high, as in situations with a steep Keynesian cross or a prolonged liquidity trap.

This result may help gauge when central bankers should stop talking about interest rates and start talking about unemployment. More broadly, our analysis sheds new light on the role played by conventional, strong assumptions about agents' depth of knowledge and rationality. In our model, these assumptions make the choice between fixing the policy instrument and fixing the targeted outcome irrelevant precisely because agents can, without mistakes, understand and invert the mapping between the two. Realistic, structured relaxations of these assumptions not only allow such mistakes to exist, but also deliver sharp insights on how a policymaker can minimize their impact on aggregate outcomes. These insights extend to the design of more sophisticated policy rules.

Framework. To fix ideas, consider the following context nested in our abstract model. There are many investors, or firms, whose joint behavior determines aggregate output (the targeted outcome), and a policymaker, who can provide a production subsidy (the policy instrument). Holding constant everything else, an individual invests more when she expects a higher subsidy. Because of an aggregate demand externality (the GE feedback), she also invests more when she expects higher aggregate output. The latter in turn depends on the aggregate investment as well as the rate of taxation. This is because taxes discourage labor supply in addition to investment.

The policymaker wishes to minimize the distance of the policy instrument and the targeted outcome from some ideal point, which can be interpreted as the first best. This point moves with a shock

[1] The first set of issues is the subject of a large literature; see, *inter alia*, Chari, Christiano and Eichenbaum (1998), Atkeson, Chari and Kehoe (2007), Barro and Gordon (1983), Poole (1970), and Friedman (1990).

that is observed by the policymaker and unobserved by the private agents. The agents do not care about the shock itself, because it enters their payoff only through the policymaker's chosen subsidy. The sole purpose of this shock is therefore to make sure that the agents do not *a priori* know what the policy is going to be; rather, it is essential that the policymaker "talks."

We finally let the policymaker choose between two strategies for managing expectations. In the first, she announces, and commits to, a value for the subsidy (the policy instrument). In the second, she does the same with a target for aggregate output (the relevant outcome). We refer to the former strategy as *instrument communication* and to the latter one as *target communication*.[2]

REE benchmark. Our frictionless benchmark imposes, like most of the literature, Rational Expectations Equilibrium (REE). In this benchmark, the aforementioned choice is irrelevant: the implementable combinations of policies and outcomes, and the expectations that support them, are invariant to whether the policymaker follows the one strategy or the other. In other words, there is no room for managing expectations via the aforementioned choice.

This irrelevance result, which mirrors the equivalence of the "primal" and "dual" approaches to the textbook Ramsey exercise, depends on two assumptions about agents' rationality and depth of knowledge. First, every agent is *herself* rational, aware of the structure of the economy, and attentive to the policy announcement. Second, this fact is common knowledge: to any order of reasoning ("I think that you think..."), the previous property is known. REE embeds both of these assumptions. The forms of bounded rationality considered in this paper relax the second, stronger assumption.

Anchored beliefs. Our main specification lets agents doubt the attentiveness or the knowledge of others; a variant has them question their rationality. In the former case, REE is replaced by Perfect Bayesian Equilibrium with heterogeneous priors about others' information; this builds on Angeletos and La'O (2009) and is tightly connected to a literature that studies the inertia of higher-order beliefs in common-prior settings.[3] In the latter case, REE is replaced by Level-k Thinking.[4]

These two approaches display a few subtle differences, which lead us to marginally favor the first one. Both of them, however, help capture essentially the same friction in expectations: they anchor the expected responses of others. In the first, an agent expects others to respond less than in the REE benchmark because he believes that some of them are inattentive. In the second, the same expectation is justified by the belief that others are less sophisticated. This friction seems consistent with both survey evidence on expectations and experimental evidence.[5] Our contribution is to study *how* the policymaker should go about managing expectations in the presence of this friction.

[2] Our framework equates each type of communication with a commitment to a different policy. But since we abstract from time-inconsistency problems, and since we have in mind situations where the policymaking is "98% talk" (per Bernanke's quote), we prefer the interpretation in terms of communication and expectations management.

[3] Abreu and Brunnermeier (2003), Angeletos and Lian (2018), Morris and Shin (2006), Nimark (2008), Woodford (2003*a*).

[4] Nagel (1995), Garcıa-Schmidt and Woodford (2018), Farhi and Werning (2017).

[5] See Coibion and Gorodnichenko (2012) and Coibion et al. (2018) for the former, and Crawford, Costa-Gomes and Iriberri (2013) and Mauersberger and Nagel (2018) for the latter.

Main results. Our first lesson regards the impact of anchored beliefs on the implementability constraint faced by the policymaker. The sign of this impact depends crucially on the communication strategy. The letter controls the nature of the strategic interaction between the agents, which in turn regulates whether the belief distortion attenuates or amplifies the GE effect of the policy.

With instrument communication, the investors play a game of strategic complementarity: conditional on the tax, the individual incentive to invest increases with the expected aggregate investment. In such a game, the belief friction produces attenuation: when an agent expects the others to respond less, she responds less herself. As a result, the implementability constraint is steeper: a larger change in taxes is needed in order to induce the same change in output.

With target communication, everything flips. The investors now play a game of strategic substitutability. Conditional on an announced GDP target, a firm that expects a higher aggregate investment also expects a lower required subsidy to that target, which reduces the incentive to invest. As a result, the belief friction produces amplification rather than attenuation: when an agent expects the others to respond less, she responds more. By the same token, the implementability constraint is now flatter.

Our second lesson relates to the interaction between the communication choice and the underlying GE effect. One the one hand, the mode of communication regulates which object the private agents have to forecast. Instrument communication lets the private agents know the tax at the expense of facing uncertainty about aggregate output; and the converse is true with target communication. On the other hand, the GE effect regulates which of these two objects are relatively more important in shaping actual behavior. When the GE effect is weak, investors care relatively more about taxes; when the GE effect is strong, investors care more about aggregate demand.

Combining these observations, we reach the following result: when the GE effect is weak, the bite of the belief friction on implementability is minimized by fixing the policy instrument; when the GE effect is strong, the same goal is accomplished by fixing the targeted outcome. Under the assumption that the first best is attainable with rational expectations,[6] this naturally translates to the policy recommendation stated in the beginning: target communication is the best means for managing expectations if and only if the GE effect is sufficiently strong.

Forward guidance. A recent literature has argued that the kind of anchored beliefs we accommodate here may help explain why forward guidance was much less effective in practice than what predicted by the New Keynesian model.[7] This literature presumes that forward guidance takes the form of a commitment on the future path of the policy instrument and proceeds to study how the belief friction influences aggregate spending and income. Our paper flips the question to a normative one: what is

[6]This assumption is not needed at all for our results regarding implementability, but sharpens our normative conclusions by letting bounded rationality be the *only* source of distortion relative to the first best.

[7]Angeletos and Lian (2018) attribute the anchored beliefs to lack of common knowledge; Garcıa-Schmidt and Woodford (2018) and Farhi and Werning (2017) to Level-k Thinking; and Gabaix (2017) to "cognitive discounting." Other resolutions to the forward-guidance puzzle, which do not center on beliefs, include those articulated in Del Negro, Giannoni and Patterson (2015), McKay, Nakamura and Steinsson (2016) and Kaplan, Moll and Violante (2016).

the optimal form of forward guidance in the presence such a friction?

Although our framework is too abstract to accommodate relevant institutional details or permit a quantitative evaluation, our results suggest the following: as the economy transitions from normal times to a liquidity trap, the central bank should stop talking cautiously about interest rates and instead start promising to do what "whatever it takes" to bring unemployment down. This is because a liquidity trap switches on strong GE feedback effects between income, spending and inflation, which in turn tilt the balance in favor of target communication.

Erratic beliefs. While most of our analysis focuses on the case with anchored beliefs, the main lessons also apply to a variant featuring erratic beliefs, or animal spirits. By this we mean situations in which the agents' behavior be swayed by waves of optimism and pessimism about the effectiveness of the policy. Such waves resemble sunspots but do not rely on multiple equilibria. They reflect correlated shocks to the agents' higher-order beliefs, or to their sophistication and reasoning.[8]

Regardless of whether bounded rationality takes the form of anchored or erratic beliefs, the key object is the wedge between the agents' subjective expectations of the endogenous outcomes, or the actions of others, and the frictionless REE counterparts. This wedge is correlated with the policy announcement in the first case (anchored beliefs) and uncorrelated in the second one (erratic beliefs). In both cases, however, the impact of the wedge on implementability and welfare is minimized with target communication if and only if the GE feedback is large enough. It follows that our main insights directly extend from the one form of bounded rationality to the other.

Broader scope. The main insights extend to more sophisticated communication strategies, such as when the policymaker commits to a certain relation between the instrument and the outcome (e.g., a Taylor rule for monetary policy). This allows to identify a new function for policy rules in general.

Our analysis also adds to the discussions of the "tightness" with which different policies can steer the economy. Such discussions typically focus on the policymaker's uncertainty about the underlying fundamentals (Poole, 1970; Friedman, 1990), with some authors adding a commitment problem (Atkeson, Chari and Kehoe, 2007). Our emphasis on bounded rationality and higher-order beliefs as the source of "looseness" introduces different economics to these discussions.

Layout. Section 2 discusses the related literature. Section 3 introduces our framework. Section 4 studies the rational expectations benchmark. Section 5 shows how this benchmark hides a layer of strategic interaction that is directly affected by the communication choice. Section 6 studies our main specification of bounded rationality, anchored beliefs with heterogeneous priors. Section 7 considers the variant with Level-k Thinking. Section 8 discusses the application to forward guidance. Section 9 considers the variant with erratic beliefs. Section 10 extends our analysis to additional policy options. Section 11 concludes.

[8] Similar mechanisms are at the core of the literature on speculation and bubbles (Allen, Morris and Postlewaite, 1993*a*; Scheinkman and Xiong, 2003). The particular formalization adopted here mirrors Angeletos, Collard and Dellas (2018*a*), while the re-interpretation in terms of sophistication helps build a bridge to Akerlof and Shiller (2009).

2 Related Literature

Our focus on the interplay between policy communications and higher-order beliefs is reminiscent of the literature that was spurred by Morris and Shin (2002); see, *inter alia*, Amador and Weill (2010), Angeletos and Pavan (2007), Chahrour (2014), James and Lawler (2011), Morris and Shin (2007), Svensson (2006), and Walsh (2007). In this literature, communication means varying the precision of a public signal about exogenous fundamentals, holding constant the agents' strategic interaction. In our paper, instead, communication means regulation of the agents' strategic interaction and of the equilibrium impact of bounded rationality via commitment to a specific policy rule. Related in this respect is Angeletos and Pavan (2009), which studies how certain policy rules that influence the agents' strategic interaction also influence the decentralized use of dispersed private information, its aggregation through prices or other statistics, and the equilibrium impact of noise.

Our paper also contributes to the literature on "targets vs instruments" that followed Poole (1970). This literature constraints the ways in which the policymaker can react to different shocks (such as supply and demand shocks) under different policy regimes (such as fixing the interest rate or the growth rate of money), and proceeds to study how the optimality of the different regimes depends on the composition of shocks. The modern literature on optimal Taylor rules (Woodford, 2003*b*) essentially follows the same lead. Our paper highlights a novel aspect, namely how different policy strategies can regulate the impact of bounded rationality and higher-order beliefs. The same basic point also distinguishes our contribution from Weitzman (1974)'s classic on "prices vs quantities."

Another arm of the literature, following the tradition of Barro and Gordon (1983), focuses on time inconsistency. In particular, Atkeson, Chari and Kehoe (2007) studies how the choice of targets vs instruments regulates the bite of the commitment problem by influencing the ability of the market to detect deviations. Although in this paper we are not interested in this issue (we assume full commitment), our results hint that bounded rationality may itself be a source of time inconsistency.

Last but not least, our paper bridges two methodological traditions. On the one hand, our main specifications, which assume heterogeneous priors and build on Angeletos and La'O (2009) and Angeletos, Collard and Dellas (2018*b*), can be thought of interchangeably as forms of bounded rationality and as convenient short-cuts for the anchoring and volatility effects that higher-order uncertainty generates *even* in common-prior, rational-expectations settings. Such anchoring is documented by, *inter alia*, Abreu and Brunnermeier (2003), Angeletos and Lian (2018), Morris and Shin (2006), Nimark (2008), and Woodford (2003*a*), whereas the link between higher-order uncertainty and animal spirits is the theme of Angeletos and La'O (2013), Benhabib, Wang and Wen (2015), and Huo and Takayama (2015). On the other hand, our Level-k variants follow the lead of Nagel (1995). Recent applications of Level-k Thinking to monetary policy include Garcıa-Schmidt and Woodford (2018), Farhi and Werning (2017), and Iovino and Sergeyev (2017). Our main contribution *vis-à-vis* all these works is to show how the communication of different policy commitments can regulate the agents' strategic interaction and thereby the bite of the belief distortions, regardless of their precise micro-foundations.

3 Framework

This section introduces the physical environment, the objective of the policymaker, the incentives of the private agents, and the timing of actions. We postpone, however, the specification of the agents' depth of knowledge and of their higher-order beliefs until later.

Basic 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 $\alpha \in (0, 1/2)$ 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.[9]

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 an investment subsidy or the interest rate set by the central bank, and the expectations of an aggregate outcome, such as aggregate output or demand. 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 these decisions. In applications, this interdependence often emerges from 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 underlying GE feedback, or the "macroeconomic complementarity."

The third assumption and the parameter α, on the other hand, play a more mechanical function. Had $1-\alpha$ been zero, the policymaker could not possibly commit to a specific target for Y "no matter what" (i.e., regardless of K). Letting $\alpha < 1$ simply makes sure that such a commitment is viable.

[9] The restriction $\alpha < 1/2$ serves a mostly technical purpose, namely to guarantee that the procedure of iterating on best responses converges. Footnotes 17 and 22 explain the precise role of this restriction, as well as why it is without serious loss of generality to relax this restriction and let α take any value in $(0,1)$.

To illustrate these points, and to fix ideas, we next discuss how our framework can nest a stylized neoclassical economy, in which τ represents a subsidy, or the negative of a tax instrument, K represents investment, and γ indexes the strength of an aggregate-demand externality. A more topical application, which is discussed in Section 8 and allows our results to speak to the context of forward guidance, recasts τ as the negative of the interest rate set by the central bank, K as consumer spending, and γ as the Keynesian income-spending multiplier. Yet another, possibility, which we do not explore, is a GE financial accelerator operating through asset prices, in the tradition of Bernanke, Gertler and Gilchrist (1999) and Kiyotaki and Moore (1997).

A micro-foundation. The economy is populated by a continuum of monopolistic entrepreneurs, who are the key players in the economy. In addition, there is a representative final-good firm and a representative worker, who only play auxiliary roles for our purposes. There are three periods, $t \in \{0, 1, 2\}$. The first period, $t = 0$, identifies the moment at which the policymaker commits to, and announces, a policy plan. Each entrepreneur, indexed by $i \in [0, 1]$, consumes at $t \in \{1, 2\}$ and her utility is given by $c_{i,1} + c_{i,2}$, where $c_{i,t}$ denotes his consumption in period t. The worker lives, works, and consumes only at $t = 2$, and his utility is given by $C_w - \frac{1}{1+\phi} N^{1+\phi}$, where $C_w = wN$ is his consumption, N is his labor supply, w is the real wage, and $\phi > 0$ parameterizes the Frisch elasticity of labor supply.

At $t = 1$, each entrepreneur allocates a unit endowment of the current final good toward either consumption, $c_{i,1}$, or the production of a differentiated capital good, x_i, which can be used in the production of the final good next period. The entrepreneur's budget at $t = 1$ is therefore given by $c_{i,1} + x_i = 1$ and the choice of x_i represents an investment decision, which, as it will be shown shortly, directly maps to the choice of k_i in our abstract framework.

At $t = 2$, the entrepreneur sells his good to a competitive final-good firm and consumes the profits. His budget in that period is therefore given by $c_{i,2} = p_i x_i$, where p_i denotes the price at which he sells his good to the final-good firm. That firm also hires the representative worker. Its output—that is, the period-2 quantity of the final good, or GDP—is given by $Q = X^\eta N^{1-\eta}$, where $X \equiv \left(\int_i X_i^{1-\rho} \, \mathrm{d}i \right)^{1/1-\rho}$ is a CES (constant elasticity of substitution) aggregator of the differentiated capital goods, N is the labor input, and $\rho \in [0, 1]$ and $\eta \in [0, 1]$ are fixed parameters determining, respectively, the elasticity of substitution of the differentiated goods and the income share of capital.

We let the policymaker impose a uniform tax, at rate r, on labor supply at $t = 2$. The proceeds of the tax are used to finance the production of a public good, which enters the utility of all agents in a separable way. We finally require that the entrepreneurs know the structure of the economy (including all the equations described above) but allow them to have possibly irrational expectations about one another's behavior at $t = 1$, when they maker their investment decision.

This economy is highly stylized, yet it captures essential aspects of the kind of richer, neoclassical models often used in business-cycle research. The simplifications made here—few periods, linear utility, an uninteresting worker, a single tax instrument—help nest this economy to our more abstract framework. The detailed derivations can be found in Appendix B.1. Here, we give the bottom line.

Consider a log-linear approximation of the equilibrium and let $k_i \equiv \log x_i$, $K \equiv \int k_i \, \mathrm{d}i$, and $Y \equiv \log Q$. The production function at $t = 2$ can be written, up to a constant, as $Y = \eta K + (1-\eta) \log N$. Taking the aggregate investment, K, and the tax rate, r, as given, and solving out the equilibrium value for N, we obtain the following expression for aggregate output:

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

where $\alpha \equiv \eta(1+\phi)/(\phi+n)$ and $\tau \equiv \frac{1}{\phi}\log(1-r) - \log\phi$. This offers a micro-foundation for condition (1) in our abstract framework. The dependence of Y on τ captures the distortionary effect of taxation on labor supply, whereas its dependence on K captures the use of capital in production.

Consider next the investment decisions of the entrepreneurs at $t = 1$. Optimality requires that the marginal cost of investment is equated with its expected marginal revenue, *net* of taxes. After some manipulation, this condition can be expressed as follows:

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

where $\gamma \equiv \alpha - \frac{1-\eta}{\rho}\left(\alpha - (1+\phi)^{-1}\right)$. This offers a micro-foundation of condition (1). The dependence of k_i on $\mathbb{E}_i[\tau]$ captures the distortionary effect of taxation on investment, whereas the dependence on $\mathbb{E}_i[Y]$ captures the combination of two GE effects. On the one hand, because of the aggregate-demand externality embedded in the production of the final good, the marginal revenue of an entrepreneur tends to increase with aggregate output, Y. On the other hand, because a higher Y raises the demand for labor, which in turn raises the real wage and depresses the return to capital, individual investment may decrease with the expected Y. The conflict between these two GE effects explains why, in this context, the reduced-form parameter γ can be either positive or negative.

In the rest of the paper, we adopt the lens of this micro-foundation under the restriction that aggregate-demand externality is strong enough that $\gamma > 0$.[10] We also note that varying the deep parameter ρ, which controls the aggregate-demand externality, varies the reduced-form parameter γ without varying the reduced-form parameter α. The micro-foundation provided here therefore justifies the practice, followed in the rest of the paper, of interpreting the comparative statics of our abstract model with respect to γ as the implications of varying the strength of the underlying GE feedback.

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

$$L = (1-\chi)\left(\tau - \tau^{\text{fb}}\right)^2 + \chi\left(Y - Y^{\text{fb}}\right)^2$$

where $\chi \in (0,1)$ is a fixed scalar and τ^{fb} and Y^{fb} are random variables that represent the policy-maker's "ideal" or first-best values of, respectively, the policy instrument and the outcome. The

[10] The alternative possibility, in which the combined GE effect turns negative ($\gamma < 0$), is discussed in Appendix D. Although some of the results and intuitions have to be modified in this case, the take-home message remains largely the same.

micro-foundations of this objective are left outside our analysis. For our purposes, the key is that the randomness of the pair $(\tau^{\text{fb}}, Y^{\text{fb}})$ introduces random variation in the policymaker's choices.

We next let

$$\tau^{\text{fb}} = Y^{\text{fb}} = \theta \tag{3}$$

for a single random variable θ, which has an unconditional mean normalized to 0, and rewrite the policymaker's objective as

$$L = L(\tau, Y, \theta) \equiv (1-\chi)(\tau - \theta)^2 + \chi(Y - \theta)^2. \tag{4}$$

The restriction imposed in (3) is not needed for our main results regarding the bite of bounded rationality on the implementability constraint faced by the policymaker, but sharpens the normative exercise by guaranteeing that policymaker can attain her first best (zero loss) in the frictionless, REE benchmark studied in the next section. In other words, bounded rationality is the only source of welfare distortion.

Throughout, we assume that θ is observed by the policymaker but unobserved by the private agents. This informational asymmetry does not serve the role of introducing incentive problems. Furthermore, because θ does not enter conditions (1) and (2), the agents do not care to know θ *per se*. Instead, they only care to know what the policymaker intends to do and how this may affect the behavior of others. 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 and why the policymaker has to engage in some form of communication.[11]

This point also explains why the precise stochastic properties of θ do not play a central role in our analysis. Rather, what matters is what the policymaker says and what the agents forecast, or reason, about the responses of others. These aspects of our model are described later on.

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

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

1. Each agent makes his investment choice.

2. The realized 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 $Y = \hat{Y}$.

[11] The assumption that θ does not enter conditions (1) and (2) may be at odds with applications, in which the first best typically depends on fundamentals that have direct effects on the agents' behavior, such as technology. Put differently, a literal interpretation of our model seems to require that θ is a pure externality. However, we view this assumption only as a simplification that allows us to disentangle the communication of policy commitments, and the associated regulation of the agents' strategic interaction, from the revelation of information about fundamentals that affect the agents' behavior while holding constant their strategic interaction. Our paper is interested in the first mechanism; the second is the subject of a large literature that follows the lead of Morris and Shin (2002).

This structure, which is common knowledge, embeds the assumption that the policymaker has full commitment: what the policymaker does in stage 3 coincides with what he communicates in stage 1. This structure also restricts the policymaker's strategy space to the two types of commitment we are interested in. More sophisticated strategies, such as the announcement of a rule that specifies a linear relation between τ and Y, seem less practical, especially if the intended audience is the general public rather than financial traders. They are nevertheless considered in an extension (Section 10) and are shown to preserve the essence of our insights.

The question of interest, and what's next. The question mentioned in the beginning of the Introduction can now be formulated as follows: which communication strategy minimizes L? The following remarks help understand the precise meaning of this question.

In our context, different communications are equated to different commitments: in what follows, "instrument communication" can be reread as "forward guidance in the form of a commitment to a value for τ" and, similarly, "target communication" can be reread as "forward guidance in the form of a commitment to a target for Y." However, unlike Atkeson, Chari and Kehoe (2007) and the related literature on time inconsistency, the choice between these two alternatives has nothing to do with incentive constraints faced by the policymaker. This choice is also not aimed at inducing different responses to different fundamentals, which is the theme of Poole (1970) and the pertinent literature on Taylor rules. Instead, as we will make clear, this choice only matters in regulating the equilibrium bite of the allowed departures from rational expectations.

Thus, in the sequel we first show that, in our model, the communication strategy is entirely irrelevant under rational expectations: whether the policymaker announces, and commits to, a value for τ or a value for Y does not affect either the pairs of (τ, Y) that can be implemented as part of an REE or the associated expectations. We next demonstrate that this is no longer true when we move away from that benchmark: in the presence of bounded rationality, what the policymaker says ends up influencing how the agents think and what they actually do.

4 REE Benchmark

In this section, we explain why communication is irrelevant in a benchmark that, like the standard Ramsey paradigm, imposes Rational Expectations Equilibrium (REE). We also pave the way to the analysis in the rest of the paper by decomposing this benchmark to two separate assumptions: one regarding the agents' own knowledge and rationality, and another regarding their higher-order beliefs.

The first assumption, which we maintain throughout the paper, is the following.

Assumption 1. *Every agent is rational and attentive in the following sense: every agent 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 is has full commitment and acts so as to minimize (4), and perfectly observes any message sent by the policymaker.*

This assumption puts non-trivial discipline on subjective beliefs. In particular, an agent's subjective beliefs about the triplet (τ, Y, K) must consistent with condition (2). Furthermore, because the policymaker has full commitment and the agents know it, the following restriction on subjective beliefs is also imposed: if the policymaker says "I will set a value $\hat{\tau}$ for the instrument," an agent expects τ to equal $\hat{\tau}$ with probability one; and if the policymaker says "I will target a value $\hat{Y}$ for the outcome," an agent expects Y to equal $\hat{Y}$ with probability one.

This discipline, however, is not enough to pin down the agents' subjective beliefs or to equate them with rational expectations. This is because Assumption 1 leaves open the possibility that agents doubt the rationality and/or the attentiveness of *others* and, as a result, form possibly arbitrary beliefs about how others will respond to any given policy message. Such doubts are ruled out in the standard policy paradigm, but are at the center of the forms of bounded rationality considered in this paper. For our purposes, this paradigm is therefore captured by the addition of the following, stronger assumption.

Assumption 2. *It is common knowledge that all agents are rational and are attentive (in the sense of Assumption 1).*

Assumption 2 implies that agents can reason, with full confidence and no mistake, that the restrictions imposed by Assumption 1 extends from their own beliefs to the beliefs of others, to the belief of others about the beliefs of others, and so on, ad infinitum. It is such *boundless* knowledge and rationality that our frictionless benchmark encapsulates by imposing Assumption 2. In our model, this is equivalent to imposing complete-information Nash equilibrium, or REE, as the solution concept.[12] In the rest of this section, we explain why this also trivializes the answer to the question of interest.

We start by defining the sets of the combinations of the policy instrument, τ, and the outcome, Y, that can be implemented under each communication strategy.

Definition 1. A pair (τ, Y) is implementable under instrument (respectively, target) communication if there is an announcement $\hat{\tau}$ (respectively, $\hat{Y}$) and a profile of beliefs and actions for the agents such that condition (1) and Assumptions 1 and 2 are satisfied.

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 problem can then be expressed as follows:

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

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.

[12] This is an implication of the following properties. First, the combination of Assumptions 1 and 2 rules out any combination of subjective beliefs and actions that fall outside the set of complete-information rationalizable outcomes. Second, the set of rationalizable outcomes always contains the set of Nash equilibria. And third, at least under the restriction that $\alpha < 1/2$, our environment admits a unique rationalizable outcome under both types of communication. Without this restriction, Assumptions 1 and 2 may fail to pin down a unique rationalizable, but the REE remains unique and continues to be characterized by the results obtained in this section.

As already noted, imposing Assumptions 1 and 2 is effectively the same as imposing REE. Along with the absence of any heterogeneity, this means that our benchmark admits a representative agent whose optimal behavior is described by the following condition:

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

where $\mathbb{E}[.]$ is the rational expectations conditional on the information of the representative agent. This condition follows from condition (2) after replacing the subjective beliefs with the rational expectation and aggregating. 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

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

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

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

These facts are true regardless of the mode of communication. With instrument communication, we also have $\mathbb{E}[\tau] = \tau = \hat{\tau}$. 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 $\mathbb{E}[Y] = Y = \hat{Y}$. 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 2. $\mathcal{A}^*_\tau = \mathcal{A}^*_Y = \mathcal{A}^* \equiv \{(\tau, Y) : \tau = Y\}$.

That $\mathcal{A}^*$ is a linear locus with slope 1 is a simplifying feature of our environment. More generally, one should think of $\mathcal{A}^*$ as the graph of a monotone mapping between the policy instrument, τ, and the equilibrium outcome, Y. The result then states that this mapping is invariant to whether the policymaker announces and commits to a value for the policy instrument or a target for the outcome.[13] Clearly, the same invariance applies to the expectations that support the implementable (τ, Y) pairs in each case. There is therefore no room for managing either expectations or actual behavior via the mode of communication.

The following result is then immediate:

Proposition 3. *The policymaker attains her first best* $(L = 0)$ *by announcing* $\hat{\tau} = \theta$*, as well as by announcing* $\hat{Y} = \theta$*. The optimal mode of communication is therefore indeterminate.*

In fact, the first best is attained even if the policymaker only announces θ itself, as opposed to announcing the policy or the targeted outcome. 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.

[13] As mentioned in the Introduction, this result mirrors the equivalence of the "dual" and "primal" approaches in the Ramsey literature. In our setting, $\mathcal{A}^*_\tau$ corresponds to the primal problem and $\mathcal{A}^*_Y$ corresponds to the dual.

These findings suggest where we are heading next. The rest of the paper is devoted on understanding how this kind of flawless reasoning breaks apart and how the mode of the policymaker's communication and commitment becomes relevant once we relax the strong assumptions made by our REE benchmark about agents' rationality and depth of knowledge.[14]

5 Beyond REE

Any departure from rational expectations has to be done in a structured way, or else "anything goes." The structure adopted in this paper is to maintain Assumption 1 and only relax Assumption 2 in the manners detailed in the subsequent sections. As already noted, this approach aims at isolating any mistakes agents make when trying to predict or reason about the responses of others to policy—or, equivalently, when trying to calculate the GE effects of the policy.[15]

With this goal in mind, this section develops two insights that hold true for *any* possible relaxation of Assumption 2. The first is that the mode of communication controls the agents' strategic interaction and, thereby, the equilibrium impact of *any* distortion in their expectations or reasoning about the behavior of others and the GE effects of the policy. The second is that such expectations or reasoning can be mapped to higher-order beliefs. Under this lens, our REE benchmark represents a tight restriction on higher-order beliefs and the forms of bounded rationality, or the belief distortions, studied in the rest of the paper can all be understood as relaxations of this restriction.

5.1 Instrument communication

Consider first the case in which the policymaker announces, and commits on, a value $\hat{\tau}$ for the instrument. By Assumption 1, every agent is attentive to the announcement and believes that $\tau = \hat{\tau}$ with probability one. It follows that, for every i, $\mathbb{E}_i[\tau] = \hat{\tau}$ and the best-response condition (2) reduces to

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

This makes clear that, when the policymaker fixes and announces the value of τ, agents only need to form expectations of Y. The question then is *how* these expectations are formed.

Because every agent understands the validity of condition (1), his beliefs about τ, Y, and K must be consistent with it. Taking the expectation of each side of this condition and using the fact that

[14] Before proceeding, let us make the following clarification about the results of this section. The fact that the policymaker can attain $L = 0$ under rational expectations hinges on the property that $(\tau^{\text{fb}}, Y^{\text{fb}}) \in \mathcal{A}^*$ for every θ. If we relax this assumption, $L = 0$ is no more attainable. It nevertheless remains true that communication is irrelevant under rational expectations. This is because Proposition 2, which characterizes implementability, does not depend at all on the assumptions made regarding the policymaker's objective.

[15] By contrast, relaxing Assumption 1 shifts the focus to other possibilities, such as letting agents question the credibility of the policymaker, hold incoherent beliefs about the triplet (τ, Y, K), or make optimization mistakes conditional on their beliefs. These possibilities may be interesting on their own right, but are not the theme of our paper.

$\mathbb{E}_i[\tau] = \hat{\tau}$, we thus get the following restriction on beliefs:

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

This makes clear that forming expectations about the response of Y to the policy announcement is the same as forming expectations about the responses of others, or the response of K.

Combining the above two conditions, we reach the following result.

Lemma 4. *Let $\delta_\tau \equiv \alpha\gamma \in (0,1)$. When the policymaker announces and commits to a value $\hat{\tau}$ for the instrument, agents play a game of strategic complementarity in which best responses are given by*

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

Note that the intercept of the best responses (or the payoff-relevant "fundamental") 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 tax relative to its PE effect. [16]

5.2 Target communication

Consider now the scenario in which the policymaker announces, and commits on, a target $\hat{Y}$ for the outcome. Such an announcement is credible for any value of $\hat{Y}$ because, regardless of the realized K, the policymaker can always adjust τ so as to attain $Y = \hat{Y}$.

By Assumption 1 we now have that, for all i, $\mathbb{E}_i[Y] = \hat{Y}$ and therefore condition (2) reduces to

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

This makes clear that, under target communication, agents are confident that aggregate output will equal $\hat{Y}$ but need to form expectations of the tax that will support this target.

As noted before, the subjective beliefs about τ, Y, and K must be consistent with condition (1). In the previous case (instrument communication), this requirement helped characterize the expectations of Y. In the present case (target communication), it helps characterize the expectations of τ. Indeed, by taking the expectation of both sides of (1), using $\mathbb{E}_i[Y] = \hat{Y}$, and solving for $\mathbb{E}_i[\tau]$, we get

$$\mathbb{E}_i[\tau] = \tfrac{1}{1-\alpha}\hat{Y} - \tfrac{1}{1-\alpha}\mathbb{E}_i[K].$$

[16] The game obtained above is similar to the beauty-contest games studied in, *inter alia*, Morris and Shin (2002), Woodford (2003*a*), 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 engineer higher-order uncertainty by letting the fundamental be observed with noise, here $\hat{\tau}$ is perfectly observed. Despite this fact, higher-order beliefs turn out to play an equally central role in our context because of the relaxation of Assumption 2.

Following any announcement $\hat{Y}$, an agent who expects a lower K therefore also expects a higher τ. Intuitively, 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 the above two conditions, we reach the following counterpart to Lemma 4.

Lemma 5. *Let* $\delta_Y \equiv -\frac{\alpha}{1-\alpha}(1-\gamma) < 0$. *When the policymaker announces and commits to a target* $\hat{Y}$ *for the outcome, agents play a game of strategic substitutability in which best responses are given by*

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

This game is similar to that obtained in Lemma 4 in the following respect: in both cases, the policymaker's announcement controls the intercept of the best responses. This captures the *direct* control that the policymaker has on the incentives of an individual, regardless of the mode of communication and the agents' beliefs about one another's behavior (or, equivalently, the perceived GE effect).

The two games are nevertheless different in the following key respect: whereas the game obtained in Lemma 4 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 higher K maps to lower τ and hence to lower returns for given Y.

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

Corollary 6. *Switching from instrument communication to target communication changes the game played by the agents from one of strategic complementarity to one of strategic substitutability.*

This exact form of this result depends on the assumption that $\gamma \in (0,1)$. If instead we had allowed $\gamma < 0$, both games display strategic substitutability (i.e., both δ_τ and δ_Y are negative when $\gamma < 0$). It remains true, however, that the mode of communication changes the strategic interaction (i.e., δ_τ and δ_Y are different even when $\gamma < 0$). Furthermore, as explained in Appendix D, letting $\gamma < 0$ complicates the exposition but does not upset our take-home message. We thus continue to maintain the assumption $\gamma > 0$ in the main text.

5.3 The role of rational expectations, or Assumption 2

The results developed above (Lemmas 4 and 5 and Corollary 6) are valid in our REE benchmark, for they follow directly from Assumption 1. But they turn out to be irrelevant because of the benchmark's stronger assumption regarding agents' depth of knowledge and rationality, namely Assumption 2. We next explain how this assumption imposes a tight restriction on higher-order beliefs, which in turn drives the irrelevance of the aforementioned results and the irrelevance of the mode of communication. In so doing, we also lay the foundations for what comes next: all the forms of bounded rationality, or belief friction, considered in the rest of the paper can be understood as relaxations of this restriction;

and Lemmas 4 and 5 then help understand how the mode of communication regulates the impact of the allowed friction on actual outcomes.

With $X \in \{\tau, Y\}$ indexing the mode of communication, the best responses obtained in Lemmas 4 and 5 are nested in the following form:

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

Recall that the sign of δ_X depends on the mode of communication. Regardless of this, however, common knowledge of rationality—which is one half of Assumption 2—implies that every agent can aggregate and iterate condition (9) to obtain her forecast of the aggregate action as follows:

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

where $\bar{\mathbb{E}}^h[\cdot]$ denotes the h-th order average forecast. This is defined recursively by letting $\bar{\mathbb{E}}^1[\cdot] \equiv \int \mathbb{E}_i[\cdot]\mathrm{d}i$ and $\bar{\mathbb{E}}^h[\cdot] \equiv \bar{\mathbb{E}}[\bar{\mathbb{E}}^{h-1}[\cdot]]$ for all $h \geq 2$.[17]

Consider how an agent's expectation of the behavior of others, $\mathbb{E}_i[K]$, varies with the message $\hat{X}$. This captures the agent's perception of the GE implications of the policy communications. Condition (10) allows us to represent this perceived GE effect as a function of the higher-order beliefs about $\hat{X}$. This explains why, at least for our purposes, imposing a structure on higher-order beliefs is synonymous to imposing a structure on how agents form expectations or reason about the GE effects of the policy. We next detail the specific, and rather tight, structure imposed by Assumption 1.

Assuming that every agent knows that *other* agents are attentive guarantees that

$$\mathbb{E}_i[\mathbb{E}_j[X]] = \mathbb{E}_i[X] = \hat{X},$$

for every i and every $j \neq i$. That is, the announcement moves by exactly the same amount the agents' own forecasts of X and their forecasts of the forecasts of others. By induction, common knowledge of attentiveness—which is the second half of Assumption 1, the first being common knowledge of rationality—imposes that the same is true for all higher-order forecasts. That is,

$$\bar{\mathbb{E}}^h[X] = \hat{X} \quad \forall k \geq 1. \tag{11}$$

[17]Condition (10) and the argument given in the rest of this section, which depends on this condition, requires $|\delta_X| < 1$. Had we allowed α to take any value in $(0,1)$, instead of restricting $\alpha \in (0, 1/2)$, we would have maintained $\delta_\tau \in (0,1)$, and hence $|\delta_\tau| < 1$, but we would have opened the door to the possibility that $\delta_Y \leq -1$, and hence $|\delta_Y| \geq 1$. When this happens, the method of iterated best responses fails to converge. This underscores that, whenever $\delta_Y \leq -1$, the game induced by target communication does not have a unique rationalizable outcome. Nevertheless, except for the degenerate case in which $\delta_Y = -1$, this game continues to admit a unique REE, and this equilibrium is such that $K = \hat{X}$. The analysis that follows is therefore robust to the possibility $\delta_Y < -1$ insofar as we focus on the unique REE. Restricting $\alpha < 1/2$ only carries the extra benefit of guaranteeing that this REE is "globally stable" in the sense of being obtainable as the limit of iterating on best responses. A similar point applies the cases with bounded rationality studied in subsequent sections.

Next, substituting (11) into (10) gives $\mathbb{E}_i[K] = \hat{X}$. By (9), we then also have

$$k_i = K = \hat{X} = \mathbb{E}_i[K].$$

This makes clear that the key role played by Assumption 2 is to let every agent be confident that the every other agent will react to the announcement in exactly the same way as himself.

What does this mean for the set of implementable policy-and-outcome pairs? With instrument communication, we have $\tau = \hat{\tau}$ and, from the above result, $K = \hat{\tau}$; condition (1) then gives $Y = \hat{\tau}$; and because $\hat{\tau}$ can be any number, we infer that a pair (τ, Y) is implementable under instrument communication if and only if $\tau = Y$. With target communication, on the other hand, we have $Y = \hat{Y}$ and, form the above result, $K = \hat{Y}$; condition (1) then implies $\tau = \hat{Y}$; and because $\hat{Y}$ can be any number, we infer that a pair (τ, Y) is implementable under target communication if and only if $\tau = Y$. Combining these observations proves that, in our frictionless benchmark, the set of implementable pairs is invariant to the mode of communication.

Although this invariance property has already been stated in Proposition 2, the more detailed argument given here spells out exactly *how* this property depends on the strong assumptions our REE benchmark makes about agents' depth of knowledge and rationality, as captured by Assumption 2. In addition, by representing the expectations, or the reasoning, about the behavior of others in terms of higher-order beliefs and by revealing the tight structure that Assumption 2 imposes on higher-order beliefs, the above argument lays the common foundation of the subsequent analysis: all the forms of bounded rationality considered in our paper can be understood as relaxations of this tight structure.

6 Anchored Beliefs

We now turn to the core of our contribution, which is to characterize the optimal strategy for managing expectations in the presence of a particular form of bounded rationality—one that anchors the beliefs about the responses of others to the policy announcement. As noted in the Introduction, the kind of anchored beliefs we are after seems consistent with both survey and experimental evidence. One arm of the literature attempts to rationalize such anchoring as the by-product of dispersed private information; another interprets it as the symptom of bounded rationality.[18] In this paper, we adopt the latter interpretation and capture the friction by replacing Assumption 2 with the following.

Assumption 3 (Anchored beliefs)**.** *Every agent believes that all other agents are rational but only a fraction $\lambda \in [0,1]$ of them is attentive. In particular, 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.*

[18] The available evidence does not necessarily differentiate among these interpretations, but clearly supports the kind of anchored beliefs we are after. Furthermore, as discussed at the end of Section 10, our main insights extend from the one interpretation to the other.

Relative to Assumption 2, Assumption 3 maintains common knowledge of rationality but drops common knowledge of attentiveness. This amounts, essentially, to changing the solution concept from REE to Perfect Bayesian Equilibrium with heterogeneous priors regarding the information of others. To see this more clearly, just recast the above assumption in the following terms. First, let each agent i receive a private signal s_i of the announcement. And second, specify her prior about the joint distribution of $(s_i, \{s_j\}_{j \neq i}, \hat{X})$ as follows: s_i is a drawn from a Dirac measure at $\hat{X}$ with probability 1; and 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 $1 - \lambda$.[19]

A similar heterogeneous-prior specification has been used in Angeletos and La'O (2009) to introduce belief inertia in the New Keynesian model. As already noted, this specification is grounded on a literature that studies the role of higher-order uncertainty in common-prior settings. The heterogeneous-prior approach, not only affords a higher level of tractability, but also allows an extra degree of freedom: the degree of the belief anchoring is not necessarily tied to the cross-sectional dispersion in forecasts. This extra degree of freedom represents a departure from rational expectations—which, for the present purposes, is a feature, not a bug.

As explained in Section 7, Level-k Thinking produces similar results as this specification—in effect by relaxing the part of Assumption 2 that pertains to common knowledge of rationality, as opposed to the part that pertains to the common knowledge of attentiveness. 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.

There is, however, a subtle difference. As it will become clear shortly, the present specification allows the scalar λ to parameterize the degree of belief anchoring in a continuous and monotone manner: at the one extreme, $\lambda = 0$ identifies a situation in which agents expect the others not to respond at all; at the other extreme, $\lambda = 1$ nests our frictionless, REE benchmark, in which the expected response of others is maximal; in between, the expected response varies is continuous and monotone in the agents' depth of knowledge, as measured by λ. By contrast, as explained in Section 7, Level-k Thinking introduces a discontinuous and non-monotone pattern, which we find unappealing. This is the main reason why, even though we embrace the *spirit* of Level-k Thinking, we prefer on the margin the modeling approach taken here.[20]

[19] Note that priors are heterogeneous because, for any pair (i, j) such that $j \neq i$, agents i and j disagree about the distribution of the pair (s_i, s_j). Also note that the equivalence between the heterogeneous-prior PBE and Assumption 3, just like that between REE and Assumption 2, depends on the restriction $|\delta_X| < 1$.

[20] Another reason is the aforementioned connection between the approach taken here and the literature that studies the effects of higher-order uncertainty in common-prior, rational-expectation settings. This connection helps explain why appropriate adaptations of our main insights apply to such settings as well.

6.1 Beliefs and behavior

Fix a communication mode $X \in \{\tau, Y\}$ and consider the game played among the agents. As already noted, individual best responses are given by

$$k_i = (1 - \delta_X)\mathbb{E}_i[X] + \delta_X \mathbb{E}_i[K]. \tag{12}$$

Furthermore, because we have maintained common knowledge of rationality, it remains true that, as long as $|\delta_X| < 1$,[21] the beliefs about K satisfy the following restriction:

$$\mathbb{E}_i[K] = \mathbb{E}_i \left[(1 - \delta_X) \sum_{h=1}^{\infty} (\delta_X)^{h-1} \bar{\mathbb{E}}^h[X] \right]. \tag{13}$$

As in the frictionless benchmark, this follows from iterating the best responses of the agents and gives the optimal investment of an agent as a function of his first- and higher-order beliefs about X.

What is different from that benchmark is the structure of higher-order beliefs. Because every agent believes that only a fraction λ of the other agents is attentive, 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 $h \geq 1$, the $(h+1)$-th order beliefs satisfy

$$\mathbb{E}_i \left[\bar{\mathbb{E}}^h[X] \right] = \mathbb{E}_i \left[\mathbb{E}_{j_1} \left[... \mathbb{E}_{j_h}[X] \right] \right] = \lambda^h \hat{X}. \tag{14}$$

Comparing the above to their frictionless counterparts, we see that higher-order beliefs are tilted towards zero, and the more so the higher their order. This is the hallmark of the introduced friction: whereas the fact that each agent is *himself* attentive guarantees that the announcement has maximal effect on first-order beliefs, the perception that others may be inattentive arrests the impact of the announcement on the higher-order beliefs of X and, thereby, on the beliefs of K.

Indeed, by substituting (14) into (13), we reach the following characterization of the beliefs about the responses of others.[22]

Lemma 7. *The typical agent's expectation of K following announcement $\hat{X}$ is given by*

$$\mathbb{E}_i[K] = \frac{\lambda - \lambda \delta_X}{1 - \lambda \delta_X} \hat{X}, \tag{15}$$

where $X \in \{\tau, Y\}$ depending on the mode of communication.

[21] The role of this condition was discussed in footnote 17: it guarantees that the game admits a unique rationalizable outcome, and that the associated beliefs are described by condition (13), but is not strictly needed for our main results.

[22] Following up the discussion in footnotes 17 and 21, note that the characterization of beliefs provided in Lemma 7 and all the other results in this section are robust to $\delta_Y < -1$, provided that we concentrate on the unique linear PBE. Adding the restriction $\alpha < 1/2$ guarantees that this PBE is the unique rationalizable outcome for any $\gamma \in [0, 1]$ and any $\lambda \in [0, 1]$.

From Lemmas 4 and 5, we have that instrument communication corresponds to $\delta_X \in (0,1)$, whereas target communication corresponds to $\delta_X < 0$. In either case, however, the responsiveness of $\mathbb{E}_i[K]$ to $\hat{X}$ is bounded between 0 and 1 :

$$0 \leq \frac{\lambda - \lambda\delta_X}{1 - \lambda\delta_X} \leq 1.$$

Furthermore, the above ratio is strictly increasing in λ. And since $\lambda = 1$ nests the frictionless benchmark, the following is true.

Corollary 8. *Regardless of the communication mode, a lower λ reduces each agent's expectation of the response K to the announced $\hat{X}$.*

This result justifies the interpretation of Assumption 3 as anchored beliefs and verifies our earlier claim that the equilibrium degree of belief anchoring is continuous and monotone in λ. It also underscores that the friction is qualitatively the same between the two modes of communication. In this sense, the replacement of Assumption 2 with Assumption 3 does not a priori favor any one mode of communication.

As shown next, however, the impact of the friction on actual behavior and on the set of implementable outcomes is qualitatively different between the two modes of communication. This is because the mode of communication determines the nature of the strategic interaction and the sign of the impact of the beliefs of K on the actual K.

Indeed, replacing (15) in (12) and aggregating across agents, we reach the following result.

Lemma 9. *The realized aggregate investment following announcement $\hat{X}$ is given by*

$$K = \frac{1 - \delta_X}{1 - \lambda\delta_X}\hat{X}, \tag{16}$$

where $X \in \{\tau, Y\}$ depending on the mode of communication.

Recall that the frictionless benchmark had $K = \hat{X}$. 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. The following is therefore true.

Corollary 10. *Letting $\lambda < 1$ attenuates the response of K under instrument communication, and amplifies it under target communication. Furthermore, a larger departure from the frictionless benchmark (lower λ) translates to larger attenuation in the first case and to larger amplification in the second case.*

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

6.2 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 (16) 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 Y is itself pegged at $\hat{Y}$. Condition (16) 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 observations, 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 11 (Implementation with anchored beliefs)**.** *For any $\lambda \in [0,1]$, let $\mathcal{A}_\tau(\lambda)$ and $\mathcal{A}_Y(\lambda)$ denote the sets of the pairs (τ, Y) that are implementable under, respectively, instrument communication and target communication. Then,*

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

where

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

The frictionless benchmark is nested by $\lambda = 1$ and results in $\mu_\tau(1) = 1 = \mu_Y(1)$ and $\mathcal{A}_\tau(1) = \mathcal{A}_Y(1)$. By contrast, for any $\lambda \in [0,1)$, we have $\mu_\tau(1) \neq \mu_Y(1)$ and therefore $\mathcal{A}_\tau(\lambda) \neq A_y(\lambda)$. That is, the two implementable sets cease to coincide as soon as we move away from the frictionless benchmark.

The next proposition offers a sharper characterization of how $\mu_\tau(\lambda)$ and $\mu_Y(\lambda)$, the slopes of the two implementability constraints, compare to one another, as well as to the frictionless counterpart.

Proposition 12 (Slope of budget lines)**.** *Suppose $\gamma \in [0,1]$ and $\lambda \in [0,1]$. Then,*

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

2. *$\mu_Y(\lambda) \leq 1$ with equality only when $\lambda = 1$ or $\gamma = 1$*

3. *For fixed (α, γ), $\mu_\tau(\lambda)$ increases in λ and $\mu_Y(\lambda)$ decreases in λ.*

Proof. See Appendix A.1.1. □

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. Figure 1 illustrates this property.[23]

The intuition behind Proposition 12 is best illustrated in the extreme case in which $\lambda = 0$, that is, when the expectations of K are completely unresponsive to policy communication.

First, consider instrument communication: let the government announce and commit to some $\tau = \hat{\tau} > 0$. Because the agents play a game of strategic complementarity (4), the expectation that others will not react causes each agent to react *less* than in the frictionless benchmark. In particular, the realized aggregate investment is $K = (1 - \alpha^2\gamma)\hat{\tau}$, a fraction of the frictionless case $K = \hat{\tau}$. Since a lower K maps to a lower Y for any given τ, the realized output is similarly a fraction of the frictionless counterpart. A policymaker with a fixed output target thus needs a more aggressive policy (higher $\hat{\tau}$) in the anchored case relative to the frictionless one. Thus $\mu_\tau(0) > \mu_\tau^* = 1$.

Next, consider target communication: let the government announce and commit to some $Y = \hat{Y} > 0$. Because agents now play a game of strategic substitutability (5), the expectation that others will not react causes each agent to invest *more* than in the frictionless benchmark. And because a higher K always maps to a higher Y for any given τ, the policymaker can now achieve any given increase in Y with a smaller increase in τ than in the frictionless world. Thus $\mu_Y(0) < \mu_Y^* = 1$.

The same logic applies also for any $\lambda \in (0, 1)$, except that the anchoring of the expectations of K is then weaker and, by the same token, the effect on the implementability sets is less acute. This claim is made precise in the last part of Proposition 11 and is illustrated in Figure 1. The dotted line in this figure corresponds to $\mathcal{A}^*$, the implementability set under rational expectations ($\lambda = 1$). The blue and red solid lines correspond to the bounded rationality implementability sets under instrument and target communication, respectively.

6.3 Role of GE feedback

Let us now turn attention to the role played by the GE feedback parameter γ. Recall that γ is directly connected to the strength of the aggregate demand externality in the micro-founded example of Section 3. In an application to forward guidance that we discuss in Section 8, it corresponds to the Keynesian income-spending multiplier. More generally, we think of γ as a proxy for a variety of GE feedbacks that regulate the effectiveness of macroeconomic policy.

[23] The figure and the discussion in the main text presumes $\mu_Y > 0$, which is necessarily the case under the maintained assumption that $\alpha < 1/2$. Otherwise, $\mu_Y < 0$ is possible. This is a somewhat perverse case in which the economy would overshoot the target if τ were kept at 0 and, therefore, the policy instrument must be used ex post to "cool off" the economy. More precisely, strategic substitutability is so high that αK increases more than one-for-one with the announcement $\hat{Y}$. Proposition 21 in Appendix A.1.2 states the parameter values for which this possibility emerges. The restriction $\alpha < 1/2$ rules out this possibility, essentially for the same reason that it guarantees the "global stability" of the equilibrium explained earlier. Our take-home message, though, is not disrupted by allowing $\alpha > 1/2$ or $\mu_Y < 0$.

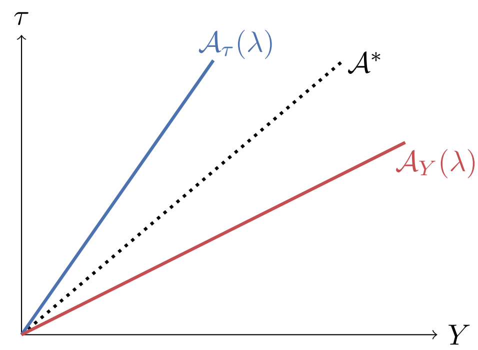

Figure 1: Implementable sets $\mathcal{A}_\tau(\lambda)$ and $\mathcal{A}_Y(\lambda)$ with bounded rationality ($\lambda < 1$).

With this in mind, the next proposition sheds light on how such GE feedbacks also regulate the impact of the introduced friction on the implementability constraints faced by the policymaker.

Proposition 13. *Fix any $\lambda \in (0,1)$ and $\alpha \in (0,1)$. As γ increases, both $\mu_\tau(\lambda)$ and $\mu_Y(\lambda)$ increase. Furthermore, $\mu_\tau(\lambda) \to \mu_{\tau,1} > 1$ and $\mu_Y(\lambda) \to \mu_Y^* = 1$ as $\gamma \to 1$, whereas $\mu_\tau(\lambda) \to \mu_\tau^* = 1$ and $\mu_Y(\lambda) \to \mu_{y,0} < 1$ as $\gamma \to 0$.*

Proof. See Appendix A.2. □

As the GE effects gets stronger (γ increases), the impact of the friction is therefore *exacerbated* under instrument communication, in the sense that $\mu_\tau(\lambda)$ gets further away from μ_τ^*, whereas its impact is *alleviated* under target communication, in the sense that $\mu_Y(\lambda)$ gets closer to μ_Y^*. These properties will prove instrumental for our characterization of the optimal communication strategy, which follows in the next subsection. The logic behind them 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 partial-equilibrium, 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 $k_i = \hat{\tau}$ for all i,and therefore also that $K = \hat{\tau}$, regardless of λ. Condition (1) then gives $Y = \hat{\tau}$, also regardless of λ. The policymaker can thus implement the first-best (frictionless) set of policy-and-outcome pairs: $\mathcal{A}_\tau(\lambda) = \mathcal{A}_\tau^*$, for all $\lambda < 1$. It is straightforward to verify that this is not the case with target communication: $\mathcal{A}_Y(\lambda) \neq \mathcal{A}_Y^*$, for all $\lambda < 1$. Target communication transforms the game played among the agents from one with a null strategic interaction one with a non-zero strategic substitutability (indeed, note that $\delta_\tau = 0$ but $\delta_y < 0$ when $\gamma = 0$), thus also allowing the belief distortion to enter the implementability restriction.

Consider next the case in which the GE effect is maximal, or $\gamma = 1$. Behavior is pinned down exclusively by expectations of the outcome: $k_i = \mathbb{E}_i Y$ for all i. Announcing and committing on an outcome target $\hat{Y}$ guarantees that aggregate investment is given by $K = \hat{Y}$, the required policy is $\tau = \hat{Y}$ regardless of λ, and thus the implementable set is the same as the undistorted one ($\mathcal{A}_Y(\lambda) = \mathcal{A}_Y^*$). By

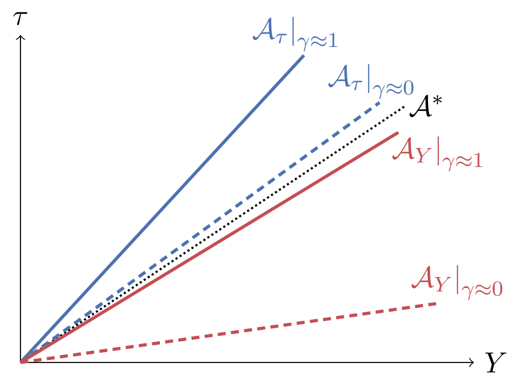

Figure 2: The effect of γ, or the GE feedback, on the implementability constraints.

contrast, instrument communication entails a distortion ($\mathcal{A}_Y(\lambda) \neq \mathcal{A}_Y^*$), because it induces non-zero strategic interaction ($\delta_\tau > 0$).

In the intermediate case in which $\gamma \in (0, 1)$, the friction impacts the implementability constraint under both modes of communication. But the logic extends in the sense that a stronger GE effect increases the effect of the friction on the slope of the implementability restriction between Y and τ under instrument communication, and decreases it under target communication.

These properties are illustrated in Figure 2. This figure is similar to Figure 1, except that now we fix the value for λ and show the implementability constraints in the (τ, Y) space for two different values of γ. As before, the color of a line indicates the mode of communication: blue for instrument communication, red for target communication.

The dashed lines correspond to a weak GE feedback (γ close to zero), the solid lines correspond to a strong GE feedback (γ close to 1). Clearly the the distance from $\mathcal{A}^*$, the frictionless implementability line, is minimized with instrument communication when the GE feedback is weak and with target communication when the GE feedback is strong.

6.4 Optimal communication and the second best

We are now ready to state, and prove, our main result regarding the optimal communication choice, as well as to characterize the optimal pair (τ, Y) that gets implemented with that choice.

Our previous discussion established that, in the extreme cases of $\gamma \in \{0, 1\}$, one mode of communication replicated the undistorted implementable set and the other did not. It follows immediately that the undistorted method (instrument for $\gamma = 0$ and target for $\gamma = 1$) is optimal for these parameters. Each strategy, in its "most favorable" extreme case, sidesteps the friction entirely by eliminating agents' need to forecast, or reason about, others' actions.

What about the intermediate cases in which $\gamma \in (0, 1)$? In this case, neither mode of communication eliminates the need to forecast or reason about the behavior of others. With instrument communication, the agents worry about the behavior of others because they need to predict the outcome;

with target communication, they worry because they need to predict the value of the instrument that will be required to honor the target. The policymaker can no longer bypass the friction and can no longer attain the first best.

The optimal communication strategy is therefore not obvious away from the aforementioned two extremes. Nevertheless, 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 proposition verifies this intuition.

Theorem 14 (GE feedback threshold)**.** *For any $\lambda < 1$, there exists a threshold $\hat{\gamma} \in (0,1)$ such that, for $\gamma \in (\hat{\gamma}, 1]$, target communication is strictly optimal given any realization of θ; and for $\gamma \in [0, \hat{\gamma})$, instrument communication is strictly preferred.*

A detailed proof is provided in Appendix A.3. Below we sketch the main argument and also identify the second best, namely the pair (τ, Y) that gets implemented by the optimal strategy.[24]

For any realization of 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 point 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 11, 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 strategy 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 $Y^{\text{sb}} = (r^{\text{sb}}/\mu^{\text{sb}})\theta$.

[24] Here, we continue to assume $\alpha < 1/2$, which, as explained earlier, eliminates the possibility of $\mu_Y < 0$. The more general case is covered in Appendix A.3.

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

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

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 communication choice. 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.

6.5 Comparative statics

Because the model is analytically tractable, we can characterize the dependence of the optimal communication strategy, not only on γ, but also on all other parameters.

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 the effect of α. As α approaches 1, τ has as vanishingly small effect on Y for given K. The policymaker may therefore need to make very large adjustments in τ to hit any given outcome target. This suggests that target communication becomes less desirable as α increases. We verify this property in Appendix A.6.

Finally, and perhaps more interestingly, consider the effect of λ. For any given (α, γ), raising the belief friction (lowering λ) intensifies the distortion under both modes of communication. As shown next, however, the additional friction "bites harder" with target communication than under instrument communication:

Proposition 15. *For fixed (α, χ), the threshold $\hat{\gamma}$ is a decreasing function of λ. That is, the range of γ for which target communication is optimal increases as the friction gets smaller.*

Proof. See Appendix A.6. □

Furthermore, as the friction vanishes, the threshold $\hat{\gamma}$ has a well-defined limit given by

$$\lim_{\lambda \uparrow 1} \hat{\gamma} = \frac{1}{2-\alpha} \in \left(\frac{1}{2}, 1\right)$$

Whereas *exact* rational expectations (nested as $\lambda = 1$) leaves optimal communication indeterminate, *near* rational expectations (i.e., λ arbitrarily close to, but strictly lower than, 1) gives a non-trivial result. Put differently, a policymaker with small uncertainty—in either a Bayesian or a Knightian sense—about the parameter λ around the benchmark $\lambda = 1$ may reach qualitatively similar conclusions to the policymaker who is confident that the friction is large.

7 Level-k Thinking

The key mechanism in the previous section is agents' under-forecasting of others' responses to an announcement (Lemma 7). 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 are allowed to 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

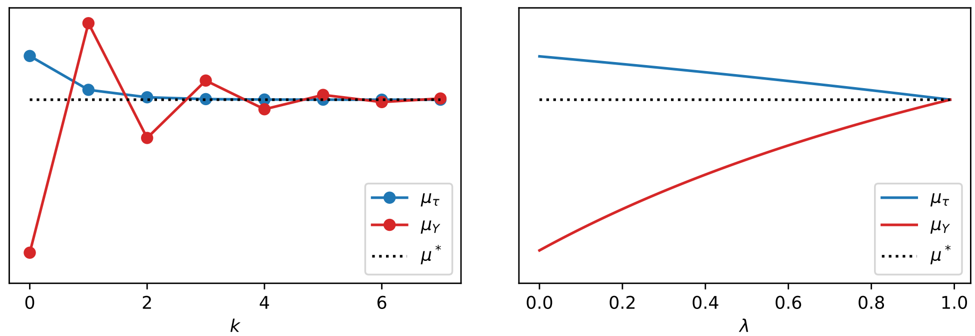

Figure 3: The implementability coefficients (μ_τ, μ_Y) under Level-k Thinking (left) and anchored beliefs (right).

(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.

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) correspond to $K = 0$. Because every agent believes that all other agents are of cognitive order $k-1$, the expectation of K is now given by[25]

$$\bar{\mathbb{E}}[K] = \left((1-\delta_X) \sum_{h=0}^{k-1} \delta_X^h \right) \hat{X} \tag{17}$$

Comparing this expression to (13), which gave expected investment as a function of higher-order beliefs about X, reveals that Level-k Thinking is isomorphic to the following belief hierarchy about the policy announcement:

$$\bar{\mathbb{E}}^h[X] = \hat{X}, \ \ \forall h < k \qquad \text{and} \qquad \bar{\mathbb{E}}^j[X] = 0, \ \ \forall h \geq k$$

That is, it is *as if* agents know that others know that... others have heard the announcement only up to order k;beyond that order, beliefs are pegged at zero.

This is similar to the structure of higher-order beliefs considered in Section 6. Both approaches allow the announcement's effect on the h-th order belief to decay with h. Before, the decay was exponential in h; now it is a step function jumping from 1 to 0 at the specific order $h = k$.

This similarity suggests that the lessons derived earlier extend to Level-k Thinking. Indeed, $k = 1$ corresponds exactly to $\lambda = 0$ in our earlier analysis. Furthermore, for any odd number $k \geq 3$, one can find a $\lambda \in (0, 1)$ such that the implementability sets under Level-k Thinking coincide with those in our earlier analysis (see Appendix C for the exact construction).

The equivalence, however, breaks down for even k because of the concept's "oscillatory" behavior in games of strategic substitutability. Consider the game following target communication. 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

[25]The formula applies for $k \geq 2$; for $k = 1$, $\bar{\mathbb{E}}[K] = 0$.

less than in the frictionless benchmark and thus moves *more*. A level-2 agent expects K to move *more* than in the frictionless benchmark and therefore chooses to move *less* himself. Whereas $k = 1$ amplifies the actual response of investment, $k = 2$ attenuates. The left panel of Figure 3 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 6.

We find this oscillatory, non-monotone pattern to be conceptually unappealing and suspect that it is an unintended consequence of a particular formalization that was developed in the experimental literature for games of complements, but may not be ideal for games of substitutes. Seen from this perspective, our heterogeneous-prior formalization captures the essence of Level-k Thinking while bypassing this "pathological" feature. In Appendix C, we show that the same goal can be achieved with a "smooth" version of Level-k Thinking along the lines of Garcıa-Schmidt and Woodford (2018). With these qualifications in mind, one can see these alternative approaches to bounded rationality as essentially interchangeable.

8 Forward Guidance for Monetary Policy

Consider the question of how aggregate demand responds to forward guidance when the latter takes the form of a unconditional commitment for keeping interest rates low after the economy has exited a liquidity trap. This question has already been addressed by Angeletos and Lian (2018), Garcıa-Schmidt and Woodford (2018), and Farhi and Werning (2016).[26] Our paper inverts the question: should the policymaker engage in this type of forward guidance, or should she instead commit to a credible target for GDP and unemployment?

Even though our framework is too stylized to nest the New Keynesian model, our results suggest that the answer to the above question depends critically on the strength of the underlying GE feedback mechanisms. But what are these mechanisms and what determines their strength?

Three such mechanisms are at work in the context of the baseline New Keynesian model: the positive feedback between aggregate income and aggregate spending, or the Keynesian cross, which underlies the Dynamic IS curve; the dynamic strategic complementarity in the firms' price-setting decisions, which underlies the New Keynesian Philips cure; and the inflation-spending feedback that is captured by the interaction of the Dynamic IS curve and the New Keynesian Philips cure.[27]

All these effects are purely forward-looking: the feedback goes in one direction from expectations of future outcomes to current behavior. Adding capital, habit persistence or commitments in consumption, or wealth effects can introduce an opposite-direction feedback from current behavior to

[26]The first paper models the friction as higher-order uncertainty in a rational-expectations, common-prior setting, the other two model it as Level-k Thinking. In line with our results regarding instrument communication, these papers find that the belief distortion reduces the power of the aforementioned type of forward guidance.

[27]See Angeletos and Lian (2018) for a detailed discussion of these mechanisms and for game-theoretic representations that reveal the connection to the more abstract framework used here.

future outcomes. The combination delivers a two-way interaction similar to that stylized by our abstract framework, with K corresponding to current aggregate spending and Y to economic conditions in the not-so-far future.

Finally, building on the results of Angeletos and Lian (2018), one may expect the combined effect of all these mechanisms, proxied by γ in our framework, to increase with the horizon of forward guidance. The reason is that longer horizons entail longer chains of dynamic feedback effects. The results of Farhi and Werning (2016), on the other hand, suggest that the effective γ may increase with liquidity constraints, insofar as such constraints map to a large income-spending multiplier.

Combining these insights with the results of our paper suggests the following policy implication. Consider a situation in which the liquidity trap is expected to be sufficiently long and/or the Keynesian cross is sufficiently steep. It is precisely in this situation that traditional forward guidance is severely constrained, as argued by the aforementioned papers. But it is also then that the policymaker may bypass the friction by communicating, and committing to, a path for future employment and GDP rather than a path for the policy rate.[28]

We corroborate these intuitions in Appendix B.2, within the context of a stylized New Keynesian economy. We take a few shortcuts in order to keep the analysis tractable and to nest the economy to our abstract framework. These shortcuts do not necessarily drive our results, but preclude a quantitative evaluation or a richer understanding of the determinants of the optimal forward guidance. The micro-foundation of the welfare objective is another task left for future investigation.[29]

9 Erratic Beliefs

The analysis so far has abstracted from the possibility that bounded rationality is the source, not only of belief inertia vis-a-vis the policy communication, but also of random shifts in "market psychology." We now capture this possibility by considering the following, different relaxation of Assumption 2.

Assumption 4 (Erratic beliefs). *Every agent believes that the other agents are rational but worries that a fraction $1-\sigma$ of them receives a randomly distorted message and is unaware of the distortion. In particular, every i believes that, for every $j \neq i$, $\mathbb{E}_j[X] = \hat{X}$ with probability σ and $\mathbb{E}_j[X] = \hat{X} + \varepsilon$ with probability $1-\sigma$, where $\sigma \in (0,1)$ is fixed scalar and ε is a random shock, drawn from a Normal distribution with mean zero and variance one, orthogonal to θ, and unobserved by the policymaker at the moment of his announcement. These facts and the value of σ are common knowledge.*

[28] We let the reader decide whether this is good news for policymakers, in the sense that the appropriate form of communication can indeed contain the friction under consideration, or bad news for the aforementioned papers, in the sense that the forward guidance puzzle has been translated to a different dimension rather than been truly resolved.

[29] Our analysis also abstracts from shocks, or constraints, that may interfere with the policymaker's control of the targeted outcome. As explained in Appendix F, such considerations tilt the balance in favor of instrument communication for any given γ, but do not affect our result that a higher γ (stronger GE effect) increases the relative value of target communication.

To understand what this assumption does, recall our earlier characterization of the expectations of K in terms of the higher-order beliefs of X:

$$\mathbb{E}_i[K] = \mathbb{E}_i\left[(1-\delta_X)\sum_{k=1}^{\infty}(\delta_X)^{k-1}\bar{\mathbb{E}}^k[X]\right] \tag{18}$$

The above continues to apply here, just as it applied in the frictionless benchmark and in the case considered in the previous section. What changes as we move across these three scenarios is how higher-order beliefs relate to first-order beliefs.

In the frictionless benchmark, Assumption 2 forces all higher-order beliefs to collapse to first-order beliefs ($\bar{\mathbb{E}}^k[X] = \hat{X}$), thus giving $\mathbb{E}_i[K] = \hat{X}$. In the scenario studied in Section 6, Assumption 3 allows the higher-order beliefs to move less than one-to-one with first-order beliefs ($\bar{\mathbb{E}}^k[X] = \lambda^{k-1}\hat{X}$), but rules out any orthogonal variation in the gap between first- and higher-order beliefs, thus giving $\mathbb{E}_i[K] = b\hat{X}$ for some $b < 1$. This captures "anchored beliefs" but rules out "erratic beliefs."

The scenario studied now, under Assumption 4, does the opposite: it lets ε drive random variation in higher-order beliefs. Because each agent believes that only a fraction σ of the population heard the actual announcement, whereas the remaining friction heard the distorted message $\hat{X} + \varepsilon$, her second-order belief is now given by

$$\mathbb{E}_i[\bar{\mathbb{E}}[X]] = \sigma\hat{X} + (1-\sigma)(\hat{X}+\varepsilon) = \hat{X} + (1-\sigma)\varepsilon.$$

By induction, the h-th order average belief is given by

$$\bar{\mathbb{E}}^h[X] = \hat{X} + a_h\varepsilon \tag{19}$$

with $a_1 = 0$ and $a_h = \sigma a_{h-1} + (1-\sigma)$ for $h \geq 2$.[30]

Using (19) in (18) yields the following expression of the expectations of K :

$$\mathbb{E}_i[K] = \hat{X} + \frac{1-\sigma}{1-\sigma\delta_X}\varepsilon. \tag{20}$$

from which it becomes evident that ε introduces extrinsic waves of optimism and pessimism about the activity of others.

From this perspective, ε resembles a sunspot. But instead of being the product of multiple equilibria, it is the product of a correlated bias in higher-order beliefs. And while the form adopted here may look exotic, related forms of variation in higher-order beliefs have been used before in the literature to capture the role of confidence in business cycles (Angeletos, Collard and Dellas, 2018*a*; Angeletos and La'O, 2013; Benhabib, Wang and Wen, 2015; Huo and Takayama, 2015) and speculative fluctuations in asset markets (Allen, Morris and Postlewaite, 1993*b*; Scheinkman and Xiong, 2003). Finally, similar belief fluctuations are produced in an extension of the Level-k setting of Section 7 that features a random level-0 point and/or a random cognitive order k. In this sense, ε can also be thought as randomness in how agents reason about the behavior of others and the GE effects of the policy.

[30]Note that a_h increases with h, which means that the impact of the shock increases with the order of beliefs. In the limit, as $h \to \infty$, $a_h \to 1$.

9.1 Implementability

As long as $|\delta_X| < 1$, $\frac{1-\sigma}{1-\sigma\delta_X}$ is positive and increasing in δ_X; and as long as $\alpha \leq 1/2, 1 < \delta_Y < 0 < \delta_\tau < 1$. It then follows that the random waves in the agents' expectations of K are necessarily stronger under instrument communication than under target communication. What matters for implementability, however, is not these belief waves per se, but rather the fluctuations they trigger in actual investment.

Replacing the expectation of K from condition (20) to the best-response condition (12), which is still valid, we get the following characterization of the realized K :

$$K = \hat{X} + \frac{\delta_X(1-\sigma)}{1-\sigma\delta_X}\varepsilon \tag{21}$$

Note that the mode of communication regulates not only the magnitude, but also the sign of the effect of ε on K. With instrument communication, the agents play a game of strategic complementarity. In this case, optimism about the beliefs and the behavior of others feeds to more investment. With target communication, instead, the agents play a game of strategic substitutability. In this case, optimism about the beliefs and the behavior of others feeds to less investment. This echoes the differential effect of anchored beliefs under the two modes of communication, which was documented in Section 6.

Proceeding in a similar manner as in that section, we can map the realized K to the pairs of τ and Y that are induced by any given announcement. We can thus reach the following characterization of the implementability constraints faced by the policymaker.

Proposition 16. *A pair (τ, Y) is implementable if and only if*

$$\tau = Y + \psi_X \varepsilon$$

where $X \in \{\tau, Y\}$ indexes the mode of communication and where

$$\psi_\tau \equiv -\frac{\alpha^2\gamma(1-\sigma)}{1-\sigma\alpha\gamma} \leq 0 \qquad \text{and} \qquad \psi_Y \equiv \frac{\alpha(1-\alpha)(1-\gamma)(1-\sigma)}{1-\alpha(1-\sigma(1-\gamma)))} \geq 0.$$

In the scenario with anchored beliefs, the mode of communication regulated the slope of the implementability restriction between τ and Y. In the present scenario, this slope is pegged to 1, as in the frictionless benchmark, but the implementability constraint is perturbed away from that benchmark by the sunspot-like shock ε. The mode of communication now regulates the impact of this shock.

The next result sheds further light on the implementability constraints faced by the policymaker by studying the comparative statics of ψ_τ and ψ_Y with respect to γ.

Proposition 17. *The following are true:*

1. *ψ_τ is non-positive and strictly decreasing in γ, and equals zero at $\gamma = 0$.*

2. *ψ_Y is non-negative and strictly decreasing in γ, and equals zero at $\gamma = 1$.*

These properties are illustrated in Figure 4 and have a similar flavor as those documented earlier in Figure 2: a stronger GE feedback reduces the impact of erratic beliefs under target communication, and increases it under instrument communication. (The frictionless benchmark corresponds to $\psi = 0$.)

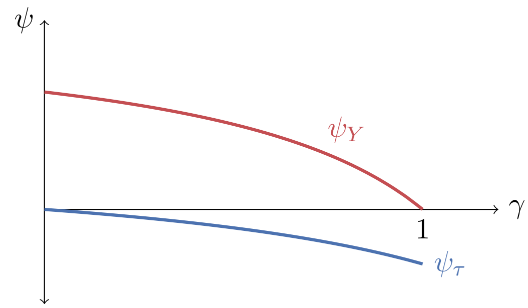

Figure 4: Dependence of (ψ_τ, ψ_Y) on GE feedback γ.

9.2 Optimal communication

Conditional on instrument communication, the policymaker chooses a message $\hat{\tau}$ so that

$$\hat{\tau} \in \arg\min_{\tau} \int L\left(\tau, \tau - \psi_\tau \varepsilon, \theta\right) \varphi(\varepsilon)\, \mathrm{d}\varepsilon,$$

where φ is the p.d.f. of the sunspot-like shock. Conditional on target communication, the policymaker instead chooses a message $\hat{Y}$ so that

$$\hat{Y} \in \arg\min_{Y} \int L\left(Y + \psi_Y \varepsilon, Y, \theta\right) \varphi(\varepsilon)\, \mathrm{d}\varepsilon.$$

In both cases, the applicable implementability constraint has already been incorporated in the objective and the integration over ε captures the restriction that the message cannot be contingent on ε. The optimal mode of communication is then determined by comparing the minimal losses obtained by the solution to the above two problems.

Because of the quadratic specification of L and the Gaussian specification of the ε shock, it is straightforward to solve for the message and the policymaker's loss in each case. With instrument communication, the policymaker picks $\hat{\tau} = \theta$ and obtains a loss equal to $L_\tau \equiv \chi \mathrm{Var}[Y - \theta] = \chi \psi_\tau^2$, for all θ. With target communication, on the other hand, the policymaker picks $\hat{Y} = \theta$ and obtains a loss equal to $L_Y \equiv (1-\chi)\mathrm{Var}[\tau - \theta] = (1-\chi)\psi_Y^2$, for all θ. It follows that, regardless of θ, target communication is preferred to instrument communication if and only if $L_Y < L_\tau$, or equivalently $(1-\chi)\psi_Y^2 < \chi\psi_\tau^2$.

We now study how this comparison depends on γ. From Proposition 17, we have that, as we increase γ continuously from 0 to 1, ψ_τ^2 increases continuously from 0 to a positive number, whereas ψ_Y^2 decreases continuously from a positive number to 0. It follows that there exists a threshold $\tilde{\gamma} \in (0,1)$ such that $(1-\chi)\psi_Y^2 < \chi\psi_\tau^2$ if and only if $\gamma > \tilde{\gamma}$, which proves the following.

Theorem 18. *For any $\sigma > 0$, there exists a threshold $\tilde{\gamma} \in (0,1)$ such that the following is true: for $\gamma \in [0, \tilde{\gamma})$, instrument communication is strictly optimal for all realizations of θ; and for $\gamma \in (\tilde{\gamma}, 1]$, target communication is strictly optimal for all realizations of θ.*

Despite the different nature of the belief distortion under consideration, the take-home policy lesson is essentially the same as that obtained before in Theorem 14: target communication is optimal if and only if the GE feedback is strong enough.

The common thread behind the two theorems is how the GE feedback and the mode of communication interact in shaping the nature and the strength of the strategic interaction among the agents. The sharpest possible version of this point is made by considering, once again, the extremes in which $\gamma = 0$ and $\gamma = 1$. When $\gamma = 0$, instrument communication eliminates the impact of *either* kind of belief distortion simply by guaranteeing that the behavior of each agent is independent of her beliefs of the behavior of other agents. When instead $\gamma = 1$, the exact same thing is achieved by target communication. Finally, in between these two extremes, the impact of either distortion is non-zero under both modes of communication, but the basic logic survives in the sense that a higher γ tilts the balance in favor of target communication.

We close this section by commenting on the resemblance of the exercise conducted in this section to that in Poole (1970). In both cases, the policymaker tries to minimize the impact of "unwanted" shocks on the economy. The same basic logic drives the analysis of optimal Taylor rules in the DSGE literature. However, this resemblance is somewhat superficial. Poole's analysis, and the modern DSGE literature alike, does not require a departure from either rational expectations or the representative-agent framework. Instead, the relevant friction is the inability to condition the policy instrument directly on the underlying shocks to payoff-relevant fundamentals such as preferences, technology, and monopoly power. Here, instead, the aforementioned departures are central precisely because the relevant friction is how agents form beliefs, or reason, about the behavior of others. This also explains why our result, unlike those featured in Poole (1970) and the DSGE literature, is robust to letting the policymaker face no uncertainty about the agents' beliefs. This point is self-evident in the earlier setting with anchored beliefs, extends to a variant of the present setting that lets ε be observed by the policymaker, and is consistent with the logic given above.[31]

10 Other Policy Strategies

So far, we have focused on two policy options: "simple" forward guidance about either the policy instrument or the policy target. We now broaden the scope to a larger toolkit for managing expectations. First, we show that attempting forward guidance about either the aggregate action K or the policy parameter θ is ineffective—or, at least, ill-posed—in our model. Next, we show that the consider forms of forward guidance may be dominated by a more sophisticated one that has the policymaker commit to a linear policy rule that links the instrument with the outcome. Such a strategy allows the policymaker

[31] Appendix F works out several variations of our model with additional shocks, which introduce similar considerations as those raised in Poole (1970). These considerations may tilt the balance toward either instrument or target communication, but are orthogonal to the role played by higher-order beliefs and therefore do not affect any of our main lessons. In particular, the comparative statics with respect to the GE feedback (γ) and degree of the belief distortion (λ or σ) are unchanged.

substantially more flexibility in determining the relationship *between* different expectations, which in turn helps further contain any belief distortion. Our discussion thus provides a new perspective on the function that policy rules, including Taylor rules for monetary policy, may serve in the presence of bounded rationality or higher-order uncertainty.

10.1 Communicating θ or K

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.

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.

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 6 and let the policymaker announce and commit to a value $\hat{K}$ for aggregate investment. In Appendix E, we show that there exists a system of beliefs about τ and Y that is consistent with the belief that K will equal $\hat{K}$ if and only if $\lambda = 1$ (i.e., rational expectations). When instead $\lambda < 1$, there does not exist an equilibrium in which the policymaker fixes a target $\hat{K}$ for aggregate investment. The reason is that, unlike in the case of a Y target, the policymaker has not have the power to persuade the agents that she can attain a K target "no matter what."

We alluded to this kind of problem when we noted the necessity of letting τ have a direct, mechanical effect on Y (or the fragility of target communication as $\alpha \to 1$). The same basic logic applies. To make sense of commitments on K, we would have to add a new policy instrument that can *directly* control the investment decisions of the firms. That is, we would have to modify (2) to

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

where z is the new policy instrument. 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.

Put differently, it is precisely the absence of such an instrument that justifies the focus on "managing expectations." In the context of forward guidance studied in Section 8, this simply means the following: if the central bank could use current interest rates (the analogue of z) to control aggregate demand, there would naturally be less need for engaging in forward guidance of any type.

10.2 Policy rules

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:

$$\tau = A - BY \tag{22}$$

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 (22) and (1) jointly for τ and Y and substitute the solution into (2) to obtain the following game representation:

$$k_i = \zeta(A, B; \alpha, \gamma) + \delta(B; \alpha, \gamma)\mathbb{E}_i[K] \tag{23}$$

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 (23).

Consider now the case with anchored beliefs (as in Section 6). 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.[32]

We summarize these lessons in the following result.

Proposition 19. *Suppose that the policymaker can announce and commit on a policy rule as in (22) and let Assumptions 1 and 3 hold with $X = (A, B)$.*

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

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

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

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

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

[32] 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 19 as a complement to our main analysis, not a sign that the choice between instrument and target communication was too narrowly framed. Proposition 19 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.

11 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 aggregate employment?

We pose this question in a stylized model in which agents form mis-specified expectations, either anchored to a reference point or subject to erratic impulses. Our main result shows 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.

The mechanism is intuitive. Instrument communication pins down the expectations of the policy instrument itself, but leaves the agents with the task of having to predict, or reason, how aggregate outcomes will be determined. Target communication does the opposite: it pins down beliefs of the aggregate outcome but leaves the agents with the task of figuring out what policy will support this outcome. Which type of communication is optimal depends on whether mistakes in the former kind of reasoning are most costly that mistakes in the latter kind of reasoning.

When the GE feedback is strong, the actual outcome depends relatively more on expectations of the outcome itself and relatively less on expectations of the policy instrument. In this case, it is optimal to minimize the mistakes in the expectations of the outcome, which is precisely what target communication achieves. When the GE feedback is weak, the opposite is true.

Put more succinctly, the optimal communication strategy minimizes agents' need to "reason about the economy," precisely because this reasoning produces distortions. The micro origin of such distortions—whether belief bias, imperfect computation of equilibrium, or incomplete information—is less important for the result.

Along the way, we uncovered additional insights, such as how Taylor rules can play a new role in regulating the bite of bounded rationality, or how the latter may itself be the source of a commitment problem. In all these cases, our analysis suggested interesting trade-offs but 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—e.g., anchored beliefs offsetting financial amplification. 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.

Allen, Franklin, Stephen Morris, and Andrew Postlewaite. 1993*a*. "Finite Bubbles with Short Sale Constraints and Asymmetric Information." *Journal of Economic Theory*, 61(2): 206–229.

Allen, Franklin, Stephen Morris, and Andrew Postlewaite. 1993*b*. "Finite bubbles with short sale constraints and asymmetric information." *Journal of Economic Theory*, 61(2): 206–229.

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*a*. "Quantifying Confidence." *Econometrica*, 86(5): 1689–1726.

Angeletos, George-Marios, Fabrice Collard, and Harris Dellas. 2018*b*. "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.

Barro, Robert J., and David B. Gordon. 1983. "Rules, Discretion and Reputation in a Model of Monetary Policy." *Journal of Monetary Economics*, 12(1): 101 – 121.

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

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.

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

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

Chari, V.V., Lawrence J. Christiano, and Martin Eichenbaum. 1998. "Expectation Traps and Discretion." *Journal of Economic Theory*, 81(2): 462 – 492.

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*.

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.

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

Farhi, Emmanuel, and Ivan Werning. 2016. "Fiscal Multipliers: Liquidity Traps and Currency Unions." *Handbook of Macroeconomics*, 2: 2417–2492.

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

Friedman, Benjamin M. 1990. "Targets and Instruments of Monetary Policy." In *Handbook of Monetary Economics*. Vol. 2. 1 ed., , ed. B. M. Friedman and F. H. Hahn, Chapter 22, 1185–1230. Elsevier.

Gabaix, Xavier. 2017. "A Behavioral New Keynesian Model." *Harvard mimeo*.

Garcıa-Schmidt, Mariana, and Michael Woodford. 2018. "Are Low Interest Rates Deflationary? A Paradox of Perfect-Foresight Analysis." *American Economic Review*, forthcoming.

Huo, Zen, and Naoki Takayama. 2015. "Higher Order Beliefs, Confidence, and Business Cycles." Yale University miméo.

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.

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

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

Mauersberger, Felix, and Rosemarie Nagel. 2018. "Levels of Reasoning in Keynesian Beauty Contests: A Generative Framework." In *Handbook of Computational Economics*. Vol. 4 of *Handbook of Computational Economics*, , ed. Cars Hommes and Blake LeBaron, Chapter 10, 541 – 634. Elsevier.

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

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

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

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

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

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

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

Scheinkman, Jose A, and Wei Xiong. 2003. "Overconfidence and speculative bubbles." *Journal of Political Economy*, 111(6): 1183–1220.

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.

Walsh, Carl E. 2007. "Optimal Economic Transparency." *International Journal of Central Banking*, 3(1): 5–36.

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

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

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

A Proofs

A.1 Slopes of Budget Lines

A.1.1 Proof of Proposition 12

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 12,

$$\mu_\tau = \left((1-\alpha) + \alpha\frac{1-\delta_\tau}{1-\lambda\delta_\tau}\right)^{-1} \tag{24}$$

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

$$\mu_Y = \frac{1-\alpha b}{1-\alpha} = \frac{1-\lambda\delta_y - \alpha(1-\delta_y)}{(1-\alpha)(1-\lambda\delta_y)} \tag{25}$$

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, $\partial\mu_Y/\partial\gamma < 0$.

A.1.2 Further results

Lemma 20 (Sign of μ_Y)**.** *$\mu_Y > 0$ if and only if $\lambda \geq \alpha$ or $\gamma > \frac{1+\alpha(\lambda-2)}{\alpha(\lambda-\alpha)}$.*

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

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

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

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

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

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

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

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

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

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

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

Lemma 21. *Assume that $\mu_Y > 0$ and $\alpha\gamma < 1$. Then $\mu_\tau > \mu_Y$.*

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

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

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

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

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

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

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

A.2 Proof of Proposition 13

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

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

Meanwhile, the slope with target communication is

$$\mu_{y,1} = 1$$

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

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

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 12.

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

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

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

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 interaction δ_y is

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

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

A.3 Proof of Theorem 14

Let $r \equiv Y/\theta$. 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

$$r^*(\mu) := \frac{\mu^2(1-\chi)+\mu\chi}{\mu^2(1-\chi)+\chi} \tag{26}$$

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 20.

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 12.

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 $\mu > 1$ (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:

$$\lim_{\gamma \to -\infty} \mu_\tau = \mu_{\tau, -\infty} \equiv \frac{\lambda}{\lambda + (1 - \lambda)\alpha} \in (0, 1) \tag{27}$$

and

$$\lim_{\gamma \to -\infty} \mu_Y = \mu_{Y, -\infty} \equiv \frac{\lambda(1 - \alpha/\lambda)}{\lambda(1 - \alpha)} \tag{28}$$

A.3.1 Other results

Theorem 22. *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 20) or $\chi < 1/2$, $\dot{\gamma} = -\infty$.*

Proof. First, maintain Lemma 20 and its assumptions. Note that the second case ("more general") of the proof of the previous section does not use $\gamma > 0$. Hence the result is proved for $\dot{\gamma} = -\infty$ in this case.

Now relax those assumptions. Our best bound on the loss with target communication, for $\mu_Y < 0$, is $\min\{\mathcal{L}(\mu_{y,-\infty}), 1-\chi\}$, 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 $\underset{\cdot}{\gamma} = -\infty$ 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 $\gamma \in (-\infty, \dot{\gamma})$. □

A.4 Proof of Proposition 17

Instrument communication. Recall that

$$\psi_\tau = -\frac{\alpha^2 \gamma (1-\sigma)}{1-\sigma\alpha\gamma}$$

For $\gamma \in [0,1]$ and $\sigma \in [0,1]$, the numerator is non-negative. Additionally, given $\alpha \in (0,1)$, the denominator is strictly positive. Hence $\psi_\tau \leq 0$ on this domain.

The partial derivative with respect to γ is the following:

$$\frac{\partial \psi_\tau}{\partial \gamma} = \frac{-\alpha^2(1-\sigma)}{(1-\sigma\alpha\gamma)^2} < 0$$

so this function is decreasing for all values of γ. More transparently, the numerator of $|\psi_\tau|$ always increases and the denominator always decreases as γ increases.

Target communication. Recall that

$$\psi_Y = \frac{\alpha(1-\alpha)(1-\gamma)(1-\sigma)}{1-\alpha(1-\sigma(1-\gamma)))}$$

and the derivative is

$$\frac{\partial \psi_\gamma}{\partial \gamma} = -\frac{\alpha(1-\alpha)^2(1-\sigma)}{(1-\alpha(1-\sigma(1-\gamma)))^2} < 0$$

A.4.1 Other results

Proposition 23. *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 $\delta_Y < \delta_\tau \le 0$ for $\gamma \le 0$. Thus $\delta_\tau = -\alpha g(\delta_\tau) < -\alpha g(\delta_Y) < -\alpha g(\delta_y)/(1-\alpha) = \psi_Y$. □

A.5 Proof of Theorem 18

Let $\mathcal{L}_\tau(\gamma)$ and $\mathcal{L}_Y(\gamma)$ denote the policymaker's loss evaluated at the optimal message as a function of GE feedback parameter γ. As mentioned in the main text, $\mathcal{L}_\tau(\gamma) = \chi\psi_\tau^2(\gamma)$ and $\mathcal{L}_Y(\gamma) = (1-\chi)\psi_Y^2(\gamma)$. It is straightforward to deduce from the expressions for (ψ_τ, ψ_Y) and from Proposition 17 that the following are true:

1. $\partial\mathcal{L}_\tau/\partial\gamma = 2\chi\psi_\tau(\partial\psi_\tau/\partial\gamma) \ge 0$, $\mathcal{L}_\tau(0) = 0$, and $\mathcal{L}_\tau(1) > 0$.
2. $\partial\mathcal{L}_Y/\partial\gamma = 2(1-\chi)\psi_Y(\partial\psi_Y/\partial\gamma) \le 0$, $\mathcal{L}_Y(0) > 0$, and $\mathcal{L}_Y(1) = 0$.

It follows that there exists as single crossing point $\hat{\gamma} \in (0,1)$ such that instrument communication is preferred for lower γ and target communication is preferred for higher γ.

A.6 Proof of Comparative Statics

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 $\gamma \in [0,1]$:

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

With this expression, we can do analytical comparative statics.

A.6.1 Policy parameter α

Proposition 24. *For fixed (λ,χ), the threshold $\hat{\gamma}$ is an increasing function of α. That is, the range of γ for which target communication is optimal decreases as $\partial Y/\partial K$ gets larger.*

Proof. 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 like terms,

$$(1-\alpha)^2 > 0$$

which is trivial for $\alpha \in (0,1)$. 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 $\partial\hat{\gamma}/\partial\alpha > 0$. □

A.6.2 Attentive fraction λ (Proposition 15)

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.

A.6.3 Output gap parameter χ

Proposition 25. *For fixed (α,λ), the threshold $\hat{\gamma}$ is an decreasing function of χ. That is, the range of γ for which target communication is optimal increases as the policymaker pays more attention to the output gap.*

Proof. 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. The first is the neoclassical economy introduced in the setup of our framework (Section 3). The second is the stylized New Keynesian economy mentioned in our discussion of forward guidance (Section 8).

B.1 A neoclassical economy with aggregate demand externalities

In this section we fill in the details of the micro-founded example discussed in the main text. The set up was described in Section 3, p.3. Here, we solve the model and explain how it is nested in our abstract framework.

B.1.1 Solution

It is easiest to solve this model backward in time.

Period 2. The final goods producer's demand for intermediates is the following:

$$p_i = \alpha Y X^{\rho-1} x_i^{-\rho}$$

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

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

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

Finally, where do the wages come from? Labor supply has the following form:

$$w = \frac{(1+\varphi)N^{\varphi}}{1-r}$$

When $\varphi = 0$, wages are inelastically supplied at the inverse tax rate. Labor demand is set by the final-goods firm:

$$wN = (1-\alpha)Y$$

Taking stock, the entrepreneurs make more second-period profit per unit of capital precisely when output is high and wages are low. Wages are low (in partial equilibrium, for fixed N) precisely when taxes are low.

Period 1. The entrepreneur invests until the marginal return on capital is one:

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

The first-order condition re-arranges to

$$x_i^{\rho} = \beta\eta(1-\rho)\mathbb{E}\left[X^{\eta+\rho-1} N^{1-\eta}\right] \tag{29}$$

Investment solves this fixed-point equation. Given sufficiently big demand externalities, $\rho > 1-\eta$, the first term encodes the strategic complementarity in the economy. The second term, once one solves for equilibrium labor, will encode the substitutability.

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

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

It is immediate that output is linear in labor:

$$Y = X^\eta N^{1-\eta} = (\beta\eta(1-\rho))^{\frac{\eta}{1-\eta}} N$$

Of course, it is also linear in investment by the same token:

$$Y = (\beta\eta(1-\rho))^{-1} X$$

It remains only to solve for one of them (more easily, N) as a function of primitives. Setting labor supply to labor demand gives

$$(1-\eta)\frac{Y}{N} = \frac{(1+\varphi)N^\phi}{1-r} \tag{30}$$

Plugging in the previous expression of the form $Y = A_N N$, this reduces to

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

Note first that $\phi > 0$ is necessary for this to be well-defined. Without convex labor supply, the scale of this economy is not determined: two times the investment yields exactly twice the output. We might have instead assumed some decreasing returns in the production of intermediate goods or some curvature in the utility of consumption.

Second, a monotonic transformation of the labor subsidy, $Z := (1-r)^{\frac{1}{\phi}}$, shows up in the expression for N. Higher taxes contract the labor supply, which both directly affects output and reduces first-stage investment.

B.1.2 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 29 can no longer be solved without expectations. To make progress, we will take log-linear approximations. Let lowercase letters denote the log deviation quantities.

Aggregate production is exactly log-linear:

$$y = \eta x + (1-\eta)n$$

The expression equating labor supply and labor demand, (30), is also exactly

$$y - n = \phi(n - z)$$

where, again, $z := \log(1-\tau)/\varphi$ is a rescaled labor subsidy. Combining these two expressions yields the following expression for output as a function of investment and policy:

$$y = \frac{(1-\eta)\phi}{\phi+\eta} z + \frac{\eta(1+\phi)}{\phi+\eta} x \tag{31}$$

This maps, in our abstract model, to

$$\alpha := \frac{\eta(1+\phi)}{\phi+\eta} \tag{32}$$

The direct effect of policy is stronger ($1-\alpha$ is larger) when labor is a larger share of output and when labor supply is fairly inelastic (small φ). Note that the Dixit-Stiglitz aggregation and aggregate demand externality play no role in these expressions, which hold also under complete information.

Let us now turn to the investment decision (29). To a log-linear approximation, it is

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

We substitute out x and n, respectively, using the expressions (31) and (30):

$$x_i = \left(1 - \frac{1-\eta}{\rho}\right)\mathbb{E}_i\left[\alpha y - \frac{\alpha}{1-\alpha} z\right] + \frac{1-\eta}{(1+\varphi)\rho}\mathbb{E}_i[y + \varphi z]$$

Collecting terms, we get

$$x_i = (1-\gamma)\,\mathbb{E}_i[z] + \gamma\mathbb{E}[y] \tag{33}$$

for feedback parameter

$$\gamma := \alpha - \frac{1-\eta}{\rho}\left(\alpha - (1+\phi)^{-1}\right) \tag{34}$$

For this model to behave in a stable way (i.e., produce $\gamma \in (-\infty, 1]$), we want to assume $\alpha > (1+\phi)^{-1}$, or $\phi > 1/\eta - 2$. This condition is trivial if η, the weight on intermediates in the production function, is greater than or equal to $1/2$. Otherwise, it requires a sufficiently high level of φ, which controls the effective scarcity of labor.

On this domain, it is clear that the GE feedback parameter increases in ρ. In particular, as $\rho \to 1$, γ approaches a limit $\bar{\gamma} := \eta\alpha + (1+\phi)^{-1} \in (0,1)$. As $\rho \downarrow 0$, γ decreases to $-\infty$.

B.2 A New Keynesian economy

In this section we present an alternative example of a stylized New Keynesian economy.

Set-up. Let all variables be in log deviation from a steady state. Consider a three-period economy ($t \in \{0, 1, 2\}$) with two types of consumers. A "type d (discretionary)" agent consumes in periods 1 and 2. She earns income y only in period 1 and has an opportunity to save at interest rate r for period 2. The consumption $c_{i,d}$ for one such individual obeys the following Euler equation:

$$c_{i,d} = (1-\beta)y - \beta r$$

where β is a discount factor (and thus $1-\beta$ is a permanent-income marginal propensity to consume). There are mass χ of such agents.

Our "reduced-form" period 2 can be thought of as a substitute for the infinite future in a fully-specified model (see, for instance, Angeletos and Lian, 2018). The simplification made in our Euler equation is that, at least in expectation, all future income and interest rates are at the steady state level. Thus, when we substitute the lifetime budget constraint into the standard Euler equation, we get the previous simple expression.

A second type of agent, of mass $1-\chi$, also consumes only in periods $t \in \{1, 2\}$, but precommits at $t = 0$ to her consumption at $t = 1$ ("type p"). She also earns income y in period 1 and faces interest rate (from period 1 to period 2) r. She tries in expectation to follow the same reduced-form Euler equation.

$$c_{i,p} = (1-\beta)\mathbb{E}_i[y] - \beta\mathbb{E}[r]$$

These agents understand the structure of the economy but may have mis-specified beliefs about each others' actions.

Output, and hence income, at $t = 1$ is determined by total demand: $y = \int_i c_i \, \mathrm{d}i = (1-\chi)c_p + \chi c_d$, where (c_p, c_d) denote the cross-sectional average consumption of each agent. This scenario might represent a liquidity trap, which the economy will escape in period 2.

Finally, at $t = 0$, the policymaker can make an announcement about the intended value of r or target for y.

Solution. Let us re-arrange the equations slightly to clarify the feedback between each group's consumption. Substituting the Euler equation into the market clearing expression yields the following expression for output as a function of pre-committed consumption and interest rates:

$$y = \frac{\chi\beta}{1-\chi(1-\beta)}(-r) + \frac{1-\chi}{1-\chi(1-\beta)}c_p \tag{35}$$

This fits our abstract framework with $\tau = -r$ (a renormalization) and

$$\alpha = \frac{1-\chi}{1-\chi(1-\beta)} \tag{36}$$

The "direct effect" of decreasing the interest rate, via the Euler equation of the discretionary fraction χ, is $\chi\beta$. In equilibrium, decreasing the interest rate has an "extra kick" over this direct effect ($1-\alpha >$

$\chi\beta$) because of a familiar Keynesian cross: more consumption also produces more income, and then even more consumption. Increasing pre-committed consumption does something similar: it also increases income, and hence spending, for the discretionary consumers. In that way, any "sentiment shock" that causes more pre-committed consumption causes an economic boom.

Recall that pre-committed consumption already matches our abstract model's form:

$$c_p = \beta\bar{\mathbb{E}}[-r] + (1-\beta)\bar{\mathbb{E}}[y] \tag{37}$$

for $\gamma := (1-\beta)$. Now, the pre-committed consumption increases when agents are optimistic about policy and income at $t = 1$. To sum up, there are two-way feedbacks between pre-committed and discretionary consumption, going through the fact that output is demand determined.

We can think of comparative statics in β controlling feedback in this economy. A larger β may correspond to a longer horizon or larger "effective" discounting induced by, for instance, liquidity constraints. Both of these amount to having a larger "pre-committed" Keynesian cross.

In our baseline formulation, it is impossible to affect this GE feedback effect without affecting the second-period Keynesian cross as well. In particular, a higher β will correspond to a larger direct effect of interest rates on the economy. These forces could be "mechanically" decoupled by assuming that pre-commitment and discretionary consumers have different discount rates (and/or MPCs).

C Equivalence with "Smooth" Level-k Thinking

Let $K = b_{X,k}\hat{X}$ in the level-k economy in which fundamental $X \in \{\tau, Y\}$ is announced and all agents compute to order k. The coefficients $(b_{X,k})_{k=0}^{\infty}$ follow the following recursion:

$$b_{X,k} = 1 + \delta_X\left(b_{X,k-1} - 1\right) \tag{38}$$

and we recover that $\lambda_{X,k} := b_{X,k}/b_{X,k-1}$ as the ratio between realized K and expected K. It follows that we could plug $\lambda_{X,k}$ into the inertial beliefs model and recover the same relationship between K and $\hat{X}$. For instrument communication, the sequence $\lambda_{\tau,k}$ is bounded above by 1 and decreases in k. For target communication, the sequence $\lambda_{Y,k}$ switches sides of one for even ($\lambda_{Y,k} < 1$) and odd ($\lambda_{Y,k} > 1$) periods. This oscillatory behavior is a familiar behavior of level-k models with strategic substitutability. Our model agrees with the level-k model for only the even k.

This last property can be "smoothed out" in a more refined solution concept. Consider the following model based on the "reflective equilibrium" in Garcıa-Schmidt and Woodford (2018). Let $\mathcal{B}(b_X; X)$ map the conjecture $\mathbb{E}_i[K] = b_X X$ to the equilibrium relationship $K = b_X^* X$ for a given communicated fundamental. Assume that the agents' conjecture is a smooth function of "cognitive depth" t solving the following differential equation:

$$\frac{\mathrm{d}b_X(t)}{\mathrm{d}t} = \mathcal{B}(b_X(t);X) - b_X(t)$$
$$b_X(0) = b_{X,k=0} \tag{39}$$

It is straightforward to verify that, given an initially conservative guess ($b_X(0) < \mathcal{B}(b_X(0);X)$), the solution is also conservative ($b_X(t) < \mathcal{B}(b_X(t);X)$ for all $t > 0$). The dampening parameter $\lambda_X(T) := b_X(T)/\mathcal{B}(b_X(t);X)$, which maps to λ in our model, remains less than one for both communication modes.

Theorem 14 remains true for fixed T. There is no disruption to the logic from having different dampening λ_X conditional on communication choice.

D Positive vs. Negative GE Feedback

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 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 $\delta_Y < \delta_\tau < 0$. This directly translates into a "harder" implementability constraint under target communication, whether bounded rationality takes the form of anchored beliefs (in which case we have $\mu_Y < \mu_\tau < 1$) or erratic beliefs (in which case we have $|\psi_Y| > |\psi_\tau|$).

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. Theorem 22 in the Appendix elaborates on the details.

In the model with erratic beliefs, the shape of the loss function always matters. The policymaker cares asymmetrically about variation in the instrument gap (as induced by target communication) versus variation in the policy gap (as induced by instrument communication), and lacks any tools (like the choice of r in the first model) to shift the distortions between the gaps. 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.

In all cases, for sufficiently low γ, we rely on possibly undesirable unstable equilibria for which $\delta_X < -1$. We might strengthen the equilibrium concept for implementability to eliminate these cases. Then all the previous statements could only hold for $\gamma \geq -(\alpha\lambda)^{-1}$ or $\gamma \geq -(\alpha\sigma)^{-1}$ in the two models, respectively. Below this point, the optimal communication choice is ill-posed, since neither announcement induces an interpretable equilibrium to determine K.

E Communicating K

Imagine that a policymaker announces $\hat{K}$ and agents have the "inertial beliefs" friction of Assumption 3. 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, as noted in the main text, 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 6.5).

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.

F 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.

F.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 4 and 5 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.

F.2 Measurement errors

The same logic as above applies if we introduce measurement errors in the policymaker's estimates of τ and Y. To see this, consider a variant of our framework that lets the policymaker observe only $(\tilde{\tau}, \tilde{Y})$, 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. Accordingly, 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. 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.