

NBER WORKING PAPER SERIES

MANAGING EXPECTATIONS:  
INSTRUMENTS VS. TARGETS

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, Revised March 2020

An earlier version of this paper was entitled “Managing Expectations without Rational Expectations.” We are grateful to Gaetano Gaballo, Gauti Eggertsson and Marcus Hagedorn for insightful conference discussions; and to Roland Benabou, V.V. Chari, Martin Eichenbaum, John Leahy, Stephen Morris, and various seminar participants for valuable feedback. Angeletos acknowledges the financial support of the National Science Foundation (Award #1757198). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2018 by George-Marios Angeletos and Karthik A. Sastry. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Managing Expectations: Instruments vs. Targets  
George-Marios Angeletos and Karthik A. Sastry  
NBER Working Paper No. 25404  
December 2018, Revised March 2020  
JEL No. D82,D84,E03,E52,E58

## **ABSTRACT**

Should a policymaker offer forward guidance in terms of a path for an instrument such as interest rates or a target for an outcome such as unemployment? We study how the optimal approach depends on bounded rationality. People make mistakes in reasoning about the behavior of others and the equilibrium mapping between policy and outcomes. The policymaker wishes to minimize the effects of such mistakes on implementability and welfare. This goal is achieved by target-based forward guidance if and only if GE feedback is strong enough, as when faced with a prolonged liquidity trap, a steep Keynesian cross, or a large financial accelerator.

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

# 1 Introduction

Forward guidance is not comprehensive. Even if a central bank can shape expectations about future interest rates, it remains up to the public to predict the consequences for aggregate employment and income. Under what circumstances is it better to do the opposite, anchoring expectations about the targeted economic outcomes and leaving the public to ponder the supporting policy?

During the Great Recession, many central banks faced such a choice. In December 2012, the Federal Reserve famously switched from providing a timeline over which rates would remain near zero to highlighting a 6.5% target for unemployment.<sup>1</sup> Earlier that same year, ECB President Mario Draghi provided perhaps the purest example of target-based communication: a commitment to do “whatever it takes” to save the Euro. This statement is widely credited with restoring market confidence despite a lack of clarity about the policy instruments that would support this target—or perhaps because of it.<sup>2</sup>

This paper offers a new gauge for when to *optimally* sacrifice clarity about instruments for clarity about targets—and hence also a new rationale for why the previously mentioned communication strategies may have been timely. We introduce a meaningful policy trade-off over the public’s expectations of policy instruments and policy targets by relaxing the rational-expectations solution concept: there is a bound on the public’s depth of knowledge or depth of rationality. People consequently make mistakes in reasoning about others’ behavior and hence about the policy’s equilibrium implications.

The policymaker cannot eliminate these mistakes. The challenge is to convince a public that has the “wrong model” about how equilibrium works to nonetheless do the “right thing.” Our main result is that this is achieved by switching from instrument-based to target-based forward guidance, or from anchoring the public’s expectations of interest rates to anchoring its expectations of unemployment, when general equilibrium (GE) feedback loops intensify, as they would during a severe liquidity trap.

**Framework.** We employ an abstract, minimalistic model in order to convey the main insights as transparently as possible. But we also nest a micro-foundation of the following context. The economy is in a recession at the zero lower bound (ZLB). Aggregate demand depends on expectations of future interest rates and aggregate income. Aggregate income, in turn, is demand-determined. The resulting feedback loop between income and spending, or the Keynesian cross, is an example of the GE mechanism we have in mind. A financial accelerator, or positive feedback loop between asset prices and demand, is another.

Conventional monetary policy is unavailable but the central bank can offer a policy commitment for the future. Such “Odyssean” forward guidance can take one of two forms: instrument communication or target communication. The former means a commitment to keep interest rates low until a certain date or engage in quantitative easing (QE) of a certain size. The latter means a credible promise to “do whatever it takes” to achieve a certain target for aggregate employment or income.<sup>3</sup>

---

<sup>1</sup>The target regime is sometimes called the “Evans rule,” in reference to its strong advocate, Charles Evans. The exact statement involved targets also for inflation, but the spotlight was on unemployment; see the discussion in Section 2.4.

<sup>2</sup>The Draghi example is sometimes associated with the logic of selecting the “no bank run” equilibrium in models with multiple, self-fulfilling equilibria. Although this is *not* the route taken here, we later discuss how our approach can be viewed as an extension of that logic to unique-equilibrium models once we relax the rational expectations solution concept.

<sup>3</sup>Following [Campbell et al. \(2012\)](#), the literature refers to the communication of pure commitments as “Odyssean” forward

**A rational-expectations benchmark and beyond.** Consider a “textbook” benchmark in which agents not only know the correct equilibrium mapping between policy and economic outcomes but also have no doubts about others’ knowledge, rationality, and responsiveness to forward guidance. This precludes a meaningful policy trade-off between anchoring expectations of the instrument (interest rates) and anchoring expectations of the target (unemployment). Agents flawlessly reasons back and forth between these two objects, without any consequence on implementability and welfare.

By direct implication, the instrument-versus-target choice is irrelevant under rational expectations. This irrelevance is essentially a translation into our setting of the equivalence of the primal and dual versions of rational-expectations policy problems (Chari and Kehoe, 1999; Lucas and Stokey, 1983). In the particular problem studied here, rational expectations also achieves the first best.<sup>4</sup>

We depart from this benchmark by limiting the private sector’s depth of knowledge and/or rationality.<sup>5</sup> Agents make mistakes when trying to reason through the equilibrium relation between policy, aggregate behavior, and macroeconomic outcomes. The policymaker cannot eliminate these mistakes but can regulate their footprint on different *kinds* of expectations—expectations of interest rates versus expectations of aggregate employment and income—and thereby on actual behavior.

In our preferred specification, these mistakes amount to a form of myopia and to under-estimation of the responses of others. Such a friction is the core common element of a related literature on monetary policy and bounded rationality (Angeletos and Lian, 2018; Farhi and Werning, 2019; Gabaix, 2019; Garcia-Schmidt and Woodford, 2019), and is broadly consistent with both the kind of “shallow” reasoning observed in laboratory experiments (Nagel, 1995; Crawford, Costa-Gomes and Iribarri, 2013) and salient features of surveys of macroeconomic expectations (Coibion and Gorodnichenko, 2012, 2015). But mistakes of the opposite kind, a form of belief over-reaction and hyperopia, are also considered.

**Results.** We start by explaining *why* and *how* the form of forward guidance helps regulate the bite of bounded rationality. By committing to different future policy reactions, instrument and target communication induce sharply different strategic interactions within the private sector. In the former case, agents play a game of strategic complements: when an agent expects the others to spend less for a given path of interest rates, she responds less herself because she expects lower aggregate income. In the latter case, everything flips: conditional on an announced employment or GDP target, an agent that expects higher aggregate spending also expects tighter monetary policy in the future, which reduces the incentive to spend.

That such different strategic considerations balance out to the same behavior in the textbook policy paradigm is a knife-edge consequence of *infinitely deep* knowledge and rationality. Away from this case, the strategic considerations instead determine how any given error in the public’s equilibrium reasoning

---

guidance—with the alternative, “Delphic” forward guidance, being contaminated with signal extraction about the state of the economy. We are concerned exclusively with the former type. By the same token, we abstract from the “informational effect” emphasized in Nakamura and Steinsson (2018), or the kind of communication featured in Morris and Shin (2002, 2007).

<sup>4</sup>This isolates the departure from rational expectations as the only source of inefficiency, which is ideal for our theoretical purposes. Our liquidity-trap application, however, allows for the additional distortion caused by nominal rigidity and the ZLB.

<sup>5</sup>“Depth of knowledge” relates to what agents think others believe (and so on, to higher orders). “Depth of rationality” refers to whether agents think others are rational (and so on, to higher orders). Sections 4.1 and 6.2 fill in the details.

influences its actual behavior and thereby the policymaker's implementability constraint. And the more salient such considerations are in people's decisions, the larger the distortion caused by any given error in equilibrium reasoning, or any bias in beliefs about others.

The observable content of this insight depends on the *specific* error in people's reasoning. In our (and the related literature's) preferred specification, this error amounts to under-estimating others' responsiveness to shocks. Such an error yields *under-reaction* in games of strategic complementarity but *over-reaction* in games of strategic substitutability. It follows that the power of forward guidance, as measured by the inverse of the change in interest rates needed to implement a given change in aggregate spending, is attenuated under instruments and amplified under targets.<sup>6</sup>

This highlights that the distortion flips *sign* with the form of forward guidance. As discussed momentarily, this qualifies a key prediction of the related literature. But what is most crucial for our purposes, namely for determining the optimal communication strategy, is the *size* of the distortion, not its sign.

Consider how the distortion varies with the underlying GE feedback, or the slope of the Keynesian cross. A greater reliance of individual spending on aggregate income makes the game induced by instrument communication *more* strategic and the game induced by target communication *less* strategic, in the sense of requiring more or less reasoning about the behavior of others. It follows that high GE feedback exacerbates the distortion under instrument communication and eases it under target communication.

What if the belief bias is of the opposite direction than that assumed in our (and the related literature's) preferred specification? This naturally flips the signs of the distortions under both forms of guidance, but does not upset the comparative static of their sizes with respect to the strength of GE feedback. We thus reach our main result, which is robust across various specifications of the primitive friction:

*Main lesson. When GE feedback loops are strong enough, the optimal communication strategy switches from instruments to targets, or from anchoring expectations of interest rates to anchoring expectations of aggregate employment and income. This minimizes the public's need to reason about equilibrium, thus also minimizing the distortion caused by bounded rationality.*

**Robustness.** As just noted, our main lesson is robust to whether people under- or over-estimate the responses of others. The same applies to a third specification with purely random mistakes in equilibrium reasoning. By the same token, a policymaker who suspects that the public "does not fully understand GE" but is not sure of the precise mis-specification thereof could still apply our main lesson.

Our main lesson is also robust to introducing measurement error, policy trebles, and uncertain fundamentals. These elements, which are the focus of the classic by [Poole \(1970\)](#) and the literature reviewed in [Friedman \(1990\)](#), naturally enter the costs and benefits of different policy options. But they do not tie the optimal choice to the relative importance of PE and GE effects, nor do they capture the kind of expectation management trade-off on which we focus.

Our main lesson is also robust to letting the policymaker engage in more sophisticated forms of forward guidance than the two starker forms considered above. This in turn hints at a broader contribution of our paper, which is to offer a new perspective on the design of optimal policy rules.

---

<sup>6</sup>In our abstract model, attenuation and amplification are both "bad." The logic is qualified for the ZLB context in due course.

**Extension to policy rules.** To illustrate the last point, we let the policymaker commit to and announce an arbitrary relation between the policy instrument, the targeted outcome, and the state of nature. This is our model's analogue of a flexible, state-contingent, Taylor rule for future monetary policy.

Such a state-contingent policy rule may be harder to communicate and thus less effective than the starker alternatives considered in our baseline analysis. In the words of then Minneapolis Fed President [Narayana Kocherlakota \(2012\)](#), attempting to communicate an entire policy reaction function in a crisis would be “letting the perfect be the enemy of the good.” Absent this consideration, the added flexibility of course allows the policymaker to attain a better outcome. Still, our main insights generalize as follows.

Under rational expectations, the contingency of the policy on the relevant outcome is redundant given the contingency on the state of nature. But once we bound the public's depth of knowledge or rationality, this redundancy disappears. The logic is similar to the one that broke the instruments-targets irrelevance in our baseline setting: the outcome contingency helps regulate the mistakes in equilibrium reasoning.

The optimal such regulation calls for a policy rule that continuously shifts weight from anchoring expectations about instruments (“interest rates”) to anchoring expectations about targets (“unemployment”) as the GE feedback gets larger. This result smooths out the discrete switch from instruments to targets featured in our main analysis. And at a higher level, it hints at how parameters that govern GE multipliers, such as the marginal propensity to consume or the extent of input-output linkages, should enter the design of “sophisticated” policy rules ([Atkeson, Chari and Kehoe, 2010](#)) in a way not previously understood.

**Application to the Great Recession.** Let us return to our motivating context, forward guidance in a liquidity trap. [Angeletos and Lian \(2018\)](#), [Farhi and Werning \(2019\)](#), [Garcia-Schmidt and Woodford \(2019\)](#), [Gabaix \(2019\)](#) and [Wiederholt \(2016\)](#) have argued that a belief friction like that captured in our preferred specification diminishes the power of forward guidance.<sup>7</sup> But whereas these works study exclusively instrument-based guidance, the discussion at the Federal Reserve largely regarded the question of whether a switch to target-based guidance would make a difference. In this context, we make four contributions.

First and foremost, we offer a gauge for determining when to *optimally* switch to target-based guidance: once Keynesian multipliers are sufficiently large. In our stylized micro-foundation, this is the case when the marginal propensity to consume out of income is large enough. In more sophisticated models, the right timing may correlate with widespread financial distress and deflationary spirals.

Second, we qualify the aforementioned works' lesson about the ineffectiveness of forward guidance in the presence of bounded rationality: forward guidance may *still* be effective, albeit in a different from.

Third, we offer a formal backdrop for discussions of “clarity” and “public interpretation,” which are front-and-center in retrospectives of policy deliberations.<sup>8</sup> Under our lens, instrument communication offers clarity about interest rates at the expense of letting agents miscalculate, and possibly disagree on, the policy's implications for employment and income. The converse is true with target communication. Such

<sup>7</sup>Although [Angeletos and Lian \(2018\)](#) focus on the friction in equilibrium reasoning, or higher-order beliefs, they combine it with inattention, or a friction in first-order beliefs. A similar confounding is present in [Gabaix \(2019\)](#) and [Wiederholt \(2016\)](#). We discuss the separate role of inattention in Section 6.4.

<sup>8</sup>See, for example, [Blinder \(2018\)](#) and [Feroli et al. \(2017\)](#), discussions by then Fed Governor [Jerome H. Powell \(2016\)](#) and then San Francisco Fed President [John C. Williams \(2016\)](#), and the now public transcripts of the December 2012 FOMC meeting.

as a trade-off in anchoring different expectations seems consistent with evidence provided by [Andrade et al. \(2019\)](#) and [Ehrmann et al. \(2019\)](#) from the recent ZLB experience.

Finally, we offer a new perspective on what the Fed's 2012 pivot to target-focused language may have accomplished. Although this switch seems to have been motivated by the Fed's uncertainty about the date at which it could start lifting interest rates, it is anyone's guess how much of this uncertainty had to do with the economy's "hard" fundamentals versus the question of when "confidence" would be restored. Both kinds of uncertainty can justify the switch, but, as we shall explain, only the second predicts that the switch increased policy effectiveness and possibly shortened the time spent at the ZLB.

The last prediction is, of course, highly speculative. But under the lens of our analysis, it is a logical extension of the apparent effectiveness of Mario Draghi's "whatever it takes" speech.

**Other related literature.** Our paper fits into the broader literature on the "forward guidance puzzle," as introduced in [Del Negro, Giannoni and Patterson \(2015\)](#) and [McKay, Nakamura and Steinsson \(2016\)](#). These works have argued that the puzzle is eased by accommodating finite horizons and liquidity constraints (see also [Kaplan, Moll and Violante, 2018](#); [Werning, 2015](#)), but have maintained rational expectations, thus also maintaining the irrelevance of instruments versus targets. We instead break this irrelevance by relaxing rational expectations. But we also point out that the considerations put forward in these works tend to favor target-based forward guidance, for they tend to increase GE multipliers.

The irrelevance results that serve as our point of departure echo related results from the Ramsey literature ([Chari and Kehoe, 1999](#); [Correia, Nicolini and Teles, 2008](#)). From this perspective, a high-level contribution of our paper is to illustrate both how such results hinge on *infinite* depth of rationality and how more "sophisticated" policies can regulate the bite of bounded rationality. The exploration of our insights in other contexts, such as that of Ricardian equivalence ([Barro, 1974](#)), is left for future work.

Our paper also connects to the literature on policy communication spurred by [Morris and Shin \(2002\)](#); see, *inter alia*, [Amador and Weill \(2010\)](#), [Angeletos and Pavan \(2007, 2009\)](#), [Chahrou \(2014\)](#), [Cornand and Heinemann \(2015\)](#), and [Myatt and Wallace \(2012\)](#). We share this literature's emphasis on higher-order beliefs but drop rational expectations. We also change the meaning of policy communication. In this literature, policy communication means revelation of information about the economic fundamentals, holding constant the agents' strategic interaction. In our paper, instead, it means regulation of that interaction via different kinds of "Odyssean" forward guidance (i.e., via different policy commitments).<sup>9</sup>

[Kydland and Prescott \(1977\)](#), [Barro and Gordon \(1983\)](#) and a large follow-up literature study time-inconsistency issues under rational expectations. In the context of instruments and targets, [Atkeson, Chari and Kehoe \(2007\)](#) and [Halac and Yared \(2014\)](#) emphasize a trade-off between commitment and flexibility; and in the context of forward guidance, [Bassetto \(2019\)](#) emphasizes the interaction of commitment and signaling. We take a different approach than all these papers, taking commitment for granted and

---

<sup>9</sup>Among the aforementioned literature, [Angeletos and Pavan \(2009\)](#) and [Cornand and Heinemann \(2015\)](#) deserve special mention. The first allows a policymaker to regulate the agents' strategic interaction but maintain rational expectations and focus, instead, on the use and the aggregation of information. The second introduces Level-k Thinking but abstract from policy and focus, instead, on the social value of information. These papers therefore foreshadow two key ingredients of our analysis, policy and bounded rationality, but do not share our context or results.

focusing on bounded rationality. But we also show how commitment is crucial for regulating the bite of bounded rationality and, conversely, how the latter could itself be the source of time inconsistency.

[Caballero and Simsek \(2019\)](#) study an economy where, from the policymaker's perspective, private agents have “wrong” beliefs about a fundamental such as TFP but “correct” beliefs about one another's behavior. This precludes the kind of “flawed GE reasoning” on which our paper focuses, but shares theme of finding a policy that persuades the public to do the right thing despite its wrong beliefs.

The focus on GE reasoning, or higher-order beliefs, also distinguishes our paper from a larger literature that relaxes rational expectations but preserves the representative-agent assumption, such as [Hansen and Sargent \(2008\)](#) and [Woodford \(2010\)](#). Closer in this respect is [Bergemann and Morris \(2016\)](#), which studies the robustness of a mechanism to the designer's uncertainty about the players' higher-order beliefs.

## 2 Framework and Context

In this section we set up our framework, explain how it captures a New Keynesian economy at the ZLB constraint, and relate our instruments-versus-targets problem to the actual policy experience.

### 2.1 Basic structure and interpretation

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 average of which is denoted by  $K \equiv \int k_i \, di$ . The policymaker controls a policy instrument  $\tau \in \mathbb{R}$  and is interested in manipulating an aggregate outcome  $Y \in \mathbb{R}$ .

The workings of the economy are described by two key equations. The first relates the aggregate outcome to the policy instrument and the aggregate behavior of the agents:

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

where  $\alpha \in (0, 1)$  is a fixed parameter. The second describes the optimal behavior of the typical agent as a function of her expectations of the policy and the outcome:

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

where  $\mathbb{E}_i$  denotes the subjective expectation of agent  $i$  and  $\gamma < 1$  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  $\gamma$  controls how much private incentives depend on expectations of the aggregate outcome and thereby on the choices of others. In this sense,  $\gamma$  parameterizes the GE feedback (more on this shortly). The parameter  $\alpha$ , on the other hand, controls how much of the effect of  $\tau$  on  $Y$  is direct, or mechanical, rather than channeled through expectations.

This structure stylizes the dependence of economic decisions on two kinds of expectations: the expectations of a policy instrument, such as the interest rate set by the central bank, and the expectations of an aggregate outcome, such as unemployment, which itself depends on the aggregate behavior of others. The particular question we are after is how a policymaker can manage these expectations when they are

not “fully” rational (in a sense made precise later on). To shed light on this issue in the most transparent possible way, we have opted to abstract from the micro-foundations and the institutional details of any specific application. But a few concrete translations of our abstract framework are discussed next.

**Forward guidance during a liquidity trap.** Our primary application is forward guidance by a central bank during a liquidity trap. In Appendix B.1, we describe a formal mapping from this context to our abstract model, within a stylized version of the New Keynesian model.

The economy is currently at the ZLB, precluding conventional monetary policy. The instrument  $\tau$  is the extent of monetary loosening after the economy exits the liquidity trap (i.e., after the fundamentally determined natural rate rises above 0 and the ZLB ceases to bind). In particular,  $\tau$  can be mapped to either the negative of the interest rate set in the post-trap period, or to the length of time during which interest rates are kept at zero after the ZLB has ceased to bind. The action  $K$ , on the other hand, is aggregate spending in the middle of the trap and the outcome  $Y$  is an appropriate income measure, the present discounted value of aggregate output both within and outside of the trap.

The mechanics of forward guidance are familiar: lower interest rates tomorrow increase consumption today both through a PE effect (intertemporal substitution) and a GE effect (a feedback loop between aggregate spending and income). The relative strength of these forces are captured by the parameters  $1 - \gamma$  and  $\gamma$ , respectively, in the abstract model’s equation (2). With the applicable micro-foundations,  $\gamma$  is related to the marginal propensity to consume, or the slope of the Keynesian cross. This captures a central theme of the modern Heterogeneous-Agent New Keynesian (HANK) literature. It also illustrates how  $\gamma$  serves more generally as a measure of the GE feedback.

**A Neoclassical alternative.** In a second application, we consider forward guidance about future taxation in a purely Neoclassical environment (Appendix B.2). In this context,  $\tau$  is a future subsidy, or the negative of future taxation,  $K$  is investment,  $Y$  is output, and  $\gamma$  encapsulates the combination of two conflicting GE forces: competition for a scarce resource (labor) and a real aggregate-demand externality.

This example serves two goals. First, it illustrates how a substantially different, flexible-price mechanism could generate the basic structure of equations (1) and (2). And second, it captures within the same context both the case of *positive GE effects* ( $\gamma > 0$ ) and the case of *negative GE effects* ( $\gamma < 0$ ), either of which can hold as the net result of a tug-of-war between the aforementioned two forces.

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

**Parametric restrictions.** The main analysis restricts  $\gamma > 0$  (positive GE feedback), which is the relevant case for our liquidity-trap application. But as shown in Section 5.4, our main result extends to  $\gamma < 0$ . We also require that  $\alpha < \frac{1}{2-\gamma}$ . As explained in Section 4.2, this is necessary for behavior not to be unduly sensitive to beliefs of *infinite* order. Without this restriction, the model is ill behaved: the rational-expectations equilibrium cannot be obtained from iteration of best responses, and tiny relaxations in the agents' depth of knowledge or depth rationality can have arbitrarily large effects on their behavior.

## 2.2 Policy objective

The policymaker minimizes her expectation of the following loss function:

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

where  $\chi \in (0, 1)$  is a fixed scalar and  $\theta$  is a zero-mean random variable. This variable represents the policymaker's assessment of the ideal combination of  $\tau$  and  $Y$ . It could capture a shock to the policymaker's preferences, or her subjective belief about fundamentals that ought to govern policy but, for simplicity, do not directly enter conditions (1) and (2).

The micro-foundations of (3) are not central to our analysis. Nevertheless, Appendix B.1 shows how to derive such a reduced-form objective from first principles in our version of the New Keynesian model.

In this context, we let  $\theta$  be the central bank's assessment of aggregate productivity, or equivalently of the natural rate of output. Because prices are completely rigid, there are no welfare costs from inflation and the welfare loss relative to the first best can be expressed as a function of the output gaps during and right after the trap. It is these gaps that are captured by the quadratic terms in (3).<sup>10</sup>

At the same time,  $\theta$  does *not* enter conditions (1) and (2), because a consumer's optimal spending is invariant to aggregate productivity for given expectations of interest rates and income.<sup>11</sup> By the same token, there is no scope or need for informing agents about  $\theta$  per se. Instead, what consumers care to know is only what the interest-rate policy will be (and its GE implications). In other words, we focus on what is known in the literature following [Campbell et al. \(2012\)](#) as "Odyssean forward guidance."

The following clarification is due at this point. In combination with (1) and (2), (3) embeds the assumption that the policymaker's ideal point, or the first best, is attained under rational expectations equilibrium. This is the right assumption for our theoretical purposes, because it isolates bounded rationality as the only source of a welfare distortion and lets us focus on how policy can regulate its bite.<sup>12</sup> But it is at odds with our application to the Great Recession, where an additional distortion was presumably present because of nominal rigidity and a strictly binding ZLB. We address this issue in Appendix B.1 and return to its policy implications in Section 7.1.

---

<sup>10</sup>As shown in Appendix B.1, the relevant micro-founded gaps map most closely to  $(\tau - \theta)^2$  and  $(K - \theta)^2$ . But this is inconsequential for our results: Theorem 1 readily extends to this case, as shown at the end of its proof in Appendix A.

<sup>11</sup>This property extends to the full version of the New Keynesian model. Aggregate productivity does not enter the Dynamic IS curve. It only enters via the New Keynesian Philips curve, by affecting the representative firm's real marginal cost. This channel is shut down here thanks to the simplifying assumption that prices are completely rigid.

<sup>12</sup>In fact, the essential property is only that any other distortion is *separable* from that caused by bounded rationality. The logic is similar to that in [Correia, Nicolini and Teles \(2008\)](#).

## 2.3 Timing

We close the model by specifying the timing of play in the following three stages:

0. The policymaker observes  $\theta$  and, conditional on that, chooses whether to engage in “instrument communication,” announcing a commitment to set  $\tau = \hat{\tau}$  for the policy instrument, or “target communication,” announcing a commitment to achieve  $Y = \hat{Y}$  for the outcome.
1. Each agent  $i$  hears the policymaker’s announcement, forms expectations (in one of the various ways described in the sequel), and chooses  $k_i$  according to best-response condition (2).
2.  $K$  is publicly observed and the pair  $(\tau, Y)$  is 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  $\tau$  is adjusted so that condition (1) holds with  $Y = \hat{Y}$ .

This structure embeds three assumptions, which are worth emphasizing.

First, the policymaker chooses what to say after observing  $\theta$ . This allows forward guidance to reveal  $\theta$  to the public. But because, as discussed previously, agents do not care to know  $\theta$  per se, such signaling is irrelevant—there is no room for “Delphic” forward guidance.

Second, the policymaker always honors in stage 2 any promise made in stage 0. Forward guidance is thus equated to commitment—which is precisely what the literature calls “Odyssean” forward guidance. But whereas prior work has restricted such guidance to take only one form, instrument-based, we give the policymaker the option to engage in another form, target-based.

Finally, we allow the policymaker to announce a value for *either*  $\tau$  or  $Y$ , but not on a pair of values for *both* of them. As we make clear in Section 5.1, such a commitment is not viable away from the rational-expectations benchmark, because private agents have distorted views about the equilibrium mapping from  $\tau$  to  $Y$ . A similar rationale, shown in Appendix C.2, rules out a commitment to a value for  $K$ .

What remains viable, though, is commitment to a flexible *relation* between  $\tau$  and  $Y$ , namely a function  $f$  such that  $\tau = f(Y; \theta)$ . Section 8 will explain why our insights are robust to this possibility. For the main analysis, we concentrate on the two starker forms of forward guidance introduced above, not only because this is without serious loss for theory, but also because of the real-world context discussed next.

## 2.4 Why instruments vs. targets

As discussed in the Introduction, central banks used forward guidance to an unprecedented extent during the Great Recession. Table 1 offers a few salient examples.

We classify these examples in two categories. The first deals primarily with the policy instrument itself (e.g., talking about how long interest rates will remain low), and is illustrated from various stages of the crisis in the US (row 1), Canada (row 3), and Sweden (row 4). The second is forward guidance that shines the spotlight on a clear target. In rows 2 and 6, respectively from the Fed and the Bank of England, that target is (most clearly) unemployment; in row 5, from the Bank of Japan, it is inflation; and in row 7, from the ECB, it is saving the Euro.

	Institution	Date	Source	Type	Statement Excerpt
(1)	US Federal Reserve	Aug 9, 2011	Policy statement by Committee	Instrument	[T]he Committee decided today to keep the target range for the federal funds rate at 0 to 1/4 percent. The Committee currently anticipates that economic conditions—including low rates of resource utilization and a subdued outlook for inflation over the medium run—are likely to warrant exceptionally low levels for the federal funds rate at least through mid-2013
(2)	US Federal Reserve	Dec 12, 2012	Policy statement by Committee	Target	[T]he Committee decided to keep the target range for the federal funds rate at 0 to 1/4 percent and currently anticipates that this exceptionally low range for the federal funds rate will be appropriate at least as long as the unemployment rate remains above 6-1/2 percent, inflation between one and two years ahead is projected to be no more than a half percentage point above the Committee's 2 percent longer-run goal, and longer-term inflation expectations continue to be well anchored.
(3)	Bank of Canada	Apr 21, 2009	Press release	Instrument	With monetary policy now operating at the effective lower bound for the overnight policy rate, it is appropriate to provide more explicit guidance than is usual regarding its future path so as to influence rates at longer maturities. Conditional on the outlook for inflation, the target overnight rate can be expected to remain at its current level until the end of the second quarter of 2010 in order to achieve the inflation target. The Bank will continue to provide such guidance in its scheduled interest rate announcements as long as the overnight rate is at the effective lower bound.
(4)	Sveriges Riksbank (Sweden)	Apr 21, 2009	Press release	Instrument	The Executive Board of the Riksbank has decided to cut the repo rate by 0.5 percentage points to 0.5 per cent. The lower interest rate and interest rate path are necessary to dampen the fall in production and employment and to attain the inflation target of 2 per cent. The repo rate is expected to remain at a low level until the beginning of 2011.
(5)	Bank of Japan	Apr 4, 2013	Press release	Target	The Bank will achieve the price stability target of 2 percent in terms of the year-on-year rate of change in the consumer price index (CPI) at the earliest possible time, with a time horizon of about two years. In order to do so, it will enter a new phase of monetary easing both in terms of quantity and quality.
(6)	Bank of England	Aug 7, 2013	Letter from Governor Mark Carney	Target	In practice, that means the [Monetary Policy Committee] intends not to raise Bank Rate above its current level of 0.5%, at least until the Labour Force Survey headline measure of unemployment has fallen to a threshold of 7%. While the unemployment rate remains above 7%, the MPC stands ready to undertake further asset purchases if additional stimulus is warranted.
(7)	European Central Bank	Jul 26, 2012	Speech by President Mario Draghi	Target	But there is another message I want to tell you. Within our mandate, the ECB is ready to do whatever it takes to preserve the Euro. And believe me, it will be enough.

Table 1: Forward guidance in recent monetary history.

Of course, the mapping from these examples to our theory is only approximate. Actual policy communication is more complex than speaking about a single instrument or a single target. Even the Fed's announcement of an "unemployment target" in December 2012 was, as originally stated, a more sophisticated contingency on both unemployment and inflation among other indicators (see row 2).

Our focus on a starker instrument-vs-target choice is nevertheless consistent with a plausible real-world *interpretation* of that December 2012 policy shift. As [Alan S. Blinder \(2018\)](#) wrote in a retrospective analysis, much of the public seemed to hear the following much simpler edict:

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

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

During the December 2012 FOMC deliberations, San Francisco Fed President [John C. Williams](#) expressed a complementary idea based on limited attention:

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

Our coming model for belief formation does not get into the deep micro-foundations of “attention” and “salience” but helps formalize the trade-off alluded to above by letting the policymaker be effective at anchoring a *single* expectation, either that of the policy instrument or that of the targeted outcome.

What about communicating more complex objects, like a quantitative reaction function? Macroeconomic models invariably describe monetary policy in a such a manner. This is in part because, during normal times, the Fed laboriously tries to communicate to financial markets the finest details of both its information about the economy and its policy intentions.

But the experience from Mario Draghi’s famous “whatever it takes” proclamation suggests that simpler, shorter, and seemingly vaguer forms of communication may be more effective in offering clarity about what actually matters and in restoring public confidence during turbulent times, when historical experience offers few clues, policy guidance is dearly needed, and the relevant audience is wider.<sup>13</sup>

And in the running example of the December 2012 switch in the Fed’s language, several Fed members saw simpler forms of target communication as the only feasible approximation of an “ideal” rule in an uncertain policy environment. Minneapolis Fed President [Narayana Kocherlakota \(2012\)](#) commented that, absent “the perfect description of a reaction function,” attempting a more complex communication would be “letting the perfect be the enemy of the good.”

The real-world complications that make sophisticated policy rules hard to communicate are outside the scope of our analysis. In the spirit of the above quote, we think of our instrument-versus-target setting as a representation of the “good.” Nonetheless, in Section 8 we will return to a relaxed policy problem that explores “the perfect:” clear and credible communication of a fully flexible, policy reaction function.

## 2.5 Odyssean forward guidance vs. other policy communications

All our example communications in Table 1 included a commitment to a policy plan. Such commitment defines “Odyssean” forward guidance (in the language of [Campbell et al., 2012](#)) and is the sole focus of our paper. But a large fraction of the communications coming out of any modern central bank reveals something about the institution’s information and outlook without explicitly making promises.

Consider, in the Fed’s context, the many sentences about data interpretation and macro outlook that begins all modern FOMC statements, the Summary of Economic Projections, or SEP, for real variables (released quarterly since November 2007), and the “dot plot” of individual FOMC participants’ assessments

---

<sup>13</sup>Retrospective accounts rate this among the most important policy moves in the crisis. See, for example, the November 2018 Bloomberg feature “3 Words and \$3 Trillion: The Inside Story of How Mario Draghi Saved the Euro” ([Randow and Speciale, 2018](#)).

of “appropriate” monetary policy in the future (released with the SEP since early 2011). Such communications provide information about the state of the economy, without necessarily a policy commitment. They relate to “Delphic” forward guidance; they are the focus of the literature spurred by [Morris and Shin \(2002\)](#); and they can be proxied here by communication of  $\theta$ .

As stated previously, we have written a model in which this kind of communication is entirely *ineffective* and *irrelevant*. This not because we doubt its importance in the real world, but because we want to highlight novel issues related to the interaction of commitment and bounded rationality.<sup>14</sup>

### 3 Rational Expectations and Beyond

We now explain why the form of forward guidance is irrelevant in a rational-expectations benchmark that stylizes the “textbook” policy paradigm. This sets the stage for our subsequent departures from it.

#### 3.1 Implementation with rational expectations

Say there is a representative agent, who knows the structure of the economy, observes the policy announcement, and forms rational expectations.<sup>15</sup> In this benchmark,  $\mathbb{E}_i[\cdot] = \mathbb{E}[\cdot|\hat{X}]$  for all  $i$ , where  $\mathbb{E}[\cdot|\hat{X}]$  is the common, rational expectation conditional on announcement  $\hat{X}$ , with  $X \in \{\tau, Y\}$  depending on the form of forward guidance. As a result,  $k_i = K$  for all  $i$  and condition (2) reduces to the following condition for optimal behavior:

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

What matters to the policymaker is the set of combinations of the policy instrument,  $\tau$ , and the outcome,  $Y$ , that can be implemented under each form of forward guidance. This set is defined as follows:

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

This definition embeds Rational Expectations Equilibrium (REE). In the subsequent sections, we will revisit implementability under different solution concepts, that is, different specifications of how agents form expectations about the behavior of others and the equilibrium mapping between  $\tau$  and  $Y$ . In the rest of this section, we formulate and solve the policymaker’s problem in a manner that parallels the analysis in the subsequent sections.

---

<sup>14</sup>To be more explicit about the mapping to the model, we would consider the Fed’s discussion of macro conditions as an example of announcing  $\theta$ , which has no content in our model absent a simultaneous announcement of a policy plan; and we would consider the joint macro and “dot plot” projections in the SEP as an announcement for a “desired value” of  $Y$  and  $\tau$  that, absent commitment, provides no additional information to the public than the announcement of  $\theta$ . Later, we will go one step further and show that a commitment *version* of the dual-variable forecast (i.e., a strong claim that both  $\tau$  and  $Y$  will equal  $\theta$ ) is ill-posed in our model as soon as we relax rational expectations equilibrium.

<sup>15</sup>This is effectively the same as imposing, in a game, complete information and Nash equilibrium.

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

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

The choice  $\mathcal{A} \in \{\mathcal{A}_\tau^*, \mathcal{A}_Y^*\}$  captures the choice of the optimal form of forward guidance (instrument vs target), whereas the choice  $(\tau, Y) \in \mathcal{A}$  captures the optimal pair  $(\tau, Y)$  implemented under the given form of forward guidance. Both of these choices are conditional on  $\theta$ .

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

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

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

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

These properties hold regardless of the form of forward guidance. With instrument communication, we also have  $\tau = \hat{\tau} = \mathbb{E}[\tau|\hat{X}]$ . It follows that, for any  $\hat{\tau}$ , the REE is unique and satisfies  $K = Y = \tau = \hat{\tau}$ . With target communication, on the other hand, we have  $Y = \hat{Y} = \mathbb{E}[Y|\hat{X}]$ . It follows that, for any  $\hat{Y}$ , the REE is unique and satisfies  $K = Y = \tau = \hat{Y}$ . Combining these facts, we infer that, regardless of the form of forward guidance, a pair  $(\tau, Y)$  is implementable if and only if  $\tau = Y$ . We thus reach the following two results, which serve as benchmarks of comparison for our main analysis.

**Proposition 1.** *With rational expectations, the form of forward guidance is irrelevant for implementability:*

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

**Corollary 1.** *With rational expectations, the policymaker attains her first best ( $L = 0$ ) by announcing  $\hat{\tau} = \theta$ , as well as by announcing  $\hat{Y} = \theta$ .*

That  $\mathcal{A}^*$  is a linear locus with slope 1 is a simplifying feature of our environment. The relevant point here is that the implementability constraint faced by the planner is invariant to the form of forward guidance. This invariance mirrors the equivalence of the “dual” and “primal” approaches in the Ramsey literature (Lucas and Stokey, 1983; Chari and Kehoe, 1999; Correia, Nicolini and Teles, 2008): in our setting,  $\mathcal{A}_Y^*$  corresponds to the primal problem, where the planner chooses allocations, and  $\mathcal{A}_\tau^*$  corresponds to the dual problem, where she chooses supporting instruments. This analogy also explains why the irrelevance result presented here is robust to, say, an uncorrectable monopoly distortion that drives the policy optimum away from the true first best, or to the introduction of shocks in equations (1) and (2).<sup>16</sup> Rather, the key assumption that drives both the familiar primal-dual equivalence and the irrelevance result presented above, and that we seek to relax in the rest of the paper, is rational expectations.

<sup>16</sup>The last claim requires that the policymaker's commitment for  $\tau$  or  $Y$  be contingent on these additional shocks, just as they are here on  $\theta$ . Restricting the requisite contingencies brings in the considerations first laid out in Poole (1970), a point we revisit in Section 6.1, but does not upset the logic developed in our baseline analysis.

### 3.2 Unpacking the assumptions

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

**Assumption 1.** *Every agent is rational and attentive in the following sense: she is Bayesian, acts according to condition (2), understands that  $Y$  is determined by condition (1) and that the policymaker has full commitment and acts so as to minimize (3), and is aware of any policy communication.*

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

**Proposition 2.** *The REE benchmark studied in the previous subsection is equivalent to the joint of Assumptions 1 and 2.*

This will become evident in Section 4.3 below, when we show how iterative reasoning converges to the REE under the present assumptions on the agents’ depth of knowledge and rationality but not once we relax them.<sup>17</sup> With this in mind, we next discuss what Assumptions 1 and 2 mean and how they help structure the forms of “bounded rationality” considered in the rest of the paper.

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

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

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

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

## 4 Introducing Higher-Order Doubts or Bounded Rationality

We now introduce our preferred specification of the departure from the rational-expectations benchmark studied above. This specification captures the common core of [Angeletos and Lian \(2018\)](#), [Farhi and Werning \(2019\)](#), [Gabaix \(2019\)](#), and [Garcia-Schmidt and Woodford \(2019\)](#), and allows a direct comparison of our results to theirs. But our ultimate policy conclusion extends to alternative specifications, including one that turns upside down the assumed distortion in equilibrium reasoning (introduced in Section 6.2).

---

<sup>17</sup>The only subtlety is that such convergence requires  $\alpha < \frac{1}{2-\gamma}$  under target communication. Without this restriction, “reasoning does not converge” for the reasons already anticipated (infinite-order beliefs have an undue influence on behavior) and, in this sense, the analysis is not well posed.

## 4.1 The primitive friction

For our baseline analysis, we replace Assumption 2 with the following:

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

Relative to Assumption 2 (which can be nested as  $\lambda = 1$ ), this drops common knowledge of the policy communication and introduces a “crisis of confidence” on whether other agents will respond.

The precise form of Assumption 3 draws from a large literature studying lack of common knowledge in macroeconomics and finance. See [Abreu and Brunnermeier \(2003\)](#), [Morris and Shin \(1998, 2002\)](#), and [Woodford \(2003\)](#) for key early contributions and [Angeletos and Lian \(2018\)](#) and [Wiederholt \(2016\)](#) for recent applications to the ZLB context. This relation will become clear shortly, when we show how  $\lambda < 1$  arrests the response of higher-order beliefs to policy communication. But whereas this literature confounds higher-order doubts with noisy information or rational inattention, Assumption 3 isolates the former friction and equates it with a departure from REE. As further explained in Section 6.4, it is the departure from REE, not noisy information or rational inattention per se, that drives our main result.

In so doing, Assumption 3 admits an immediate re-interpretation in terms of “shallow” reasoning. A large experimental literature documents such a phenomenon in games that resemble our framework, and accommodates it by replacing REE with Level-k Thinking ([Stahl, 1993](#); [Nagel, 1995](#); [Crawford, Costa-Gomes and Iribarri, 2013](#)). This concept has been recently imported to ZLB context by [Garcia-Schmidt and Woodford \(2019\)](#) and [Farhi and Werning \(2019\)](#); see also [Iovino and Sergeyev \(2019\)](#) for an application to Quantitative Easing. The exact mapping for our setting is spelled out in Appendix C.1, but the basic idea is quite simple: doubts about others’ rationality (Level-k Thinking) have nearly identical behavioral implications as doubts about others’ awareness (Assumption 3).<sup>18</sup>

The aforementioned experimental literature thus provides indirect empirical support for the scenario studied here. Additional support can be found in surveys of macroeconomic expectations. As we will show shortly, Assumption 3 amounts to under-reaction of the average expectations of economic activity ( $K$ ) to the relevant news (the policy message). Evidence of such under-reaction has been documented by, *inter alia*, [Coibion and Gorodnichenko \(2012, 2015\)](#).<sup>19</sup> Furthermore, Assumption 3 has a similar flavor as a phenomenon documented by psychologists in various contexts: that people tend to think they are “better than average” (see, for instance, [Alicke and Govorun, 2005](#)).

---

<sup>18</sup>The only notable difference is that our formulation avoids a certain “bug” that emerges when Level-k Thinking is imported from games of strategic complementarity to games of strategic substitutability. See Appendix C.1 for details.

<sup>19</sup>To be precise, this evidence is consistent with the scenario introduced here but does not distinguish between it and noisy information: both could explain the observed under-reaction of average forecasts. [Coibion and Gorodnichenko \(2012, 2015\)](#) thus interpret their evidence in terms of noisy information, while [Gabaix \(2019\)](#) interprets the same evidence in terms of “cognitive discounting,” a cousin of Assumption 3. As explained in Appendix D, the additional evidence provided in [Bordalo et al. \(2018\)](#) and [Kohlhas and Broer \(2018\)](#) point towards a *combination* of the two frictions that preserves our main result.

Last but not least, Assumption 3 is a close cousin of the form of “cognitive discounting” put forward in [Gabaix \(2019\)](#). The relation is evident in the similarity between the form of myopia captured in that paper and the one captured in [Angeletos and Lian \(2018\)](#). In short, for our purposes, higher-order doubts, cognitive discounting, and Level-k Thinking are all one and the same.

## 4.2 Forward guidance and strategic interaction

To start shedding light on how forward guidance interacts with the friction introduced above, and more generally with the reasoning agents must make about the behavior of others, we now show how the two policy options induce fundamentally different games among the agents. This is actually true even in the REE benchmark studied in the previous section, but it is consequential only outside of it.

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

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

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

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

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

The level of the best responses in this game is controlled by  $\hat{\tau}$ , the announced value of the policy instrument, while their slope is given by  $\delta_\tau$ . The latter encapsulates how much aggregate behavior depends on the forecasts agents form about one another’s behavior relative to the policy instrument.<sup>20</sup>

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

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

The first equation highlights that, under target communication, agents need to predict the monetary policy that will support the announced target. The second shows that, for given an announced target  $\hat{Y}$ , the expected looseness of monetary policy is a *decreasing* function of the expected  $K$ : an agent who is pessimistic about early consumption expects the policymaker to use a looser monetary policy (or longer period at the ZLB) in order to meet the given output target. Combining these two equations, we reach the following counterpart to Lemma 1.

<sup>20</sup>The game obtained above is similar to the static beauty-contest games studied in, *inter alia*, [Morris and Shin \(2002\)](#), [Woodford \(2003\)](#), and [Angeletos and Pavan \(2007\)](#), with  $\hat{\tau}$  corresponding to the “fundamental,” or the shifter of best responses, in these papers. But whereas these papers let the fundamental be observed with noise, here  $\hat{\tau}$  is perfectly observed. The role of noise, or inattention, is discussed in Section 6.4.

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

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

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

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

**Corollary 2.** Switching from instrument communication to target communication changes the game played by the agents from one of strategic complementarity to one of strategic substitutability.

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

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

for  $\delta_\tau \in [0, 1]$  and  $\delta_Y \leq 0$ . Given the restriction  $\alpha < \frac{1}{2-\gamma}$ , assumed from here on out, we have further that  $\delta_X \in (-1, 1)$  for both  $X \in \{\tau, Y\}$ .<sup>21</sup>

In the REE benchmark studied before, both of these games end up yielding the same implementability restrictions on  $\tau$  and  $Y$ . But this is not the case anymore. Below, we first show how the primitive friction under each form of forward guidance translates to the *same* under-reaction in the beliefs of  $K$ , but a *different effect* on the implementability restrictions.

### 4.3 Beliefs or reasoning

The next result verifies that Assumption 3 amounts to introducing the same under-reaction in the beliefs of  $K$ , and in this sense the same lack of confidence about the behavior of others, under all policies.

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

---

<sup>21</sup>The restriction  $\delta_X \in (-1, 1)$  means that the equilibrium of both games can be obtained via iteration of best responses for any  $\lambda \leq 1$ , or that beliefs of arbitrarily high order do not have an explosive impact on behavior. Without this restriction, the rational-expectations outcome itself is extremely fragile. For instance, Level-k thinking fails to recover it in the limit as  $k \rightarrow \infty$ . See also Lemma 5 in the Appendix for the calculation of why  $\delta_Y > -1$  maps to  $\alpha < \frac{1}{2-\gamma}$ .

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

A more detailed derivation helps reveal the underlying equilibrium reasoning. By iterating the best responses, we can express the expectation of  $K$  as a weighted average of the higher-order beliefs about  $X$ :

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

Because we have let the typical agent believe that only a fraction  $\lambda$  of the other agents is aware of the policy message, her second-order beliefs satisfy

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

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

$$\bar{\mathbb{E}}^h[X] = \lambda^h \hat{X}. \quad (11)$$

Relative to the frictionless benchmark (nested here with  $\lambda = 1$ ), higher-order beliefs are therefore more rigid (i.e., anchored to 0), and the more so the higher their order. It follows that  $\bar{\mathbb{E}}[K]$ , which is a weighted average of the beliefs of order  $h = 2$  and above, is also rigid. And because  $K$  is itself a weighted average of  $\hat{X}$  and  $\bar{\mathbb{E}}[K]$ ,  $K$  responds more strongly than  $\bar{\mathbb{E}}[K]$ , or expectations are more rigid than actions.

These derivations help understand not only how Assumption 3 equals rigid higher-order beliefs but also how it connects to Level- $k$  Thinking. The latter approach amounts to having the beliefs of order  $h \leq k$  move the same as first-order beliefs, or  $\hat{X}$  itself, and the beliefs of order  $h > k$  be pegged at 0. The math is a bit different (see Appendix C.1) but the essence is the same.

#### 4.4 Attenuation vs. amplification

Although the primitive friction, in terms of how agents reason about the behavior of others, is invariant across the two policy strategies, its footprint on the expectations of  $\tau$  and  $Y$  and thereby on actual behavior is qualitatively different. Indeed, aggregating (9) and using  $\bar{\mathbb{E}}[K] = \lambda K$ , we reach the following result.

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

$$K = \kappa_X \hat{X} \quad \text{with} \quad \kappa_X \equiv \frac{1 - \delta_X}{1 - \lambda \delta_X}, \quad (12)$$

where  $X \in \{\tau, Y\}$  depending on the form of forward guidance. Furthermore,  $\kappa_\tau < 1 < \kappa_Y$  for every  $\lambda < 1$ ; and the distance of either  $\kappa_\tau$  or  $\kappa_Y$  from 1 increases as  $\lambda$  falls.

In simpler words:

**Corollary 3.** *Higher-order doubts or bounded rationality ( $\lambda < 1$ ) attenuates the actual response of  $K$  under instrument communication, and amplifies it under target communication. Furthermore, a larger friction (lower  $\lambda$ ) translates to larger attenuation in the first case and to larger amplification in the second case.*

This is a direct implication of our earlier mapping of the form of forward guidance to the form of strategic interaction. When agents play a game of strategic complementarity (which is the case under instrument-based guidance), the belief that others will be inattentive or otherwise irresponsive causes each agent to respond less than in the frictionless benchmark. When instead agents play a game of strategic substitutability (which is the case under target-based guidance), the same belief causes each agent to respond more than in the frictionless benchmark.

This result qualifies the related common finding of [Angeletos and Lian \(2018\)](#), [Farhi and Werning \(2019\)](#), [Gabaix \(2019\)](#), [Garcia-Schmidt and Woodford \(2019\)](#) and [Wiederholt \(2016\)](#). These works have argued that a similar friction as that studied here arrests the response of aggregate spending to forward guidance at the ZLB. But whereas these works restrict forward guidance to be *instrument-based*, our result clarifies that this prediction can be reversed with *target-based* forward guidance.<sup>22</sup>

## 5 Optimal Forward Guidance

This section contains our main results. We first spell out the implications of the friction for the combinations of  $\tau$  and  $Y$  that are implementable under each the form of forward guidance. We next show how the implementable sets vary with the GE feedback. We finally characterize the optimal policy.

### 5.1 Implementability

Combining Lemma 4, which characterizes the actual response of  $K$  to any policy announcement, with the policymaker's freedom to announce  $\hat{\tau}$  or  $\hat{Y}$  as any number, we reach the following result:

**Proposition 3** (Implementability). *Let  $\mathcal{A}_\tau$  and  $\mathcal{A}_Y$  denote the sets of the pairs  $(\tau, Y)$  that are implementable under, respectively, instrument and target communication. Then,*

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

where

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

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

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

<sup>22</sup>The kind of higher-order doubts or bounded rationality we have captured via Assumption 3 is the *sole* friction in [Farhi and Werning \(2019\)](#) and [Garcia-Schmidt and Woodford \(2019\)](#). But this is not the case for [Angeletos and Lian \(2018\)](#), [Gabaix \(2019\)](#) and [Wiederholt \(2016\)](#), which combine the relevant rigidity in higher-order beliefs with a rigidity in first-order beliefs, due the inclusion of noisy information, inattention, or "sparsity." This additional friction, which we allow for in Section 6.4, contributes toward a less effective forward guidance, or a lower response of  $K$  to  $\hat{X}$ , under *both* communication strategies.

**Proposition 4.** *The following properties hold for  $(\mu_\tau, \mu_Y)$ :*

- (i)  $\mu_\tau(\lambda, \gamma) \geq 1$  with equality only when  $\lambda = 1$  or  $\gamma = 0$ .
- (ii)  $\mu_Y(\lambda, \gamma) \leq 1$  with equality only when  $\lambda = 1$  or  $\gamma = 1$ .
- (iii)  $\mu_\tau(\lambda, \gamma)$  increases in  $\lambda$  and  $\mu_Y(\lambda, \gamma)$  decreases in  $\lambda$ .

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  $\lambda$ ) increases the slope, meaning that a higher variation in  $\tau$  is needed to attain any given variation in  $Y$ . With target communication, the opposite is true. The magnitude of the deviation, as measured by the absolute value of  $\mu_X(\lambda) - 1$ , increases in both cases, but the sign is different.

Recall that the policymaker picks both the form of forward guidance and the actual policy after observing  $\theta$ . It follows that, conditional on  $\theta$ , he can pick any pair  $(\tau, Y)$  in the union of  $\mathcal{A}_\tau$  and  $\mathcal{A}_Y$ . But since this union does not contain the first-best pair  $(\tau, Y) = (\theta, \theta)$  except for the zero-probability event in which  $\theta = 0$ , the following is immediate.

**Corollary 4.** *When  $\lambda < 1$ , the first best is not attainable under either instrument or target communication.*

This motivates the policy problem studied in the next subsection, which is to find which form of forward guidance minimizes the deviation from the first best. But let us first clarify the following points.

The two forms of forward guidance under consideration equate the communication that takes place in stage 0 with a commitment to implement a particular value for *either*  $\tau$  or  $Y$  in stage 2. Could perhaps the problem be bypassed by having the policymaker commit on a pair of values for *both*  $\tau$  and  $Y$ ? The answer is no—and it is instructive to understand why.

Suppose, towards a contradiction, that the policymaker promises to deliver  $\tau = Y = \theta$  and that the agents believe it. Then, the typical agent, who is himself rational and attentive, will find it optimal to play  $k = \theta$ . But then this agent will also reason as follows:

“If all agents who are like me (i.e., both rational and attentive) play the same, as they should, then the aggregate action will be  $K = \lambda\theta$ , because only a fraction  $\lambda$  of the population are like me and the rest are instead stuck at playing  $k = 0$ . But condition (1), a primitive restriction on that I know to be true, implies that  $K = \lambda\theta$  is necessarily inconsistent with  $\tau = Y = \theta$ . So, the original guidance is not credible.”

We have thus reached a contradiction: a commitment to deliver on both  $\tau$  and  $Y$  does not work.<sup>23</sup>

This clarifies why in the first place we restricted the policymaker to communicate a commitment for *either*  $\tau$  or  $Y$ , as opposed to both. In Appendix C.2.2, we further show that expanding the message space in

---

<sup>23</sup>To put this in context, a central bank cannot jointly promise to deliver both the natural rate of interest and the natural rate of unemployment ( $\tau = Y = \theta$ ) in a way that a public could believe despite the latter's incorrect belief that aggregate spending will *not* be at its natural level ( $K \neq \theta$ ). A similar argument shows, in Appendix C.2.1, why the policymaker cannot commit to deliver  $K = \theta$ ; he simply does not have the means to do this.

state 0 (e.g., directly revealing the state  $\theta$ , or providing a “rationale” for the policy plan) does not alter the implementability constraints characterized in Proposition 3. What is left open is only the possibility that the policymaker commits to a flexible *relation* between  $\tau$  and  $Y$ . We revisit this possibility in Section 8. For now, we note that the above argument in reverse proves the following: if the first best is to be attained by *any* means, it has to be that the public believes that it is *not* attained.

This gets to the heart of our departure from rational expectations equilibrium. When agents expect  $\tau = Y = \theta$ , they play  $K = \theta$ , which in turn allows the policymaker to achieve  $\tau = Y = \theta$ , in line with the public’s initial forecast. When  $\lambda = 1$ , one can close the loop to establish the existence of a rational expectations equilibrium where the public believes the government and the expectation  $\tau = Y = \theta$  is fulfilled. But once  $\lambda < 1$ , the loop doesn’t close, precisely because the public has a distorted view of the world.

The policy problem thus boils down to the following question:

*How can the policymaker persuade the public to do the right thing (i.e., play an action close to the first best) despite their wrong understanding of equilibrium?*

## 5.2 GE feedback and the optimal form of forward guidance

To answer the aforementioned question, we now turn attention to how the introduced friction interacts with  $\gamma$ , or the strength of the GE feedback. The next Proposition establishes that, as  $\gamma$  increases, the distortion of the implementability constraint is *exacerbated* under instrument communication ( $\mu_\tau$  gets further away from  $\mu^*$ ), whereas it is *alleviated* under target communication ( $\mu_Y$  gets closer to  $\mu^*$ ).

**Proposition 5.** *Fix any  $\lambda \in (0, 1)$  and  $\alpha \in (0, 1)$ . As  $\gamma$  increases, both  $\mu_\tau(\lambda, \gamma)$  and  $\mu_Y(\lambda, \gamma)$  increase. Furthermore,  $\mu_\tau(\lambda, 1) > 1$  and  $\mu_Y(\lambda, 1) = \mu^* = 1$ , whereas  $\mu_\tau(\lambda, 0) = \mu^* = 1$  and  $\mu_Y(\lambda, 0) < 1$ .*

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

The converse is true when the GE effect is maximal, or  $\gamma = 1$ . Behavior is then pinned down exclusively by expectations of the outcome:  $k_i = \mathbb{E}_i Y$  for all  $i$ . The distortion is then eliminated by, and only by, communicating a target  $\hat{Y}$  for the outcome (“offering clarity about the unemployment target”).

What about the intermediate case in which  $\gamma \in (0, 1)$ ? In this case, neither strategy completely eliminates the need to reason about others’ behavior. Furthermore, because the implementability constraint is tilted away from the frictionless counterpart under both strategies, it is no more optimal to announce either  $\hat{\tau} = \theta$  or  $\hat{Y} = \theta$ . Instead, conditional on the form of forward guidance, the optimal announcement and the corresponding outcome are both affected by the friction, in a manner that we explain in next

subsection. Still, the monotonicities of the implementability constraints with respect to  $\gamma$  documented in Proposition 5 suggest that the overall distortion is minimized by target communication if and only if the GE effect is strong enough. The next theorem verifies this intuition.

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

The basic idea is that anchoring expectations about the object that has the biggest *direct* effect on agents' behavior (which is  $\tau$  when  $\gamma$  is small and  $Y$  when  $\gamma$  is large) helps minimize the distortion caused by any mistakes in their reasoning about *indirect* equilibrium effects. A detailed proof is provided in Appendix A. Below we sketch the main argument. We also characterize the pair  $(\tau, Y)$  that is implemented by the optimal communication strategy for each realization of  $\theta$ .

### 5.3 Complete characterization of optimal policy

Given  $\theta$ , 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:

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

Focus on  $\lambda < 1$  and  $\gamma \in (0, 1)$ , and let  $(\mathcal{A}^{\text{sb}}, \tau^{\text{sb}}, Y^{\text{sb}})$  be the (unique) triplet that attains the minimum.

Given the assumed specification of  $L$  and the characterization of the implementability sets in Proposition 3, we can restate the choice of the form of forward guidance as the choice of a slope  $\mu \in \{\mu_\tau(\lambda), \mu_Y(\lambda)\}$  for the equilibrium mapping between  $\tau$  and  $Y$ . Letting  $r \equiv \tau/\theta$  and substituting the implementability constraint we reach the following simpler representation of the policymaker's problem:

$$\min_{\mu \in \{\mu_\tau(\lambda), \mu_Y(\lambda)\}, r \in \mathbb{R}} [(1 - \chi)(r - 1)^2 + \chi(r\mu^{-1} - 1)^2] \quad (13)$$

This makes clear that the optimal form of forward guidance is the same for all realizations of  $\theta$  and lets  $r$  identify the optimal covariation of  $\tau$  with  $\theta$ .

It is simple to solve for the optimal  $r$  in closed form and arrive at the following representation of the policymaker's loss as a function of  $\mu$  alone:

$$\mathcal{L}(\mu) \equiv \min_{r \in \mathbb{R}} [(1 - \chi)(r - 1)^2 + \chi(r\mu^{-1} - 1)^2] = \frac{\chi(1 - \chi)(1 - \mu)^2}{\mu^2(1 - \chi) + \chi},$$

which is a U-shaped function of  $\mu \in (0, \infty)$ , with a minimum equal to 0 and attained at  $\mu = 1$  (the frictionless case). The interpretation of this loss function is simple. The closer  $\mu$  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  $\mu$  moderates the distortion but does not upset the property that the loss is smaller the closer  $\mu$  is to 1.

The optimal form of forward guidance can now be found by studying which of the two feasible values of  $\mu$  yields the smallest value for  $\mathcal{L}(\mu)$ . Varying  $\gamma$  changes these two values without affecting the loss incurred

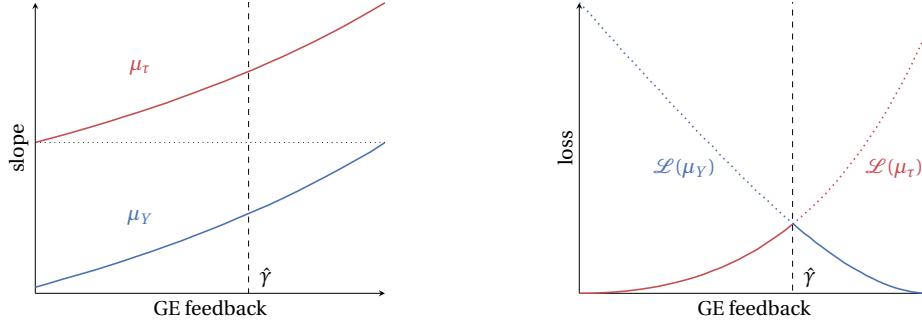


Figure 1: Slope of implementability constraint (left) and loss (right) as function of GE feedback  $\gamma$ .

from any given  $\mu$ . In particular, raising  $\gamma$  drives  $\mu_\tau$  further way from 1 and brings  $\mu_Y$  closer to 1. It follows that  $\mathcal{L}(\mu_\tau)$  is an increasing function of  $\gamma$ , 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  $\gamma$  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  $\gamma > \hat{\gamma}$ .

Figure 1 illustrates this argument in a graph, with the slopes ( $\mu_\tau, \mu_Y$ ) in the left panel and the loss functions ( $\mathcal{L}(\mu_\tau), \mathcal{L}(\mu_Y)$ ) on the right. 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 feedback is strong enough.

The next result completes the characterization of the optimal policy by describing the precise guidance offered and the result obtained for any given  $\theta$ .

**Proposition 6.** *For any  $\lambda \in [0, 1]$  and any  $\gamma \in [0, 1]$ , let  $(r^{sb}, \mu^{sb})$  be the solution to problem (13) and let  $\varphi^{sb} \equiv r^{sb}/\mu^{sb}$ . Then,  $\mu^{sb} = \mu_\tau(\lambda, \gamma)$  if  $\gamma < \hat{\gamma}$  and  $\mu^{sb} = \mu_Y(\lambda, \gamma)$  if  $\gamma > \hat{\gamma}$ . Furthermore:*

- (i) *If  $\gamma < \hat{\gamma}$ , the policymaker commits to  $\tau = r^{sb}\theta$  in stage 0, and obtains  $Y = \varphi^{sb}\theta$  in stage 2. If instead  $\gamma > \hat{\gamma}$ , she commits to  $Y = \varphi^{sb}\theta$  in stage 0, and meets this target with  $\tau = r^{sb}\theta$  in stage 2.*
- (ii)  *$r^{sb}$  displays a downward discontinuity at  $\gamma = \hat{\gamma}$ , is continuous and strictly increasing in  $\gamma$  everywhere else, and satisfies  $r^{sb} > 1$  for  $\gamma \in (0, \hat{\gamma})$  and  $r^{sb} < 1$  for  $\gamma \in (\hat{\gamma}, 1)$*
- (iii)  *$\varphi^{sb}$  displays an upward discontinuity at  $\gamma = \hat{\gamma}$ , is continuous and strictly decreasing in  $\gamma$  everywhere else, and satisfies  $\varphi^{sb} < 1$  for  $\gamma \in (0, \hat{\gamma})$  and  $\varphi^{sb} > 1$  for  $\gamma \in (\hat{\gamma}, 1)$*

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

## 5.4 Additional properties and an illustration

We conclude this section with a few clarifications and a translation of the threshold  $\hat{\gamma}$  within the context of the New Keynesian model.

**Negative GE feedback.** The restriction to positive GE feedback, or  $\gamma > 0$ , is consistent with the liquidity-trap application. But the opposite scenario is possible in other contexts: for instance, in the Neoclassical investment example studied in Appendix B.2,  $\gamma < 0$  is obtained if wage pressure overcomes the aggregate demand externality. Theorem 1 directly extends to this case. This is readily verified by noting that the proof of Theorem 1 in Appendix A does not actually use  $\gamma > 0$ .<sup>24</sup> We thus have that target communication is optimal if and only if the GE feedback is both positive in sign and large enough in magnitude.

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

**Corollary 5** (Optimal Forward Guidance with Vanishing Frictions). *In the limit of zero friction, or  $\lambda \rightarrow 1$ , the critical GE feedback from Theorem 1 depends only on  $\alpha$ :  $\hat{\gamma}_{vf} \equiv \lim_{\lambda \rightarrow 1} \hat{\gamma} = \frac{1}{2-\alpha} \in (\frac{1}{2}, 1)$ .*

The condition  $\gamma \leq 1/(2-\alpha)$  distills our analysis to a simple check of whether GE feedback is sufficiently strong relative to the policymaker's ability to honor commitments (which itself is inversely related to  $\alpha$ ).

**An illustration of  $\hat{\gamma}$  in the New Keynesian model.** We now apply Corollary 5 to translate our main result in the New Keynesian model of Appendix B.1. In this model,  $\gamma$  is positively related to the representative agent's marginal propensity to consume (MPC) out of income and maps cleanly to the slope of a Keynesian cross; but a single deep parameter controls both the previous moment and the parameter  $\alpha$ , making it unclear how optimal policy relates to the MPC. Still, as shown in Appendix C.4, the optimal policy switches to target communication for a sufficiently high MPC. For a “calibration” that matches a three-year liquidity trap, the threshold value for an annualized MPC is 0.21.

Needless to say, this calculation is purely illustrative: our model is too stylized to allow quantitative statements. But an MPC above 0.21 is not unreasonable: the available evidence suggests an average, across households, MPC of 0.30 or above (Jappelli and Pistaferri, 2010). We therefore view this back-of-the-envelope exercise as a hint for the possible value of applying our insights to richer, quantitative frameworks that emphasize high MPCs for a large fraction of the population (Kaplan and Violante, 2014) and a GE-focused transmission of monetary policy (Kaplan, Moll and Violante, 2018; Werning, 2015).

## 6 Robustness

Before re-introducing the policy context in Section 7, we pause here to discuss the robustness of our main insights to confounding shocks and imperfect policy control (Section 6.1), arbitrary mistakes in GE reasoning (Section 6.2), endogenous depth of reasoning (Section 6.3), and inattention (Section 6.4).

---

<sup>24</sup>The proof only uses the weaker restriction  $|\delta_X| < 1$ , which means that beliefs of arbitrarily high order have a vanishing effect on behavior. See Appendix C.3 for additional details.

## 6.1 Imperfect control and additional shocks

Much of the contemporary discussion of “instrument and targets” follows the durable logic of Poole (1970): that the optimal implementation device is the one best hedged against confounding shocks. Although the logic developed here is different, such hedging is front-and-center in actual policy design. In Appendix C.5, we thus consider an extension of our baseline model that *combines* bounded rationality with two Poole-like elements: (i) “uncertainty about future fundamentals,” or the existence of an unobserved shock to  $Y$  in equation (1); and (ii) “imprecise implementation,” or noisy measurement of  $Y$  or  $\tau$ .

The first element needs no motivation in the real world. But it is consequential in the theory only insofar as the commitments made for  $\tau$  or  $Y$  cannot be directly contingent of the future shocks, which is what Poole (1970) assumes. If instead these commitments could be fully contingent, the trade-off studied by Poole would disappear and the irrelevance result of our REE benchmark would apply.

The second element might relate to the imprecise mapping from primitive conditions to headline macro statistics. As an example, the Fed was very concerned in March 2014 that unemployment figures were falling toward the pre-committed 6.5% threshold for the “bad reason” that individuals were leaving the labor force, while primitive labor market conditions were not improving so much (Blinder et al., 2017).

These elements now emerge as additional determinants of the optimal policy. For instance, committing on a target for  $Y$  helps insure against unwanted future fluctuations in aggregate demand. But, unlike our approach, such Poole-like considerations do *not* induce a dependence of the optimal policy on the relative importance of PE and GE considerations, as capture by the structural parameter  $\gamma$ .

**Proposition 7.** *In the presence of Poole-like considerations, the optimal choice between instrument and target communication depends on  $\gamma$  only when  $\lambda < 1$ , and Theorem 1 continues to hold.*

What is more, the logic that the instrument-vs-target choice is irrelevant for implementability in the rational-expectations benchmark (but not away from it) generalizes in the following “average” sense:

**Proposition 8.** *Let  $\mathbb{E}_p[\tau]$  and  $\mathbb{E}_p[Y]$  be the policymaker’s expectation of  $\tau$  and  $Y$  at stage 0, where the expectation is taken over the possible realizations of the future shocks or measurement errors. When  $\lambda = 1$ ,  $\mathbb{E}_p[\tau] = \mu^* \mathbb{E}_p[Y]$  with  $\mu^* = 1$ , regardless of the policymaker’s strategy. When instead  $\lambda < 1$ ,  $\mathbb{E}_p[\tau] = \mu_\tau \mathbb{E}_p[Y]$  under instrument communication and  $\mathbb{E}_p[\tau] = \mu_Y \mathbb{E}_p[Y]$  under target communication, with  $\mu_\tau > 1 > \mu_Y$ . Furthermore,  $\mu_\tau$  and  $\mu_Y$  are the same as in the baseline model.*

With rational expectations ( $\lambda = 1$ ), and from the policymaker’s perspective at the time she has to choose whether to commit on a value for  $\tau$  or a value for  $Y$ , there continues to exist *no* trade-off in terms of how steep or flat the implementability constraint is. What changes between these two choices is only the extent of insurance provided against future shocks or measurement error. By contrast, with bounded rationality ( $\lambda < 1$ ), implementability is fundamentally altered: the average relation between  $\tau$  and  $Y$  depends on the policy rule, and on its interaction with  $\gamma$ , essentially in the same way as in our baseline model.

A corollary of this result is that, when and only when  $\lambda < 1$ , the switch from instrument to target communication is associated with a reduction in the expected value of  $\tau$  needed to achieve the desired target

in  $Y$ . This anticipates a point we make in Section 7: the 2012 shift in the Fed’s communication strategy may have, unintentionally but favorably, help shorten the length the economy had to spend at the ZLB.

## 6.2 Arbitrary mistakes in reasoning

Our main specification equated agents’ bounded rationality to lack of common knowledge of others’ awareness and rationality, which amounted to under-estimation of GE effects. As explained in Section 4, this captures the common core feature of a theoretical literature we build on. But a policymaker themselves could be forgiven for not having complete confidence that these theories are correct.

Consider the following two alternative stories. The first is that individual agents will be startled and over-estimate the response of the economy to the policy news. The second is that agents’ behavior will be swayed by animal spirits (extrinsic waves of optimism and pessimism about the behavior of others) or by purely random errors in their reasoning about GE effects (such as, perhaps, those caused by a mis-specified belief about the structure of the economy). To capture these possibilities in a structured yet flexible manner, we now consider the following specification of how the beliefs about others’ behavior:

**Assumption 4** (General distorted reasoning). *Average beliefs satisfy  $\bar{\mathbb{E}}[K] = \lambda K + \sigma \varepsilon$  for some  $\lambda > 0$  (possibly  $\lambda > 1$ ) and  $\sigma \geq 0$ , where  $\varepsilon$  is a unit-variance noise term unknown to the policymaker and independent of the policy announcement.*

Unlike the main analysis, the friction is now introduced directly in the expectations of  $K$  as opposed in the depth of agents’ knowledge and rationality. This shortcut lets us focus on how the friction matters for behavior as opposed to how it is micro-founded. But the missing details can easily be filled in.

For instance, letting  $\lambda > 1$  is akin to modifying the higher-order beliefs in Assumption 3 in the following way: let agents believe that with positive probability others will be “startled” and over-react to the policy message instead of being “sleepy” and under-reactive.<sup>25</sup> Basically the same applies to Level-k Thinking with a level-0 belief of the form  $\eta \hat{X}$ , for some  $\eta > 1$ . And just as  $\lambda < 1$  directly captures the form of “cognitive discounting” assumed in Gabaix (2019),  $\lambda > 1$  directly captures the opposite bias, or a form of “cognitive hyperopia.” As for  $\sigma > 0$ , this can be micro-founded by introducing either an “erratic” level-0 belief, or shocks to higher-order beliefs along the lines of Angeletos and La’O (2013); and it can be interpreted as animal spirits or purely random errors in equilibrium reasoning.

The upshot for implementable sets is the following:

**Proposition 9.** *A pair  $(\tau, Y)$  is implementable if and only if*

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

where  $X \in \{\tau, Y\}$  indexes the form of forward guidance,  $\mu_\tau(\lambda, \gamma)$  and  $\mu_Y(\lambda, \gamma)$  are as in Proposition 3, and

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

<sup>25</sup>The only twist is that  $\lambda$  is no more the aforementioned probability, but rather a mixture of this probability and the perceived over-reaction of others.

Compared to the case with under-reactive beliefs ( $\lambda < 1$ ), the case with over-reactive beliefs ( $\lambda > 1$ ) flips the *sign* of distortion: forward guidance is now amplified under instrument communication ( $\mu_\tau < 1$ ) and attenuated under target communication ( $\mu_Y > 1$ ). Nevertheless, the comparative statics of the *size* of distortion with respect to the strength of the GE feedback remain the same: as  $\gamma$  increases, the distortion under instrument communication gets larger and that under target communication gets smaller.

The distortions induced by random perturbations ( $\sigma > 0$ ) share this comparative static, too. The common mechanism is that a higher  $\gamma$  increases the dependence of behavior on any mistakes about  $Y$  relative to any mistakes about  $\tau$ . This is true regardless of whether these mistakes are positively correlated, negatively correlated, or uncorrelated with the announcement itself.

Putting these ideas together, it is easy to show that our main policy result and the intuition about minimizing the distortion also remain for any  $\lambda$  and  $\sigma$ .

**Proposition 10.** *When  $\lambda = 1$  and  $\sigma = 0$ , the optimal form of forward guidance is indeterminate. When instead  $\lambda \neq 1$  and/or  $\sigma \neq 0$ , Theorem 1 continues to hold: there exists  $\hat{\gamma} \in (0, 1)$  such that target communication is optimal if and only if  $\gamma > \hat{\gamma}$ .*

Finally, consider the following re-interpretation of the previous results from a policymaker's perspective. Regardless of what are the values of  $(\lambda, \sigma)$  in the context of Assumption 4, the loss functions of instrument and target communication are monotone in  $\gamma$  and the main "threshold result" is true. By implication, if a policymaker has a non-degenerate prior over  $(\lambda, \sigma)$ , in expectation the result must continue to hold. This is summarized in the following corollary:

**Corollary 6** (Robustness to unknown distortions). *Assume that, from the policymaker's perspective, the parameters  $(\lambda, \sigma) \in \mathbb{R}_+^2$  are random and drawn from some non-degenerate prior distribution  $\pi \in \Delta(\mathbb{R}_+^2)$ . Then Theorem 1 continues to hold.*

In this sense, a policymaker who suspects that the public has the wrong model of how policy works in GE but is not sure of the precise model thereof could still apply our main result.

### 6.3 Endogenous depth of reasoning

So far we have treated the parameters controlling the depth of reasoning, or  $\lambda$  from Assumption 3, as exogenous. We now discuss two ways in which  $\lambda$  might vary with the environment or the policy.

**Learning and rational expectations.** Consider first that agents learn to be "more rational" over time by playing the same game repeatedly, noticing any systematic forecast errors, and adjusting beliefs accordingly. The literature has documented such learning in the laboratory and has captured it by letting the depth of reasoning increase with the rounds of play (e.g., [Crawford, Costa-Gomes and Iribarri, 2013](#); [Mauersberger and Nagel, 2018](#)). In our context, this would suggest letting  $\lambda$  get closer to 1 with the passage of time and the accumulation of experience.

But this seems of little relevance for our application to the Great Recession and the unconventional policies central banks resorted to because of the ZLB constraint. This was an unfamiliar set of circumstances, *not* an easily learned environment for which agents would have well-tuned models. Like [Farhi](#)

and Werning (2019), we thus think that the kind of shallow reasoning observed in “early” rounds of play in laboratory experiments, or a value of  $\lambda$  well below 1, are the right place to start for modeling this scenario.

Furthermore, even when people have the chance of repeated play, “convergence to rational expectations” is not necessarily a good approximation of reality. The survey evidence reviewed in Appendix D suggests that  $\lambda < 1$  is applicable even in the “stationary” context of regular macroeconomic forecasting.

**Costly contemplation.** Consider next the possibility that agents decide how much effort to invest into equilibrium reasoning. The natural prediction of such a costly-contemplation model is that reasoning is highest when it has the most value or influence on decisions. Laboratory experiments by Alaoui and Penta (2016) and Gill and Prowse (2016) uncover related sensitivity of “depth of play” to contextual shifters that regulate the private return to strategic reasoning.

In our context, this plausibly translates to letting  $\lambda$  get closer to 1 with  $|\delta_X|$ , or the weight on the actions of others in an agent’s best response. This in turn means that  $\lambda$  may increase with  $\gamma$  under instrument communication and decrease under target communication. Such an asymmetric effect of  $\gamma$  on  $\lambda$  disrupts the monotonicities of the loss functions used to prove Theorem 1. But the basic idea survives, and so does the variant of this theorem that we prove later on in the context on flexible policy rules (Proposition 12).

What is more, our take-home message is likely to be *reinforced* if the public’s cognitive costs are internalized by the policymaker. Reducing the amount of required contemplation about the behavior of others could then be a “win-win:” it reduces distortions (the main point of Theorem 1) *and* prevents direct disutility associated with thinking about equilibrium.

## 6.4 Inattention

An important simplification in our model is that agents hear forward guidance perfectly clearly. This contrasts with ample evidence of inattention and compatible theories (Sims, 2003; Gabaix, 2014).

Consider the simplest such model, with Gaussian signal extraction. The fundamental  $\theta$  is Gaussian with mean 0; the policy announcement  $X$ , which is linear in  $\theta$  along the optimal policy strategy, is thus also Gaussian with mean 0; and each agent receives only a noisy version of the announcement, given by  $x_i = X + u_i$ , where  $u_i$  is an idiosyncratic Gaussian random variable with mean 0. Let  $x_i$  have a fixed signal-to-noise ratio with respect to  $X$ , regardless of the form of forward guidance. This signal-to-noise ratio indexes agents’ “attentiveness” to arbitrary announcements. All agents are rational and Bayesian. All these facts and the aforementioned signal-to-noise ratio are common knowledge.

In this model, which resembles Morris and Shin (2002), Woodford (2003), and the topical applications of Angeletos and Lian (2018) and Wiederholt (2016), the following properties hold. First, the average expectation of  $X$  is  $\bar{\mathbb{E}}[X] = qX$  for some  $q < 1$  (with  $q$  being simply a positive transformation of the signal-to-noise ratio). Second, the higher-order expectations of  $X$  satisfy  $\bar{\mathbb{E}}^h[X] = \lambda^{h-1}\bar{\mathbb{E}}[X]$  for all  $h \geq 2$ , with  $\lambda = q$ . And third, the implementable combinations of  $\tau$  and  $Y$  are *invariant* to the form of forward guidance.

The first two properties are commonplace in the literature. The third, which is central to our purposes and is proved in Appendix D, clarifies that noisy information *alone* does not upset the irrelevance property of the noiseless REE benchmark we studied in Section 3. A simple intuition is that, in the new context,

which has noisy but still rational expectations,  $\tau$  and  $Y$  are both functions of the same fundamental and these functions are themselves correctly understood by the agents. It follows that a signal of one is just as good as a signal of the other and, as a result, there is still no meaningful trade-off between anchoring the expectations of  $\tau$  and anchoring the expectations of  $Y$ .

By the same token, the crucial feature of Assumption 3 was not the rigidity in higher-order beliefs per se (which is present in the above model) but rather the systematic error in equilibrium reasoning (which is absent in the above model). This squares well with the basic premise that our paper is all about systematic errors in equilibrium reasoning, as opposed to mere inattention or rational confusion. But what if these conceptually distinct forces come together in practice?

In Appendix D, we augment the previous model with a behavioral twist—incorrect perception of the noise in private signals—to separately parametrize the relevant forces. The first- and higher-order beliefs continue to satisfy  $\bar{\mathbb{E}}[X] = qX$  and  $\bar{\mathbb{E}}^h[X] = \lambda^{h-1}\bar{\mathbb{E}}[X]$  for  $h \geq 2$ , but now  $\lambda \neq q$ . This immediately restores the difference between instrument and target implementation:  $\lambda \neq q$  in the present context has essentially the same effect on implementability as  $\lambda \neq 1$  in our baseline analysis. We then obtain two sets of results for optimal policy, based on different perspectives on the “right” specification of the policymaker’s objective.

Suppose, first, that the bias in equilibrium reasoning ( $\lambda \neq q$ ) remains the only source of inefficiency. By this we mean a situation in which inattention ( $q < 1$ ) is rational and internalized by the policymaker. In this case, we obtain a generalization of Theorem 1 and Corollary 6 regardless of the values of  $\lambda$  and  $q$ .

This case, which is our preferred one from a theoretical perspective, is consistent with recent work showing that rational inattention, *by itself*, does not upset market efficiency and does not call for policy intervention.<sup>26</sup> It also fully preserves the spirit of our baseline analysis: the policymaker’s optimum coincides with rational expectations equilibrium whether the latter is noiseless (baseline) or noisy (here).

If, in contrast, we treat both inattention ( $q < 1$ ) and incorrect equilibrium reasoning ( $\lambda \neq q$ ) as sources of inefficiency, the usual second-best argument applies: in principle, the policymaker might try to use one distortion to offset the other. We spell out the details in Appendix D. We proceed to argue that, even under this confounding scenario, our main policy prescription survives in the case that best fits the available survey evidence on macroeconomic expectations.

## 7 Application: Monetary Policy in a Liquidity Trap

In this section, we contextualize our model’s relevant insights for the main motivating example: communication about monetary policy when interest rates are at their zero lower bound (ZLB).

### 7.1 Justifying a switch from instrument to target communication

As noted earlier, our finding that bounded rationality, of the kind captured by Assumption 3, attenuates the power of instrument-based forward guidance is basically the same as a common finding of Angele-

---

<sup>26</sup>See [Angeletos and Sastry \(2019\)](#) for an extension of the Welfare Theorems for economies with a general form of rational inattention, and [Angeletos and La’O \(2018\)](#) for a related result in the context of optimal monetary policy.

tos and Lian (2018), Farhi and Werning (2019), García-Schmidt and Woodford (2019), Gabaix (2019) and Wiederholt (2016). Our analysis has added three new lessons, which are crucial for policy interpretation.

First, we have shown how a switch in the *form* of forward guidance may bypass or at least mitigate the problem emphasized in this earlier literature (Proposition 3).<sup>27</sup> Second, we have offered a simple guideline for when it is optimal to switch from instrument-based to target-based forward guidance (Theorem 1). And third, we have shown that this guideline is robust to alternative assumptions about the friction in GE reasoning that may flip the predictions of the aforementioned works (Proposition 10).

Let us focus on the offered policy guideline: that central banks should emphasize targets, or shine the spotlight on unemployment, in situations with large GE feedback. In the simple New Keynesian micro-foundation described in Appendix B.1, this condition corresponds to a larger marginal propensity to consume out of income in the liquidity trap, or a steeper Keynesian cross. More complicated models enrich the “deep” determinants of the relevant GE feedback, often increasing its quantitative magnitude.

For example, binding liquidity constraints and heterogeneity typically steepen the Keynesian cross (Kaplan, Moll and Violante, 2018; Farhi and Werning, 2019). Financial accelerator mechanisms introduce an additional channel for (expected future) demand to feed back into consumption via asset wealth (Kiyotaki and Moore, 1997; Bernanke, Gertler and Gilchrist, 1999; Caballero and Simsek, 2017). And in the fully dynamic version of the New Keynesian model, the interaction of consumers’ demand with firms’ price-setting decisions, and then back into consumption via inflation expectations, generates a chain of GE feedback effects whose length and ferocity increases with the horizon of the liquidity trap (Angeletos and Lian, 2018), a property that may be proxied here by mapping a longer trap to a larger  $\gamma$ .

In the light of these observations, our main result (Theorem 1) suggests that the Great Recession was the *most opportune* time to engage in target communication. This complements the policy lesson of the aforementioned literature, which argued that the ineffectiveness of instrument-based guidance is severest in precisely the scenarios that, as shown here, most favor target-based guidance. And it offers a new normative justification for the actual shift in the Fed’s language in late 2012.

At this point, we must return to the role played in our analysis by the assumption that the first best is attained under rational expectations. As noted in Section 2.2, this assumption was ideal for purely theoretical purposes, because it let isolate bounded rationality as the only distortion. But it is almost certainly not applicable in the Great Recession. In this context, it is instead more appropriate to assume that the level of output during the trap falls short of the first best even under rational expectations. In Appendix B.1.3, we show how to accommodate this scenario in our framework and explain why this *strengthens* the case for target communication under our baseline bounded-rationality specification. Achieving more response of  $Y$  to  $\tau$  usefully “keeps powder dry” and avoids an excessive distortion after the liquidity trap.

---

<sup>27</sup>This is subject to the qualifications made in footnote 22 and Section 6.4. In particular, because the approach taken in Angeletos and Lian (2018) and Wiederholt (2016) features both a friction in higher-order beliefs and a friction in first-order beliefs (or both  $\lambda < 1$  and  $q < 1$  in the notation of Section 6.4), it has the capacity to attenuate *both* forms of forward guidance. We suspect that the same is true for the version of Gabaix (2019) that combines cognitive discounting with sparsity, and for an extension of Farhi and Werning (2019) and García-Schmidt and Woodford (2019) that adds inattention. But as long as this class of models is parameterized to match the relevant survey and laboratory evidence, the analysis in Appendix D suggests that target-based guidance remains more powerful than instrument-based guidance in *relative* terms.

## 7.2 Shining the spotlight to restore confidence

San Francisco Fed president [John C. Williams](#), when discussing the merits of changing the language of forward guidance during the Fed's December 2012 meeting, said that the "spotlight" would shift from instruments to targets "for better or worse" (Section 2.4). Although we are not sure what [John C. Williams](#) had in mind or how he would himself formalize it, our model provides one complete story for why this metaphor is apt and how it can translate into a policy prescription.

As a preliminary matter, should shining the spotlight on a particular variable matter at all? With rational expectations equilibrium, given full commitment and a lack of confounding shocks, agents are equally informed about and/or confident in the future values of instruments or targets regardless of which is announced. The spotlight might shine directly on only one object, but it is reflected just as clearly onto the second in the public consciousness.<sup>28</sup>

By contrast, in the model with bounded rationality, the spotlight is necessarily brighter on the announced variable than on the variable about which the public imperfectly reasons in general equilibrium. More practically, the policymaker can never talk their way around belief disagreement to achieve "perfect clarity," defined here as full alignment of beliefs about both instruments and targets.<sup>29</sup> What optimal policy does instead is mitigate the *impact* of inevitable disagreement on equilibrium outcomes.

While "achieving perfect clarity" is not the right intuition for our main result, "maximizing confidence" is. An reasonable definition of "confidence" in our setting, taking very seriously the story of Assumption 3, might be "the importance of forecasting others' actions to take one's own." The negative magnitude of strategic interaction,  $-|\delta|$  as defined in best-response (9), is a natural unit that captures the sensitivity of one's actions to forecasts of others. The policymaker's preference for a given communication method monotonically increases in thusly defined confidence, because distortions kick in exactly when confidence is low and equilibrium reasoning is important.

## 7.3 Explaining dis-anchored beliefs in the crisis

Through the lens of our theory, the Fed's foray into target communication in December 2012 may have optimally sacrificed confidence about interest rates (and tranquility in the bond market) in order to anchor the "more useful" expectations about real outcomes.

While the data from this one episode are insufficient to test our model precisely, they are consistent with there being a trade-off in anchoring instrument and target expectations. Consider first the FOMC's explicit date-based guidance in late 2011 and early 2012. On the one hand, the Fed was able to move expectations about interest rates to an extent that it could not during the previous, weaker regime of communication (see, e.g., [Swanson and Williams, 2014](#)). San Francisco Fed president [John C. Williams \(2016\)](#) writes succinctly:

---

<sup>28</sup>In fact, even the simplest departures from full information that maintained full rationality (like the Gaussian signal-extraction example in Section 6.4) allow the spotlight to be dimmer but maintain this property of full reflection onto other objects. This leads again to irrelevance of communicating via instruments or targets.

<sup>29</sup>To see this, note that beliefs about  $K$  are misaligned essentially by assumption (Lemma 3). This necessarily translates into mis-alignment for one of the main outcome expectations.

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

On the other hand, [Andrade et al. \(2019\)](#) show in professional forecast data that the same focusing of interest rate expectations coincided with an *increase* in the dispersion of forecasts of real variables including consumption. This is consistent with our theory's proposition that instrument communication is successful in anchoring expectations of interest rates only at the expense of increasing the uncertainty about how the economy would respond to them.<sup>30</sup>

And although the December 2012 shift may not resemble the “purest” form of target communication we have in mind, it did coincide with a stabilization in GDP forecast dispersion (see again [Andrade et al., 2019](#)). This is again consistent with our theory's proposition that target communication helps anchor expectations of outcomes, but leaves the economy vulnerable to disagreement about the supporting policy.

Of course, this is not the only possible explanation of these facts. But to the extent that the apparent trade-off between anchoring expectations of interest rates and anchoring expectations of unemployment and income reflects, at least in part, imperfect reasoning about the workings of the economy, our theory provides a context for how this trade-off may be optimally resolved.

#### 7.4 Interpreting the value of an “unemployment target”

Speaking in support of target-based guidance in December 2012, then Chicago Fed president Charles Evans mentions specifically that the optimal point to depart from the ZLB might “change in response to the inevitable arrival of exogenous shocks.” Consequently, he argues, target-based guidance would provide “much more clarity” about this moving goalpost.

But what are the shocks about which the Fed was, or perhaps should have been, concerned with? Did they only reflect uncertainty about the economy's “hard” fundamentals, or also uncertainty about when and how “public confidence” would be restored?

The first kind of uncertainty is often captured in the literature by an uncertain date at which the natural rate of interest will exogenously revert from a negative value to a positive value ([Eggertsson and Woodford, 2003](#)). As noted in Section 6.1, an increase in this kind of uncertainty can alone justify a switch from instrument to target communication, via the [Poole](#)-type logic.

But it is anyone's guess if this was the *only* relevant kind of uncertainty. To the extent that some of the uncertainty regarded how the public would “interpret” forward guidance, how it would “coordinate” its response to it, and when “public confidence” would be restored, our results shed new light on why the policy shift may have been very timely.

<sup>30</sup>To be precise, this is true as soon as we introduce heterogeneity in  $\lambda$ . With such heterogeneity, our results go through with  $\lambda$  replaced by its average value in the population. At the same time, the heterogeneity in  $\lambda$  manifests as heterogeneity in the forecasts of  $K$  and thereby in the forecasts of  $\tau$  and/or  $Y$ . Instrument communication then minimizes the heterogeneity in the forecasts of  $\tau$  at the expense of maximizing the heterogeneity in the forecasts of  $Y$ ; and the converse is true with target communication. Of course, differences in beliefs about fundamentals could also be the source of heterogeneity in the aforementioned forecasts, but whether and how this interacts with the form of forward guidance is outside the scope of our analysis.

First of all, our approach makes the following prediction not shared by the [Poole](#)-like alternative: target communication becomes more desirable as GE feedback loops intensify ( $\gamma$  increases). The idea that such feedback loops were intensified during the Great Recession was front-and-center in policy discussions throughout the crisis (think of the discussions related to the risk of a “deflationary spiral”), but does not appear to have been linked to the 2012 change in Fed’s communication strategy. From this perspective, our key contribution is to make this link and to provide an additional, and previously unknown, rationale for why this shift may have been optimal.

Second, our approach suggests that this policy shift may have—unintentionally but favorably—reduced the period of time that the economy was stuck at the ZLB. This is the point we anticipated at the end of Section 6.1: the reduction in the expected value of  $\tau$  needed to induce the desired expected value of  $Y$  translates, in the present context, to less monetary loosening needed *after* the economy has exited the liquidity trap, or a faster “lift-off” date, while achieving the same stimulating effect during the trap.

Needless to say, this prediction is, highly speculative; the counterfactual is impossible to know. But under the lens of our analysis, this is a logical extension of the apparent effectiveness of Mario Draghi’s “whatever it takes” speech. In particular, the logic behind our result is closely connected with arguments that an unwavering policy commitment to be the lender of last resort, or to defend the economy against a self-fulfilling crisis, can “rule out the bad equilibrium” in a multiple-equilibrium environment. In both cases, the key is that a commitment to a value for an endogenous outcome instead of a policy instrument arrests the effects of people’s “unreliable” reasoning about the behavior of others.

## 8 Sophisticated Forward Guidance and Policy Rules

In reality, central bankers are constantly talking about instruments, targets, various intermediate quantities and their contingencies upon one another. And in theoretical models, these contingencies map to “Taylor rules” describing how monetary policy adjusts to evolving economic conditions.

We abstracted from these options in our main analysis because we believed they were less relevant in the midst of crisis (recall the discussion in Section 2.4). Here, we show that the policy trade-off featured in our baseline, instruments-versus-target setting and the main insights developed therein are robust to letting the policymaker commit to and communicate an arbitrarily flexible reaction function. We also clarify the essential role of commitment: to persuade a public that has the wrong model of how GE works to nevertheless do the right thing, the policymaker must commit to a reaction function that is ex post suboptimal. We finally use the present extension to suggest how our insights could pave the way to a new, bounded-rationality-focused perspective on the design of optimal policy rules.

### 8.1 Set-up

Assume that, after observing  $\theta$ , the policymaker can commit to and communicate a flexible *relation* between the instrument  $\tau$  and the outcome  $Y$ , given by

$$\tau = T(Y; \theta),$$

for some function  $T : \mathbb{R}^2 \rightarrow \mathbb{R}$ .<sup>31</sup> Without serious loss of generality, we restrict attention to linear reaction functions of the form

$$T(Y; \theta) = a + bY, \quad \text{with} \quad a = A(\theta) \quad \text{and} \quad b = B(\theta), \quad (14)$$

for arbitrary  $A(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$  and  $B(\cdot) : \mathbb{R} \rightarrow (\underline{b}, 1)$ , where  $\underline{b} \equiv \frac{1+\alpha\gamma}{1-2\alpha+\alpha\gamma} < -1$ . The bounds on  $b$  are necessary and sufficient for “reasoning to converge,” or for infinite-order beliefs not to have undue influence on behavior.<sup>32</sup> The simpler strategies considered in our baseline analysis are nested with  $b = 0$  and  $a = \hat{t}$  for instrument communication, and  $b \rightarrow -\infty$  and  $-a/b \rightarrow \hat{Y}$  for target communication. With the flexibility added here, forward guidance amounts to announcing, conditional on  $\theta$ , a pair of numbers  $(a, b) = (A(\theta), B(\theta))$ , or an intercept and a slope for the reaction function, instead of a single number  $\hat{t}$  or  $\hat{Y}$ .<sup>33</sup>

All assumptions about depth of knowledge and rationality now relate to agents’ understanding of the function  $T$ , or the pair  $(a, b)$ . In particular, Assumption 3 is adapted as follows: agents believe that only a fraction  $\lambda \in [0, 1]$  of the others are both rational and aware of the actual  $(a, b)$ , like themselves; the rest are expected to play the “default” action  $k = 0$ , either because of inattention or because of irrationality.<sup>34</sup>

## 8.2 Optimal policy

In our main analysis, we contrasted how the choice between instrument and target communication was irrelevant in the rational-expectations benchmark ( $\lambda = 1$ ) to how it became crucial in managing expectations once we accommodated bounded rationality ( $\lambda < 1$ ). The next result generalizes this insight to the richer policy strategy space allowed here.

**Proposition 11.** *Consider a reaction function  $T$  and let  $Y(\theta)$  and  $\tau(\theta)$  be, respectively, the induced equilibrium values of the outcomes and the supporting policies (together, “allocations”). Next, consider any other reaction function  $T'$  such that  $T'(Y(\theta), \theta) = T(Y(\theta), \theta)$  for all  $\theta$ .*

(i) *When  $\lambda = 1$ ,  $T'$  induces the same equilibrium outcomes and policies as  $T$ .*

(ii) *When instead  $\lambda < 1$ ,  $T'$  induces different equilibrium outcomes and policies than  $T$ .*

---

<sup>31</sup>Clearly, the outcomes implemented with such a policy rule coincide with those implemented with a rule of the form  $\tau = T(K; \theta)$ , since  $Y$  is a (fixed) function of  $\tau$  and  $K$ .

<sup>32</sup>See the proof of Proposition 12 for the details. In the New Keynesian framework, the analogue of  $b < 1$  is the Taylor principle, and the analogue of  $b > \underline{b}$  is the additional bound on the slope of the Taylor rule identified by [Guesnerie \(2008\)](#) as necessary and sufficient for the unique linear REE of that model to be also the unique rationalizable outcome.

<sup>33</sup>This interpretation is under the maintained timing, which has the policymaker choose and communicate the scalars  $(a, b)$  *after* observing  $\theta$ . But the same outcomes obtain also with an alternative timing that has the policymaker choose and communicate the entire mappings  $(A(\cdot), B(\cdot))$  *prior* to observing  $\theta$ . The first perspective seems more natural in the context of forward guidance and under the interpretation of  $\theta$  as the policymaker’s current assessment of the best thing to do. The second perspective is more appropriate for connecting to the macroeconomic literature on policy rules and for re-interpreting  $\theta$  as a *future* shock. Finally, note that for now we are allowing *both* the intercept and the slope of the policy rule to vary with  $\theta$ , but below we will show that optimality requires that *only* the intercept varies with  $\theta$ .

<sup>34</sup>This specification imposes  $\lambda \leq 1$ . But the results stated below readily extend to  $\lambda > 1$ , or a situation where agents over-estimate the responses of others and the GE effects of policy, along the lines of Section 6.2.

Part (i) is familiar from the existing literature on Ramsey problems, in which there is often a large family of policy rules that implement the same equilibrium allocations and policies. The analogue of this property in the 3-equation New Keynesian model is also well known: there are multiple combinations of a state-contingent intercept and a slope for the Taylor rule that implement the same equilibrium paths for output, inflation, and interest rates.<sup>35</sup>

Part (ii) shows that this kind of irrelevance breaks once we bound agents' depth of knowledge and rationality. Fix  $T$  and let  $\theta \mapsto (\tau^*(\theta), Y^*(\theta))$  be the equilibrium mapping from states to allocations implemented by  $T$ . Next take any other  $T'$  that satisfies  $T'(Y^*(\theta), \theta) = \tau^*(\theta)$ . This property guarantees that agents find it optimal to play the same action under  $T'$  as under  $T$  insofar as long as they conjecture that  $T'$  continues to induce the same allocations as  $T$ . When  $\lambda = 1$ , one can close the loop to prove this conjecture is self-fulfilling and hence that  $T'$  induces the same behavior as  $T$ . But once  $\lambda < 1$ , agents doubt that *others* make the same conjecture. This causes them to form different expectations about  $K$  under  $T'$  than under  $T$ , which in turn leads them to follow different behavior under  $T'$  than under  $T$ .

In short, the above result generalizes our earlier insights about the role of policy in regulating the error in the public's reasoning and its footprint on actual behavior. The upshot for optimality is given below:

**Proposition 12.** (i) *When  $\lambda = 1$ , the optimal rule is indeterminate and its slope can be anything: the first best is implemented if and only if the intercept satisfies  $a = (1 - b)\theta$ , for an arbitrary (possibly  $\theta$ -contingent) slope  $b$ .*

(ii) *When instead  $\lambda \neq 1$ , the optimal rule is unique and its slope is inversely related to the GE feedback: the first best is implemented if and only if*

$$b = -\frac{\gamma}{1 - \gamma} \quad \text{and} \quad a = \frac{1}{1 - \gamma}\theta \quad \forall \theta. \quad (15)$$

With rational expectations, optimality requires that  $\tau = Y = \theta$ , but there is a continuum of policy rules that induce this as an equilibrium. The analogue in the New Keynesian model (without a binding ZLB and markup shocks) is that the first best can be implemented with a continuum of Taylor rules, whose state-contingent intercept tracks the natural rate of interest and whose non-contingent slope with respect to inflation or the output gap is indeterminate.

With bounded rationality, this indeterminacy disappears. The slope of the optimal rule is now inversely tied to the strength of the GE feedback, in a way that smooths out our baseline main result (Theorem 1): as  $\gamma$  increases, the policymaker gives more emphasis on anchoring the public's expectations of  $Y$  rather than their expectations of  $\tau$ .

To see this more clearly, let us first re-express the optimal rule as follows:

$$\tau - \theta = -\frac{\gamma}{1 - \gamma}(Y - \theta).$$

From this perspective, the optimal forward guidance consists of two components: the policymaker's assessment of the "fundamentals" and of the corresponding "rational" outcome (e.g., the central bank's forecast about the natural rate of output) in the form of  $\theta$ ; and a commitment about how much she will tolerate

---

<sup>35</sup>The most applied segment of the New Keynesian literature (e.g., that on estimated DSGE models) often removes the state-contingency of the intercept of the Taylor rule. We return to this issue at the end of this section.

a gap in terms of  $\tau$  versus a gap in terms of  $Y$ . Building on our discussion of categorizing policy communication in Sections 2.4 and 2.5, the former piece might be reflected in the policymaker's overall outlook (e.g., as reflected in the Summary of Economic Projections and dot plots), whereas the latter requires an explicit discussion of policy's contingency on different outcomes. The latter piece, our analysis shows, is *required* to achieve the first best—forecasts, by themselves, cannot do the job.

The next result expands on how policy optimally manages the expectations of interest rates and aggregate employment when both of them are distorted due to bounded rationality:

**Proposition 13.** *Let  $f_\tau(\gamma) \equiv |\tau - \bar{\mathbb{E}}[\tau]|$  and  $f_Y(\gamma) \equiv |Y - \bar{\mathbb{E}}[Y]|$  denote the aggregate errors in the expectations of, respectively, the instrument and the outcome, evaluated at the optimal policy, as functions of  $\gamma$ . Then, for all  $\gamma \in (0, 1)$ : (i)  $f_\tau(\gamma) > 0$  and  $f_Y(\gamma) > 0$ ; (ii)  $f'_\tau(\cdot) > 0$ , and  $f'_Y(\cdot) < 0$ .*

The first property shows that, away from the extreme values of  $\gamma$ , the optimal policy is never completely clear: it does not eliminate the mistakes in either kind of expectations. This might be surprising given the previous discussion of “forecasts plus commitments”—in this interpretation, if the Fed had provided forecasts and a dot plot that described  $\tau = Y = \theta$ , its optimal forward guidance would induce the public to think something else.

The second property shows that policy shifts clarity from  $\tau$  to  $Y$  as the GE feedback increases. This makes even clearer how the optimal policy rule “smooths out” the main insight of Theorem 1 about switching the spotlight from instruments to targets, or from interest rates to unemployment.

Although the optimal rule does not eliminate the mistakes in people's reasoning about equilibrium, under the assumptions made thus far it insulates their actual behavior from such reasoning and recovers the policymaker's first best. The intuition is similar to the one developed for the extremes  $\gamma = 0$  and  $\gamma = 1$  in our baseline analysis, except that it now extends to interior  $\gamma$ : the optimal rule zeros out equilibrium reasoning about others' reactions.<sup>36</sup>

One should not take the present result too literally. First, there may be costs (left outside our analysis) for communicating sophisticated strategies. Second, if the policymaker is uncertain about the precise value of  $\gamma$ , the policymaker implements a “second-best approximation” of the policy described in Proposition 12: the first best is not attainable any more, but the optimal  $b$  increases (in the sense of first-order stochastic dominance) in the policymaker's beliefs about  $\gamma$ . We expect a related result to hold in a multi-decision extension with limited policy instruments: the policymaker would try to eliminate the distortion in all decisions, but might succeed in doing so for only some.

But the basic logic is always the same. The optimal policy aims at minimizing the public's need to reason about the economy. And this is achieved by shifting emphasis from anchoring the public's expectations of  $\tau$  (“interest rates”) to anchoring the public's expectations of  $Y$  (“unemployment”) as the GE feedback increases.

---

<sup>36</sup>In the language of best-response condition (9), the optimal rule ensures a zero slope on  $\mathbb{E}_i[K]$ . This implements the first best without correctly anchoring beliefs about *either* the instrument or target, but instead by making sure that the distortions in those beliefs are exactly irrelevant for choices.

### 8.3 Optimal policy without commitment

We now expand on the role played by commitment. In the absence of commitment, the policymaker chooses  $\tau$  in stage 2 so as to minimize  $L$  subject to condition (1), taking  $K$  as given. This gives the following optimality condition, which trades off the marginal effect of the policy on the two “gaps:”

$$(1 - \chi)(\tau - \theta) + \chi(Y - \theta)(1 - \alpha) = 0;$$

Rearranging gives the follow ex post optimal reaction function:

$$\tau = \frac{1-\alpha\chi}{1-\chi}\theta - \frac{\chi(1-\alpha)}{1-\chi}Y. \quad (16)$$

The coefficients of this reaction function do not depend on the parameter  $\gamma$ , which determined how expectations of  $\tau$  and  $Y$  mapped to  $K$ , because  $K$  itself is already determined. By contrast, the optimal policy rule with commitment, given in (15), depends on  $\gamma$  precisely because it internalizes the effect it has on public reasoning and thereby on  $K$ .

Except for the knife-edge case in which  $\frac{\chi(1-\alpha)}{1-\chi} = \frac{\gamma}{1-\gamma}$ , the two rules are different. Commitment is *necessary* for implementing the first best—but only insofar as there is a distortion in equilibrium reasoning ( $\lambda \neq 1$ ). When instead  $\lambda = 1$ , the following class of rules implements the first best under full commitment and rational expectations:

$$\{\tau = (1 + b) - bY, \text{for any } b\}$$

But this class now includes the policy rule in (16), which means that commitment is *not* needed under rational expectations.

When the policymaker follows (16) and agents have rational expectations, agents *correctly* expect that all others will play  $K = \theta$  and that this together with (16) will induce  $\tau = Y = \theta$ . But once  $\lambda \neq 1$ , this rule causes agents to believe that  $K \neq \theta$ , which in turn distorts their behavior away from the first best. The only way to fix this distortion is to make sure that agents find it optimal to play the first best action *regardless* of their beliefs of  $K$  (or regardless of their higher-order beliefs), which in turn is possible if and only if the policymaker commits to the rule described in (16).

We summarize these lessons below.

**Proposition 14.** *In the absence of commitment, the unique optimal policy rule is given by*

$$\tau = \frac{1-\alpha\chi}{1-\chi}\theta - \frac{\chi(1-\alpha)}{1-\chi}Y. \quad (17)$$

*This implements the first best when  $\lambda = 1$  but not when  $\lambda \neq 1$  (except for the knife-edge case in which  $\frac{\chi(1-\alpha)}{1-\chi} = \frac{\gamma}{1-\gamma}$  or  $\theta = 0$ ).*

**Corollary 7.** *Commitment is valuable only when  $\lambda \neq 1$  (away from rational expectations). In these circumstances, the rule described in (16) is ex ante optimal, even though ex post suboptimal, because and only because the commitment embedded in it helps regulate the distortion in equilibrium reasoning.*

Of course, the property that commitment is useless under rational expectations is special to our model. There is a large literature in macroeconomics studying time-inconsistency issues under rational expectations in the context of both flexible policy rules (Kydland and Prescott, 1977; Barro and Gordon, 1983) and simpler, instruments-versus-targets implementations (Atkeson, Chari and Kehoe, 2007; Halac and Yared, 2018). But by assuming away these familiar considerations, we have illustrated a *new* function that commitment can play away from rational expectations.

Circling back to our baseline analysis, this also makes clear the following point: what was crucial about the two kinds of forward guidance studied there is that they communicated a policy plan (i.e., a commitment to get something done) as opposed to information about fundamentals (e.g., the central bank's forecasts of future fundamentals).

## 8.4 A new approach to optimal Taylor rules

A strand of the New Keynesian literature often bypasses the indeterminacy of the optimal rule under rational expectations by removing its contingency on the underlying state of nature (i.e., by requiring the intercept of the Taylor rule to be non-contingent). To understand what this means in our context, recall that the optimal rational-expectations policy rule from Proposition 12 had the form  $\tau = (1 - b)\theta + bY$ , for any  $b \in \mathbb{R}$ . This nests a “direct” implementation of  $\tau = \theta$ . But when  $\tau$  cannot be directly contingent on  $\theta$ , the previous criterion selects  $b = 1$ , or  $\tau = Y$ , as the unique optimal rule.

Clearly, this is a unique optimal rule for a very different reason than our analysis suggests: its design uses the equilibrium value of  $Y$  as a “signal” of the underlying shocks to fundamentals, or as a method for replicating the optimal contingency of  $\tau$  on  $\theta$ . The same logic applies in richer applications that allow for multiple shocks (e.g., demand and supply shocks) and contingencies on multiple outcomes (e.g., a Taylor rule that ties interest rates to both inflation and output, but not the underlying shocks).<sup>37</sup>

Our own result, instead, indicates how such feedback rules can serve an entirely new function: regulating the impact of bounded rationality, or distorted equilibrium reasoning, on the transmission of shocks. This is evident in the different determinants of the *slope* of the optimal policy rule. Under the aforementioned replication approach, the slope is pinned down by the relative sensitivity of the optimal policy and the optimal outcome to the fundamental:  $b = 1$ . Under our approach, instead, it is pinned down by strategic considerations, or the relative importance of GE effects:  $b = -\frac{\gamma}{1-\gamma}$ .<sup>38</sup>

How do these insights translate in richer, dynamic, macroeconomic models used in quantitative policy

---

<sup>37</sup>Sometimes these contingencies are rich enough to permit the inversion of the equilibrium outcomes for the underlying shocks and the replication of the first best. Other times they are rough, precluding such a perfect replication, but preserving the logic: the outcomes are used as signals of the underlying fundamentals.

<sup>38</sup>Needless to say, if both considerations are present, the optimal rule will have to strike a balance between two goals, similarly to the Poole-like extension discussed in Section 6.1. Indeed, both our baseline setting and Poole (1970) correspond to policy rules that are either flat or vertical ( $b = 0$  or  $b = -\infty$ , respectively, in the language of the present extension). But whereas our baseline analysis relates the optimal choice to the goal of minimizing the distortion of bounded rationality, Poole's analysis relates it to the goal of approximating as well as possible the optimal contingency on fundamentals. In short, the same core idea that distinguishes our baseline analysis from Poole (1970) also distinguishes the present extension from the existing literature on optimal Taylor rules—which is not surprising since this literature is itself a logical extension of Poole (1970).

evaluation? In such models, a Taylor rule is a mapping from inflation and output (or the corresponding “gaps”) to the interest rate set by the monetary authority; and the optimal such rule is determined by the relative volatility of different kinds of fundamental (demand and supply shocks) so as to replicate the right contingency of the policy instrument on them. Our analysis instead suggests that the following parameters that regulate GE multipliers and that are often “buried” inside the representative-agent simplification should also enter the design of the optimal rule: the marginal propensity to consume, the degree of price rigidity, the strategic complementarity in price setting, and the extent of input-output linkages.<sup>39</sup> We leave the exploration of this insight for future research.

## 9 Conclusion

Should a policymaker offer clarity about policy instruments, for instance by outlining a time-path for interest rates or a dollar amount for fiscal stimulus, or should she instead shine the spotlight on the relevant outcome, for instance by promising to do whatever it takes to hit a target for unemployment?

In this paper, we first explained why this choice is irrelevant in a benchmark that stylizes the textbook policy paradigm. By assuming *infinite* depth of knowledge and rationality, this paradigm imposes that people have a flawless understanding of the correct mapping between policy and outcomes, or between interest rates and unemployment. This in turn guarantees that it makes no difference whether the policymaker communicates her intentions in terms of the former or the later.

We then explored the implications of accommodating bounded rationality, or incorrect equilibrium reasoning. We found that this not only breaks the aforementioned irrelevance but also yields a sharp dependence of the optimal communication strategy on the GE feedback between aggregate outcomes and individual actions:

**Main policy lesson.** *The optimal communication strategy shifts from instrument-based to target-based forward guidance, or from anchoring expectations of interest rates to anchoring expectations of unemployment and income, as Keynesian multipliers, financial accelerators, and other GE feedbacks get larger.*

Why? Instrument-focused communication leaves the public to reason about the effect of interest rates on aggregate employment and income. Target-focused communication does the opposite, sacrificing clarity about the policy for more anchoring of the expectations of targeted outcome. A larger GE multiplier makes expectations of aggregate outcomes such as employment and income more essential for private decisions, and any mistakes thereof more costly to the policymaker, which tilts the balance toward target-focused communication. More succinctly, the optimal communication strategy minimizes the public’s need to reason about equilibrium feedbacks, thus also minimizing the distortions caused by any mistakes in such reasoning.

---

<sup>39</sup>Here, we draw from [Angeletos and Lian \(2018\)](#), [Angeletos and Huo \(2018\)](#) and [Farhi and Werning \(2019\)](#), which illustrate how such parameters that are immaterial in the representative-agent, rational-expectation version of the fully-fledged New Keynesian model start mattering for its positive properties once one relaxes that model’s strong assumptions about knowledge and rationality. But these earlier works do not contain the insight offered here for the design of optimal policy rules.

There is a useful parallel, which we have emphasized throughout the paper, with the story behind Mario Draghi's famous proclamation to do "whatever it takes" to save the Euro. Evidently, according to a retrospective by [Jana Rardown and Alessandro Speciale \(2018\)](#) published in Bloomberg Markets magazine, Draghi did not yet have a precise policy in mind when he delivered his famous remarks in London. But he did have a plan:

After his pledge at Lancaster House to do whatever it takes, Draghi returns to Frankfurt and puts his staff to work turning half-formed plans into a viable program. Some heads of government and central bankers might take Draghi to task for not having a more fully formed strategy in the first place, but not Christian Noyer, the former governor of the Bank of France who was part of Draghi's inner circle. Draghi knew what he was doing, Noyer says: "He was relying on the capacity of the system to invent it. That's what I call genius intuition."

Our paper, among other things, formalizes one aspect of the "genius intuition:" that the right communication tactic could change the set of implementable outcomes due to its interaction with how the public thinks. This gets to the heart of our analysis and its lessons for forward guidance.

Apart from these context-specific lessons, our analysis makes a broader methodological contribution: to illustrate both how certain irrelevance results from the prevailing policy paradigm hinge on infinite depth of rationality, and how appropriate policies could regulate the bite of bounded rationality. As discussed in the previous section, the application of our approach to the design of optimal Taylor rules for monetary policy seems a particularly interesting direction for future work. Another such direction is the exploration of our insights in the context of Ricardian equivalence and quantitative easing.

Finally, our paper has provided a new theoretical context for a growing empirical literature on the effectiveness of central bank communication in moving the beliefs of both experts ([Campbell et al., 2012](#); [Ehrmann et al., 2019](#)) and the general public ([Coibion, Gorodnichenko and Weber, 2019](#); [Coibion et al., 2019](#)). Most of this literature focuses on information revelation: think, in particular, of the communication of the central bank's forecasts about future economic conditions. We instead have focused on the commitments embedded in central bank communications and have argued that such commitments may manage private sector expectations when the latter are not fully rational. We hope this perspective can inform more empirical work.

## References

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

**Alaoui, Larbi, and Antonio Penta.** 2016. “Endogenous Depth of Reasoning.” *The Review of Economic Studies*, 83: 1297–1333.

**Alicke, Mark D, and Olesya Govorun.** 2005. “The better-than-average effect.” *The self in social judgment*, 1: 85–106.

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

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

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

**Angelatos, George-Marios, and Alessandro Pavan.** 2009. “Policy with Dispersed Information.” *Journal of the European Economic Association*, 7(1): 11–60.

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

**Angelatos, George-Marios, and Jennifer La’O.** 2013. “Sentiments.” *Econometrica*, 81(2): 739–779.

**Angelatos, George-Marios, and Jennifer La’O.** 2018. “Optimal Monetary Policy with Informational Frictions.” *Journal of Political Economy*, forthcoming.

**Angelatos, George-Marios, and Karthik Sastry.** 2019. “Inattentive Economies.” National Bureau of Economic Research Working Paper 26413.

**Angelatos, George-Marios, and Zhen Huo.** 2018. “Myopia and Anchoring.” *MIT mimeo*.

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

**Atkeson, Andrew, Varadarajan V. Chari, and Patrick J. Kehoe.** 2010. “Sophisticated Monetary Policies.” *The Quarterly Journal of Economics*, 125(1): 47–89.

**Barro, Robert J.** 1974. “Are Government Bonds Net Wealth?” *Journal of Political Economy*, 82(6): 1095–1117.

**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.

**Bassetto, Marco.** 2019. “Forward guidance: Communication, commitment, or both?” *Journal of Monetary Economics*, 108: 69 – 86.

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

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

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

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

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

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

**Caballero, Ricardo J., and Alp Simsek.** 2019. "Monetary Policy with Opinionated Markets." MIT miméo.

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

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

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

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

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

**Coibion, Olivier, Dimitris Georgarakos, Yuriy Gorodnichenko, and Michael Weber.** 2019. "Forward Guidance and Household Expectations." *Mimeo*.

**Coibion, Olivier, Yuriy Gorodnichenko, and Michael Weber.** 2019. "Monetary Policy Communications and Their Effects on Household Inflation Expectations." *Mimeo*.

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

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

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

**Crawford, Vincent P, Miguel A Costa-Gomes, and Nagore Iribarri.** 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*.

**Eggertsson, Gauti, and Michael Woodford.** 2003. "The Zero Bound on Interest Rates and Optimal Monetary Policy." *Brookings Papers on Economic Activity*, 34(1): 139–235.

**Ehrmann, Michael, Gaetano Gaballo, Peter Hoffmann, and Georg Strasser.** 2019. "Can more public information raise uncertainty? The international evidence on forward guidance." *Journal of Monetary Economics*, 108: 93 – 112.

**Farhi, Emmanuel, and Iván Werning.** 2019. "Monetary Policy, Bounded Rationality, and Incomplete Mar-

kets." *American Economic Review*, 109(11): 3887–3928.

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

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

**Gabaix, Xavier.** 2019. "A Behavioral New Keynesian Model." *Harvard mimeo*.

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

**Gill, David, and Victoria Prowse.** 2016. "Cognitive Ability, Character Skills, and Learning to Play Equilibrium: A Level-k Analysis." *Journal of Political Economy*, 124(6): 1619–1676.

**Guesnerie, Roger.** 2008. "Macroeconomic and monetary policies from the eductive viewpoint." *Central Banking, Analysis, and Economic Policies Book Series*, 13: 171–202.

**Halac, Marina, and Pierre Yared.** 2014. "Fiscal rules and discretion under persistent shocks." *Econometrica*, 82(5): 1557–1614.

**Halac, Marina, and Pierre Yared.** 2018. "Instrument-Based vs. Target-Based Rules." National Bureau of Economic Research Working Paper 24496.

**Hansen, Lars Peter, and Thomas J Sargent.** 2008. *Robustness*. Princeton university press.

**Iovino, Luigi, and Dmitriy Sergeyev.** 2019. "Central Bank Balance Sheet Policies Without Rational Expectations." *Bocconi University mimeo*.

**Jappelli, Tullio, and Luigi Pistaferri.** 2010. "The consumption response to income changes." *Annual Review of Economics*, 2(1): 479–506.

**Kaplan, Greg, and Giovanni L. Violante.** 2014. "A Model of the Consumption Response to Fiscal Stimulus Payments." *Econometrica*, 82(4): 1199–1239.

**Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante.** 2018. "Monetary Policy According to HANK." *American Economic Review*, 108(3): 697–743.

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

**Kocherlakota, Narayana.** 2012. "Comments in the December 11-12, 2012, FOMC Meeting (Transcript)." Accessible at <https://www.federalreserve.gov/monetarypolicy/files/FOMC20121212meeting.pdf>.

**Kohlhas, Alexandre, and Ansgar Walther.** 2018. "Asymmetric Attention." *IIES miméo*.

**Kohlhas, Alexandre, and Tobias Broer.** 2018. "Forecaster (Mis-)Behavior." *IIES miméo*.

**Kydland, Finn E, and Edward C Prescott.** 1977. "Rules Rather Than Discretion: The Inconsistency of Optimal Plans." *Journal of Political Economy*, 85(3): 473–491.

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

**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.** 1998. "Unique Equilibrium in a Model of Self-fulfilling Currency Attacks." *American Economic Review*, 88(3): 587–597.

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

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

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

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

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

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

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

**Randow, Jana, and Alessandro Speciale.** 2018. "3 Words and \$3 Trillion: The Inside Story of How Mario Draghi Saved the Euro." *Bloomberg Markets*.

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

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

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

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

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

**Williams, John C.** 2012. "Comments in the December 11-12, 2012, FOMC Meeting (Transcript)." Accessible at <https://www.federalreserve.gov/monetarypolicy/files/FOMC20121212meeting.pdf>.

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

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

**Woodford, Michael.** 2010. "Robustly Optimal Monetary Policy with Near-Rational Expectations." *American Economic Review*, 100(1): 274–303.

# Appendices

## A Proofs

The proofs for the following results are contained in the main text: Lemmas 1, 2, and 3; Corollaries 1, 2, 3, 4, and 7 and Propositions 1, 2, 3, and 14. Propositions 15, 16, and 17, which are introduced in Appendix D, are also proved in that section. This appendix provides the proofs for all remaining results.

### Proof of Lemma 4

Recall that the frictionless benchmark, which corresponds to  $\lambda = 1$ , had  $K = \hat{X}$ , or  $\kappa_X = 1$ . When  $\delta_X > 0$ , the ratio  $\frac{1-\delta_X}{1-\lambda\delta_X}$  is strictly lower than 1 for every  $\lambda < 1$  and is increasing in  $\lambda$ . When instead  $\delta_X < 0$ , this ratio is strictly higher than 1 for every  $\lambda < 1$  and is decreasing in  $\lambda$ . Along with the fact that  $\delta_\tau > 0 > \delta_Y$ , this verifies the properties of  $\kappa_\tau$  and  $\kappa_Y$  stated in the Lemma.

### Proof of Proposition 4

**Instrument communication.** As shown in Proposition 3,

$$\mu_\tau(\lambda, \gamma) = \left( (1-\alpha) + \alpha \frac{1-\delta_\tau}{1-\lambda\delta_\tau} \right)^{-1} \quad (18)$$

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  $\partial\mu_\tau/\partial\lambda = -(\mu_\tau)^{-2}\partial\mu_\tau^{-1}/\partial\lambda < 0$ .

A necessary and sufficient condition for  $\mu_\tau < 1$  is  $\delta_\tau < 0$ . A sufficient condition for  $\delta_\tau < 0$ , and indeed the one that makes sense in our context, is  $\gamma < 0$ , or negative GE feedback.

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

$$\mu_Y(\lambda, \gamma) = \frac{1-\alpha\kappa_Y}{1-\alpha} = \frac{1-\lambda\delta_Y - \alpha(1-\delta_Y)}{(1-\alpha)(1-\lambda\delta_Y)} \quad (19)$$

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

To check the derivative with respect to  $\lambda$ , note that

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

and  $\partial\mu_Y/\partial\kappa_Y = -\alpha/(1-\alpha) < 0$ . Thus, by the chain rule,  $\partial\mu_Y/\partial\lambda = \partial\mu_Y/\partial\kappa_Y \cdot \partial\kappa_Y/\partial\lambda < 0$ .

### Further results

**Lemma 5.** *Let  $\gamma \in (-\infty, 1]$  and  $\alpha \in (0, 1)$ . The following statements are equivalent:  $|\delta_X| < 1$  for  $X \in \{\tau, Y\}$ ;  $\gamma > 2 - 1/\alpha$ ; and  $\alpha < 1/(2 - \gamma)$ .*

*Proof.* This is a very simple calculation. Note that  $\delta_Y < 0$  for any  $\gamma \leq 1$ , so  $\delta_Y < 1$  is guaranteed. The condition  $\delta_Y > -1$  re-arranges to

$$\delta_Y = -(1 - \gamma) \frac{\alpha}{1 - \alpha} > -1$$

This re-arranges to  $1 - \gamma < 1 - \frac{1}{\alpha}$  given  $\alpha \in (0, 1)$  and the previous re-arranges to  $\gamma > 2 - 1/\alpha$ . Finally, solving for  $\alpha$  gives  $\alpha < 1/(2 - \gamma)$ . Thus we have shown that  $|\delta_Y| < 1$ ,  $\gamma > 2 - 1/\alpha$ ; and  $\alpha < 1/(2 - \gamma)$  are interchangable statements.

Note next that  $\delta_\tau < 1$  is guaranteed by  $\alpha \in (0, 1)$ .  $\delta_\tau > -1$  requires  $\gamma > -1/\alpha$ . But this is implied by  $\gamma > 2 - 1/\alpha$  and hence by  $\delta_Y > -1$ .  $\square$

*Fix a value for  $\alpha$  and a domain  $\gamma \in G$ , with  $0 \in G$ , such that  $\delta_Y > -1$ . Then  $\mu_Y > 0$  on the same domain for all values of  $\lambda$ .*

*Proof.* Note that  $\mu_Y > 0$  when  $\kappa_Y < 1/\alpha$ . 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

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

A strictly tighter condition is the same evaluated at  $\lambda = 0$ , which re-arranges to

$$\gamma > \frac{1}{\alpha} \left( 2 - \frac{1}{\alpha} \right) \tag{20}$$

Note that the restriction that  $\delta_Y > -1$  encodes the following restriction for fixed  $\alpha$  and all  $\gamma \in G$

$$\gamma > 2 - \frac{1}{\alpha} \tag{21}$$

as calculated in Lemma 5.

Evaluating this at  $\gamma = 0$ , which is within the domain  $G$  over which the condition must apply, gives  $\alpha < 1/2$ . When  $\alpha < 1/2$ , the right-hand-side of conditions (20) and (21) are both negative, and (21) implies (20). Clearly, this must apply for all  $\gamma \in G$  and  $\alpha$  such that (21) holds. This completes the proof.  $\square$

**Lemma 6.** Assume that  $\mu_Y > 0$  and  $\lambda < 1$ . 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:

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

Note that  $\alpha < 1$  and  $\gamma \leq 1$  by assumption so  $\alpha\gamma < 1$ . Hence  $\lambda < 1$  is a sufficient condition for  $\mu_\tau > \mu_Y$ , and  $\lambda = 1$  is a sufficient condition for  $\mu_\tau = \mu_Y$ .  $\square$

## Proof of Proposition 5

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

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

Meanwhile, the slope with target communication is

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

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

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

This is less than one if and only if  $1 - \alpha < (1 - \lambda)/(1 - \alpha) < \alpha^{-1}$  or  $(1 - \alpha)^2 < 1 - \lambda < (1 - \alpha)\alpha$ . This is implied by the arguments of Proposition 4.

With instrument communication at  $\gamma = 0$ , the slope is

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

**Derivative of  $\mu_\tau$  with respect to  $\gamma$ .** For fixed  $\lambda$ , we can calculate first a derivative of the inverse slope with respect to the interaction parameter

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

which is unambiguously negative for  $\lambda < 1$ . Next,  $\partial \delta_\tau / \partial \gamma = \alpha > 0$ . Thus, by the chain rule,

$$\frac{\partial \mu_\tau}{\partial \gamma} = -(\mu_\tau)^{-2} \frac{\partial \mu_\tau^{-1}}{\partial \delta_\tau} \frac{\partial \delta_\tau}{\partial \gamma} > 0$$

**Derivative of  $\mu_Y$  with respect to  $\gamma$ .** For fixed  $\lambda$ , the partial derivative with respect to  $\delta_Y$  is

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

Next,  $\partial \delta_Y / \partial \gamma = \alpha / (1 - \alpha) > 0$ , or the best response slope increases (from negative values toward zero).

Hence

$$\frac{\partial \mu_Y}{\partial \gamma} = \frac{\partial \mu_Y}{\partial \delta_Y} \frac{\partial \delta_Y}{\partial \gamma} > 0$$

Note that this argument does not require that  $\mu_Y \geq 0$ .

## Proof of Theorem 1

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

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

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

$$\frac{\mu^2(1 - \chi) + \mu\chi}{\mu^2(1 - \chi) + \chi} \quad (22)$$

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  $\mu \in [0, 1]$ . Note that these are the only relevant domains to study according to the combination of Proposition 3 and Lemma 5.<sup>40</sup>

Finally, let  $\mathcal{L}_\tau$  and  $\mathcal{L}_Y$  denote the value of the loss function evaluated at  $r^*(\mu)$  and, respectively,  $\mu_\tau$  and  $\mu_Y$ . For fixed  $\lambda$  and  $\alpha$ , we let  $\mathcal{L}_\tau(\gamma)$  and  $\mathcal{L}_Y(\gamma)$  denote these losses as function of  $\gamma$ . We will argue that these functions cross exactly once at some  $\hat{\gamma}$ , the critical threshold of GE feedback.

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

This crossing point is unique if  $\partial \mathcal{L}_\tau / \partial \gamma > 0$  and  $\partial \mathcal{L}_Y / \partial \gamma < 0$  on the domain  $(0, 1)$ . By the chain rule,  $\partial \mathcal{L}_X / \partial \gamma = \partial \mathcal{L} / \partial \mu \cdot \partial \mu_X / \partial \gamma$  for  $X \in \{\tau, Y\}$ ; recall also  $\partial \mathcal{L} / \partial \mu > 0$  for  $\mu > 1$  and  $\partial \mathcal{L} / \partial \mu < 0$  for  $\mu \in [0, 1]$ . To show the desired monotonicities, then, it suffices to show that  $\partial \mu_\tau / \partial \gamma > 0$ ,  $\partial \mu_Y / \partial \gamma > 0$ , and  $\mu_\tau > 1 > \mu_Y > 0$ . All three are established in Proposition 4.

**Version with gap  $(K - \theta)^2$ .** As shown in Appendix B.1.3, our version of the New Keynesian model justifies the following variant specification for the policymaker's objective:

$$L(\tau, K, \theta) \equiv (1 - \chi)(\tau - \theta)^2 + \chi(K - \theta)^2, \quad (23)$$

---

<sup>40</sup>For a more general case that does not impose  $|\delta_X| < 1$  and/or  $\mu_X > 0$  for  $X \in \{\tau, Y\}$ , see Appendix C.3 for the amended statement and proof.

which is the same as (3) with  $K$  in place of  $Y$ . We now show that our main result extends to this case for essentially the same reasons as those explained above.

Let  $\beta_X$  denote the responsiveness of  $\tau$  to  $K$  under each form of communication, or

$$\beta_X \equiv \begin{cases} \frac{1}{\kappa_\tau} & \text{if } X = \tau \\ \frac{\mu_Y}{\kappa_Y} & \text{if } X = Y \end{cases}$$

where  $(\kappa_\tau, \kappa_Y)$  were defined in Lemma 4. As shown in the proof of Proposition 5, we have  $\partial\kappa_\tau/\partial\gamma < 0$  and  $\partial\kappa_Y/\partial\gamma > 0$  for  $\gamma \in (0, 1)$ . Moreover, from the statement of Lemma 4, we know that for any fixed  $\lambda \in (0, 1)$  and  $\gamma \in (0, 1)$ ,  $\kappa_\tau \in (0, 1)$  and  $\kappa_Y > 1$ . These properties combined suffice to show that  $\partial\beta_\tau/\partial\gamma > 0$ ,  $\partial\beta_Y/\partial\gamma > 0$ , and  $\beta_\tau > 1 > \beta_Y > 0$  for  $\gamma \in (0, 1)$ , given a fixed  $\lambda \in (0, 1)$ . Finally, for  $\gamma = 0$ , we have  $\beta_\tau = 1$ ; and for  $\gamma = 1$ , we have  $\beta_Y = 1$ .

Now consider the proof for Theorem 1, with  $\beta_X$  replacing  $\mu_X$ . The policy problem is

$$\min_{\beta \in \{\beta_\tau(\lambda), \beta_Y(\lambda)\}, r \in \mathbb{R}} (1 - \chi)(r - 1)^2 + \chi(r/\beta - 1)^2$$

and all the same arguments go through replacing  $\mu$  with  $\beta$ , until the very last step. At this point we use the properties  $\partial\beta_\tau/\partial\gamma < 0$ ,  $\partial\beta_Y/\partial\gamma > 0$ , and  $\beta_\tau > 1 > \beta_Y > 0$  for  $\gamma \in (0, 1)$ . This completes the proof.

## Proof of Proposition 6

This follows from (i) the results in Proposition 3 about the slope of the implementability sets; (ii) the result in Theorem 1 about optimal policy; and (iii) the formula given in the proof of the previous for the optimal choice of  $\tau = r^* \theta$ , which is re-printed here

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

## Proof of Corollary 5

For any  $\lambda < 1$ , the critical GE feedback threshold is the unique solution to  $\mathcal{L}_\tau(\hat{\gamma}) = \mathcal{L}_Y(\hat{\gamma})$ . Plugging directly into the loss function produces a quadratic equation for the threshold. Of the two roots, only one is in the correct domain,  $\gamma \in [0, 1]$ . This root is given by

$$\hat{\gamma} = \left( 1 - \alpha(1 - \chi\alpha)(1 - \lambda) + \left( \alpha(\alpha - 2\lambda - 2\alpha(1 - \lambda)\chi + (1 - \alpha(1 - \lambda)(1 - \alpha\chi))^2 \right)^{\frac{1}{2}} \right)^{-1} \quad (24)$$

The result is completed by taking the limit of this expression as  $\lambda \rightarrow 1$ .

## Proof of Propositions 7 and 8

See Appendix C.5.

## Proof of Proposition 9

This is a simple calculation involving the following two steps. First, for each of instrument and target communication, we solve the best-response fixed point

$$K = \kappa_X X + \kappa_X^\varepsilon \varepsilon = (1 - \delta_X)X + \delta_X(\lambda(\kappa_X X + \kappa_X^\varepsilon \varepsilon) + \sigma \varepsilon)$$

where  $\kappa_X$  coincides exactly with the values given in Lemma 4 and  $\kappa_X^\varepsilon$  is a new loading on the belief shock given by

$$\kappa_X^\varepsilon = \frac{\sigma \delta_X}{1 - \lambda \delta_X}$$

Then, to get the implementability constraints, we solve the fixed-point

$$\tau = \mu_X Y + \psi_X \varepsilon = \frac{1}{1 - \alpha}(Y - \alpha(\kappa_X X + \kappa_X^\varepsilon \varepsilon))$$

Which has solutions  $\mu_X$  given in Proposition 3 and  $\psi_X$  given by

$$\begin{aligned}\psi_\tau &= -\kappa_\tau^\varepsilon \frac{\alpha}{1 - \alpha + \alpha \kappa_\tau} \\ \psi_Y &= -\kappa_Y^\varepsilon \frac{\alpha}{1 - \alpha}\end{aligned}$$

The expressions given in the text follow from plugging the primitive parameters into the previous expression.

## Proof of Proposition 10

This is a simple extension of the proof of Theorem 1. Let us assume that  $\lambda \in [0, 1 + 1/\alpha]$ . This plays a similar role as the restriction  $\alpha < \frac{2}{1-\gamma}$ : it ensures convergence of iterated best-responses and thereby also that  $(\mu_\tau, \mu_Y) \in \mathbb{R}_+^2$ .

Note first, that using the same expressions from the proof of Proposition 5, that  $\mu_\tau$  and  $\mu_Y$  both *decrease* in  $\gamma$  for  $\lambda > 1$ . Importantly, the comparative statics in the distortion  $|\mu_X - 1|$  are the same.

Second, consider the comparative statics for the variance contributions  $(\psi_\tau, \psi_Y)$ . Note that

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

and  $\psi_\tau \leq 0$  as long as  $\lambda < 1/\alpha + 1$ . Finally,  $\psi_\tau = 0$  if  $\gamma = 0$ . Next,

$$\frac{\partial \psi_Y}{\partial \gamma} = \frac{\sigma \alpha^2}{(\alpha(1 - \gamma)\lambda + 1 - \alpha)^2} > 0$$

and  $\psi_Y \leq 0$  for any value of  $\lambda$ . Finally,  $\psi_Y = 0$  if  $\gamma = 1$ .

Consider now the policy loss function. After substituting in the implementability constraint, the loss function under target communication for a given value of  $\theta$  is

$$\mathcal{L}_\tau \equiv \min_{r_\tau \in \mathbb{R}} \left[ \theta^2(1 - \chi)(r_\tau - 1)^2 + \theta^2 \chi (r_\tau / \mu_\tau - 1)^2 + \chi \psi_\tau^2 \sigma_\varepsilon^2 \right]$$

where  $\sigma_\varepsilon^2$  is the variance of  $\varepsilon$ . The same problem with an unconditional expectation on  $\theta$  (i.e., an ex ante choice of method) would replace  $\theta^2$  with the fundamental's variance  $\sigma_\theta^2$ .

Let us now prove that, for fixed  $(\theta, \lambda, \sigma)$ , optimal policy is characterized by a threshold rule. For target communication, the appropriate translation of the loss function is

$$\mathcal{L}_Y \equiv \min_{r_Y \in \mathbb{R}} [\theta^2(1-\chi)(r_Y - 1)^2 + \theta^2\chi(r_Y/\mu_Y - 1)^2 + (1-\chi)\psi_Y^2\sigma_\varepsilon^2]$$

Via an identical argument to the one pursued in the proof of Theorem 1, each loss function is monotone in  $\gamma$  holding fixed the value of  $\psi_X$ . That is to say, if the loss functions were each re-written in the form  $\mathcal{L}_X = \ell_X(\mu_X, \psi_X)$  then

$$\frac{\partial \ell_\tau}{\partial \mu_\tau} \frac{\partial \mu_\tau}{\partial \gamma} > 0 \quad \frac{\partial \ell_Y}{\partial \mu_Y} \frac{\partial \mu_Y}{\partial \gamma} < 0 \quad (25)$$

Showing the equivalent monotonicity via the second channel is simple using the previously proven comparative static for the  $\psi_X$ :

$$\frac{\partial \ell_\tau}{\partial \psi_\tau} \frac{\partial \psi_\tau}{\partial \gamma} = 2\chi\sigma_\varepsilon^2\psi_\tau \frac{\partial \psi_\tau}{\partial \gamma} > 0 \quad \frac{\partial \ell_Y}{\partial \psi_Y} \frac{\partial \psi_Y}{\partial \gamma} = 2\chi\sigma_\varepsilon^2\psi_Y \frac{\partial \psi_Y}{\partial \gamma} < 0 \quad (26)$$

and combining (25) and (26) is sufficient to prove

$$\frac{\partial \ell_\tau}{\partial \gamma} > 0 \quad \frac{\partial \ell_Y}{\partial \gamma} < 0$$

Finally, note that previous argument about continuity and extreme values from the proof of Theorem 1 also carries over. In particular,

$$\mathcal{L}_\tau|_{\gamma=0} = \mathcal{L}_Y|_{\gamma=0} = 0$$

and

$$\mathcal{L}_\tau|_{\gamma=1} > 0 \quad \mathcal{L}_Y|_{\gamma=1} > 0$$

with the latter strictly true so long as  $\sigma > 0$  or  $\theta \neq 0$ . Hence, generically speaking (for  $\sigma > 0$  or  $\theta \neq 0$ ), there exists a  $\hat{\gamma} \in (0, 1)$  such that target communication is strictly optimal when  $\gamma \in (\hat{\gamma}, 1]$ , instrument communication is optimal when  $\gamma \in [0, \hat{\gamma})$ , and the policymaker is indifferent at  $\gamma = \hat{\gamma}$ .

## Proof of Corollary 6

Let us finally consider the case of uncertain distortions. Note that, for every  $(\lambda, \sigma)$  and every value of  $\theta \neq 0$ ,  $\gamma \mapsto \mathcal{L}_\tau(\cdot; \lambda, \sigma)$  and  $\gamma \mapsto \mathcal{L}_Y(\cdot; \lambda, \sigma)$  have all the relevant monotonicity and limit-value properties because of Proposition 10 (see also the proof thereof). It follows that the average  $\gamma \mapsto \mathbb{E}_{\pi(\lambda, \sigma)}[\mathcal{L}_X(\cdot; \lambda, \sigma)]$ , where the relevant expectation is over possible values of  $(\lambda, \sigma)$  in accordance with prior  $\pi$ , maintains the same properties. Hence these “expected loss functions” must cross at some  $\hat{\gamma} \in (0, 1)$ , completing the proof.

## Proofs of Propositions 11 and 12

For a policy rule to implement the first best, it is necessary that  $T(\theta, \theta) = \theta$ , which restricts  $a$  and  $b$  as follows:

$$a = (1-b)\theta.$$

We can henceforth focus on the class of policy rules that satisfy this restriction. This is a one-dimensional class indexed by  $b$ .

Solving (14) and (1) jointly for  $\tau$  and  $Y$ , using  $a = (1 - b)\theta$ , and substituting the solution into (2), we obtain the following game representation for the agents' behavior in stage 1:

$$k_i = (1 - \delta)\theta + \delta \mathbb{E}_i[K] \quad (27)$$

where

$$\delta = \delta(b; \alpha, \gamma) \equiv \frac{\alpha(\gamma + b(1 - \gamma))}{1 - (1 - \alpha)b}.$$

Note that  $\delta \in (-1, +1)$  if and only if  $b \in (\underline{b}, +1)$ , where  $\underline{b} \equiv \frac{1+\alpha\gamma}{1-2\alpha+\alpha\gamma} < -1$ . This explains the assumed bounds imposed on  $b$ : outside these bounds, “reasoning fails to converge” (this is the present analogue of the restriction  $\delta_X \in (-1, +1)$ ,  $X \in \{\tau, Y\}$ , in the baseline analysis).

Consider now the case with rigid beliefs (as in Section 4). Iterating the best response (27) yields the unique equilibrium average action as

$$K = \sum_{h=1}^{\infty} \delta^{h-1} \lambda^{h-1} \zeta \theta = \frac{1 - \delta}{1 - \lambda \delta} \theta, \quad (28)$$

with  $\delta = \delta(b; \alpha, \gamma)$  defined above. For this to coincide with the first best action, it is therefore necessary and sufficient that

$$\frac{1 - \delta}{1 - \lambda \delta} = 1$$

Clearly, this is automatically satisfied when  $\lambda = 1$  (rational expectations), regardless the value of  $\delta$ , or equivalently of  $b$ . This verifies the indeterminacy of the optimal policy rule under rational expectations. When instead  $\lambda < 1$ , the above is satisfied if and only if  $\delta = 0$ , or equivalently  $b = -\gamma/(1 - \gamma)$ . Along with  $a = (1 - b)\theta$ , this completes the characterization of the unique policy rule that implements the first best once  $\lambda < 1$ .<sup>41</sup>

### Proof of Proposition 13

Note that  $\bar{\mathbb{E}}[K] = \mathbb{E}_i[K] = \lambda K$  by an argument essentially identical to the one supporting Lemma 3. Evaluated at the equilibrium under optimal policy, this is  $\bar{\mathbb{E}}[K] = \lambda \theta$ . The expected policy instrument and outcome, evaluated at the optimal rule, are

$$\begin{aligned} \bar{\mathbb{E}}[\tau] &= \frac{\theta}{1 - \gamma} - \frac{\gamma}{1 - \gamma} \bar{\mathbb{E}}[Y] \\ \bar{\mathbb{E}}[Y] &= (1 - \alpha) \bar{\mathbb{E}}[\tau] + \alpha \lambda \theta \end{aligned}$$

Solving this system of equations gives

$$\begin{aligned} \bar{\mathbb{E}}[\tau] &= \frac{1 - \lambda \alpha \gamma}{1 - \alpha \gamma} \cdot \theta \\ \bar{\mathbb{E}}[Y] &= \frac{1 - \alpha + \alpha \lambda (1 - \gamma)}{1 - \alpha \gamma} \cdot \theta \end{aligned}$$

---

<sup>41</sup>Clearly, the argument extends to  $\lambda > 1$ , modulo the re-interpretation of the friction along the lines of Section 6.2.

Recall also that the equilibrium satisfies  $\tau = Y = \theta$ .

The instrument forecast gap as a function of  $\gamma$  is

$$f_\tau(\gamma) \equiv |\tau - \bar{\mathbb{E}}[\tau]| = \left| \frac{(1-\lambda)}{1-\alpha\gamma} \alpha\gamma \right| \cdot |\theta|$$

which satisfies  $f_\tau(0) = 0$  and  $f'_\tau(\gamma) > 0$  for  $\gamma \in (0, 1)$ . This obviously implies  $f_\tau(\gamma) > 0$  for  $\gamma \in (0, 1)$ . Similarly the target forecast gap is

$$f_Y(\gamma) \equiv |Y - \bar{\mathbb{E}}[Y]| = \left| \frac{(1-\lambda)}{1-\alpha\gamma} \alpha(1-\gamma) \right| \cdot |\theta|$$

which satisfies  $f_Y(1) = 0$  and  $f'_Y(\gamma) < 0$  for  $\gamma \in (0, 1)$ . This obviously implies  $f_Y(\gamma) > 0$  for  $\gamma \in (0, 1)$ .

Finally, going slightly beyond the original statement of the Proposition, it is simple to see that the instrument gap is always smaller for  $\gamma < 0$ :  $f_\tau(\gamma) > f_Y(\gamma)$  since  $|1-\gamma| > |\gamma|$ . This generalizes our results for the optimality of instrument communication with negative GE effects.

## B Micro-foundations

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

### B.1 A New Keynesian economy

We first model a liquidity trap with a binding zero-lower-bound through a New Keynesian model. As noted in the main text, this nesting depends on strong simplifying assumptions. The goal is only to facilitate an appealing interpretation of our insights. A careful adaptation of our analysis to the full New Keynesian model is beyond the scope of this paper.

#### B.1.1 Set-up

There are countably infinite periods, indexed by  $t \in \{0, 1, 2, \dots\}$ . As in the abstract model, period 0 exists only to index the time of forward guidance. Period 1 is the liquidity trap. Period 2 is the “future” about which the central bank would like to communicate. Periods 3 to infinity correspond to a return to steady steady state.

There is a unit measure of households, or consumers, indexed by  $i \in [0, 1]$ . Household  $i$  consumes  $C_{i,t}$  of the good and works  $N_{i,t}$  hours in period  $t$ . Their utility function is

$$\mathcal{U}_{i,t} = \mathbb{E}_i \left[ \sum_{t=1}^{\infty} \beta_t \left( \log C_{i,t} - \frac{N_{i,t}^2}{2} \right) \right]$$

where  $\beta_t = \exp(-\sum_{s=1}^t \rho_s)$ , with  $\rho_j$  denoting the discount rate between periods  $s$  and  $s+1$ . Each consumer also faces a standard flow budget constraint in terms of her asset level  $A_{i,t}$ , income  $Y_{i,t}$ , and real interest rate  $R_t$ :

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

The assets are in zero net supply. Labor hours and income are the same for all agents, so, so  $N_{i,t} \equiv N_t$  and  $Y_{i,t} \equiv Y_t$  in equilibrium.

There is a representative firm that produces with linear technology  $Y_t = e^{\theta_t} N_t$ , where  $\theta_t$  is a measure of log total factor productivity (TFP). The producer of this good charges a constant price normalized to one. Output is thus totally demand determined, and the firm hires labor to meet its demand.

A monetary authority controls the nominal interest rate. Because goods prices are perfectly rigid, this also entails direct control of the real interest rate  $R_t$ . The monetary authority faces a zero lower bound constraint on the nominal interest rate, which equals the real interest rate, so  $R_t \geq 1$ .

Standard arguments show that the efficient benchmark of such an economy features productive efficiency, with  $Y_t^* \equiv e^{\theta_t} \bar{N}$  for some constant  $\bar{N}$ . The implementing monetary policy, ignoring lower bound constraints, fully stabilizes the demand shock and allows output to fluctuate with the productivity shock, or  $R_t^* \equiv \exp(\rho_t - \theta_t)$ .

**Thought experiment: liquidity trap.** We model the “liquidity trap scenario” in the following way. For all  $t \geq 2$ , the subjective discount is constant at a steady state-value  $\rho_t = \bar{\rho} > 0$  and the gross natural rate of interest is similarly constant at  $\bar{R} \equiv \exp(\bar{\rho}) > 1$ . At  $t = 1$ , the discount rate is constant and weakly negative, or  $\rho_1 = \underline{\rho} \leq 0$ . During this liquidity trap period, a zero lower bound (ZLB) becomes binding, or  $R_1 = 1$ .

TFP is equal to its steady-state level at  $t = 1$  and for all  $t \geq 3$ . TFP at period 2 equals some random  $\theta$ . The monetary authority learns the value of  $\theta$  at  $t = 0$  and can condition their announcement of future conditions on this piece of information. Note that the level of TFP *itself* has no instrumental value for any agent in the economy to know. Consumers care just to know demand, which determines output; and firms are stuck with a constant price and forced also to track demand. We return to alternate interpretations for the shock  $\theta$  at the end of this section.

Finally, by way of normalization, we assume that the monetary authority always returns to the steady-state level of interest rates (which, by assumption, here equals the natural rate) for  $t$  after some fixed  $T < \infty$ .

The policymaker at  $t = 0$ , more specifically, has two options for forward guidance. First, they can announce a path of interest rates from  $t = 2$  to  $T$ , or  $\{R_t\}_{t=2}^T$ . Second, they can announce a targeted value for output during and just after the liquidity trap (which will be more clearly defined later). Crucially, they can adjust the interest rate path  $\{R_t\}_{t=2}^T$  to meet this target ex post.

We assume the following specific structure for beliefs and choices. First, consumers have mis-specified beliefs about each others’ knowledge and rationality *during the liquidity trap* (at  $t = 1$ ). Second, all agents have common knowledge about the common preference shock. Finally, from period 2 onward, the economy returns to common knowledge of policy (the “crisis of confidence” passes), and this fact itself is common knowledge at the beginning of the liquidity trap.

### B.1.2 Mapping to the abstract model

Let all lowercase variables now be in log deviations from the steady state in which  $\rho_t \equiv \bar{\rho}$ ,  $\theta_t \equiv 0$ , and  $R_t \equiv \bar{R} = \exp(\bar{\rho})$ . We will first solve for equilibrium in these log-deviation terms.

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

$$c_{i,t} = (1 - \beta) \sum_{j=0}^{\infty} \beta^j \mathbb{E}_{i,t}[y_{t+j}] + \beta \sum_{j=0}^{\infty} \beta^j \mathbb{E}_{i,t}[-(r_{t+j} - (\varrho - \bar{\rho}))] \quad (29)$$

where  $\beta \equiv \exp(-\bar{\rho})$  is the steady-state discount factor. This expression is obtained by substituting the consumer's Euler equation into his lifetime budget constraint and solving for consumption. It represents the optimal consumption in period  $t$ , as a function of the expected path of income and interest rates, and identifies  $1 - \beta$  with the marginal propensity to consume out of income. (More on this interpretation below.)

Let us first derive consumption and income at  $t = 2$ . We assume that, at this point, all agents have the same (rational) expectations. It is simple to apply forward-looking rational expectations in (29) to get

$$c_{i,2} = y_t = - \sum_{s=2}^T r_s$$

Let  $\tau \equiv - \sum_{s=2}^T r_s$  denote this (negative) sum of future interest rates, which is a sufficient statistic for policy. A higher  $\tau$  corresponds to looser monetary policy after the trap.

Now we can solve for consumption in period 1, after substituting out income in all periods 3 and after. Given that  $r_1 = -\bar{\rho}$  in deviation from the steady state, this reduces to the following:

$$c_{i,1} = (1 - \beta) \mathbb{E}_{i,1} [y_1 + \beta y_2 + \beta^2 \tau] + \beta \varrho \quad (30)$$

Summing up, the above is the optimal consumption of individual  $i$  in period 1, as a function of her expectations of her income and of post-trap interest rates. But recall that income is demand-determined and hence positively related on the consumption choices of others. This suggest that this model has the same basic formal structure as our abstract model. The precise mapping is spelled out below. And why this mapping is exact only with the simplified version of the New Keynesian model used here, the logic is more general. See, for instance, [Angeletos and Lian \(2018\)](#) for a detail exposition of how both the Dynamic IS curve and the New Keynesian Philips curve can be represented as a pair of dynamic games of strategic complementarity.

**Counterpart of equations (1) and (2).** Let

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

be a (normalized) measure of total output during and right after the liquidity trap. Let

$$k_i \equiv c_{i,1} - \varrho$$

be individual consumer spending in the first period of the trap relative to its no-guidance level; and let the average thereof be  $K \equiv \int_i c_i \, di - \varrho$ . Finally, as stated previously, let

$$\tau \equiv - \sum_{s=2}^T r_s$$

be the negative of the interest rates after the trap. Because  $y_2 = \tau$ , we can re-write the definition of  $Y$  in the following form:

$$Y = \frac{\beta}{1+\beta}\tau + \frac{1}{1+\beta}K$$

which matches condition (1) in our abstract framework for  $\alpha = \frac{1}{1+\beta}$ . Condition (30), on the other hand, can be written up to a constant as

$$k_i = \beta^2 \mathbb{E}_{i,1}[\tau] + (1 - \beta^2) \mathbb{E}_{i,1}[Y]$$

which matches condition (2) in our abstract framework for  $\gamma = 1 - \beta^2$ .

**Interpreting the GE complementarity.** The GE complementarity is herein given by  $\gamma = 1 - \beta^2$  and is thus higher when  $\beta$  is lower. Literally taken,  $\beta$  is the steady-state discount factor. But this is not the relevant interpretation here. Instead, inspection of condition (30) reveals that  $(1 - \beta)$  is the marginal *elasticity* to consume out of income during the liquidity trap. And because there is no capital (and hence  $C = Y$ ), this is also the marginal propensity to consume (MPC).

Of course, with the complete-markets, infinite-horizon formulation adopted here (and in the textbook version of the Permanent Income Hypothesis alike), the MPC is itself given by 1 minus the subjective discount factor. But in more realistic models, this relation does not hold. For instance, a perpetual-youth, OLG extension of the present model along the lines of [Del Negro, Giannoni and Patterson \(2015\)](#) preserves condition (30) modulo the replacement of  $\beta$  with  $\beta\zeta$ , where  $\zeta$  is one minus the probability of death. Furthermore, [Farhi and Werning \(2019\)](#) explain how the probability of “death” can be re-interpreted as the probability of “exit” from markets, or of binding liquidity constraints.

Under this light, the value of  $1 - \beta$  in our simple model is best interpretable as the MPC, and a lower  $\beta$  mimics tighter borrowing constraints. By the same token, the relation  $\gamma = 1 - \beta^2$  can be read as “the Keynesian multiplier increases with the MPC.”

**Interpreting the policy instrument.** The policy instrument is the negative sum of all future interest rates to which the central bank can commit. There are several interest rate paths that could implement a given value of  $\tau$ , and different “parametrization” of these paths might be more interpretable from a policy perspective.

Consider first the policy of holding nominal rates at zero, or setting  $r_t = -\rho$ , for  $h$  periods before returning to the natural rate,  $r_t = 0$ . In this case,  $\tau = -h\rho$  and  $\tau$  captures “the timeline for staying at the ZLB after the liquidity trap.”

Consider second the policy of committing to return to the natural rate at  $t = 2$ . In this case,  $\tau = -r_2$  and  $\tau$  captures the “aggressiveness of the lift-off.”

### B.1.3 A micro-founded policy objective

Let us now return to the policy objective. Assume first that  $\rho = 0$ , so the first-best allocation is shifted only by TFP and features  $y_1^* = \theta$ ,  $y_2^* = \theta$ , and  $\tau^* = -r_2^* = \theta$ . This is consistent with our view that bounded rationality is the only source of inefficiency, or that the first-best could be implemented with rational expecta-

tions. We take such an extreme stance to focus on normative analysis on the role of bounded rationality, though very shortly we will discuss how the  $\rho < 0$  case might play out in our example.

Assume also that policy implementation for  $t \geq 2$  is achieved only by manipulating the interest rate at  $t = 2$ , so all future gaps of output from the first-best level are 0. Up to scale, a utilitarian planner's loss function in this economy is

$$L = L(y_1, y_2; \theta) = (y_1 - \theta)^2 + \beta(y_2 - \theta)^2 \quad (31)$$

In terms of the abstract model's tuple  $(K, \tau)$ , the above can be restated as

$$L = L(\tau, K; \theta) = (K - \theta)^2 + \beta(\tau - \theta)^2 \quad (32)$$

or, in terms of  $(Y, \tau)$ , as

$$L = L(\tau, Y; \theta) = (1 + \beta)(Y - \theta)^2 + \beta(\tau - \theta)^2 - 2\beta(Y - \theta)(\tau - \theta) \quad (33)$$

**Mapping to the abstract loss function.** The loss function (33) maps essentially to our abstract objective (3). The additional “cross-term”  $-2\beta(Y - \theta)(\tau - \theta)$  does not appear in the abstract loss function we assume in the main text, but its presence would not have upset any of the insights. Indeed, we can just as easily prove our main result (Theorem 1) by using directly the micro-founded objective in (32), instead of either (3) or (33), because the trade-off between the  $\tau - \theta$  and  $K - \theta$  gaps share the same essential monotonicity properties we highlight in the proof for the trade-off between the  $\tau - \theta$  and  $Y - \theta$  gaps. This possibility is discussed at the end of the proof of Theorem 1 in Appendix A.

**Interpretation of policy shifter  $\theta$ .** Our model treats the period 2 productivity  $\theta$  as a pure shifter of the policymaker's preference. Although TFP does not *generically* fulfill this role in micro-founded models, it does in this particular model with no production decisions. This makes our abstraction from the signaling role of policy internally consistent.

We are open also to broader interpretations of what  $\theta$  means. A first possibility, which would amount to a “re-labeling” of the present structure, allows  $\theta$  to be a shock to the *policymaker's* intertemporal preference relative to the household's. This amounts to the policymaker having a random “opinion” about the extent to which fluctuations are efficient, without tying the efficient component of those fluctuations with a primitive supply shock.

A second possibility equates  $\theta$  to a shock to the central bank's “long-run outlook” about issues that plausibly have a very small impact on consumer's choices today. Examples include costs to remaining at the ZLB from financial-sector instability, central-bank balance-sheet expansion, or effects on future inflation. We do not model such issues formally here.

**First-order considerations.** Let us now allow  $\rho < 0$ , or a strictly sub-optimal demand recession at  $t = 0$ . The loss function (32) becomes

$$L(K, \tau; \theta, \rho) = (K - \theta)^2 + \beta(\tau - \theta)^2 + 2\rho K \quad (34)$$

In our baseline model, as long as the policymaker makes announcements of the form  $\hat{X} = r_X\theta$  for  $r > 0$  and  $X \in \{\tau, Y\}$ , then  $K$  admits a form  $K = \kappa_X r_X \theta$  with  $\kappa_X r_X > 0$ . Thus the extra first-order term can be re-written as  $2\rho\kappa_X r_X \theta$ .

Consider first the case of  $\theta > 0$ . The extra term in the loss function is always negative, and more so when  $\kappa_X r_X$  is large or  $K$  responds more to  $\theta$ . Intuitively, this favors target communication. When the policymaker is trying to induce a positive output level at  $t = 2$ , they would prefer a method which achieves more movement of output at  $t = 1$  to fight the distortionary demand recession. In fact, if  $\kappa_X r_X > 1$  (as it would be in our model), then target communication with bounded rationality does *better* than either method would do under rational expectations. The over-reaction afforded by bounded rationality allows the policymaker to achieve more at  $t = 1$  without unnecessarily heating up the economy at  $t = 2$  to keep the promise.

Now consider  $\theta < 0$ , which for the opposite argument favors low  $\kappa_X r_X$  and low responsiveness of  $K$  to  $\theta$ . Intuitively, this favors instrument communication. When the policymaker is needs to exogenously tighten at  $t = 2$  relative to baseline expectations (say, because there are unusually high costs of remaining at the ZLB in terms of financial instability), they would prefer a method which does not exacerbate the initial demand shortfall.

## B.2 A neoclassical economy with aggregate demand externalities

The second micro-foundation differs in its approach (Neoclassical), application (fiscal policy), and key decision (investment).

### B.2.1 Set-up

There are three periods,  $t \in \{0, 1, 2\}$ ; a continuum of firms or entrepreneurs,  $i \in [0, 1]$ , who choose investment at  $t = 1$ ; and a policymaker, who can subsidize production at  $t = 2$ . The first period,  $t = 0$ , identifies only the time of policy announcement.

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

The final-good firm operates at  $t = 2$ . Define the following CES aggregator of the intermediate goods

$$X \equiv \left( \int x_i^{1-\rho} \, di \right)^{\frac{1}{1-\rho}}$$

where  $\rho \in [0, 1]$  is the inverse elasticity of substitution. The final goods firm operates with a Cobb-Douglas technology over this intermediate and labor,  $Q = X^\eta N^{1-\eta}$ . Its revenue is

$$(1-r)Q - wN - \int p_i x_i \, di$$

where  $r$  is the rate of taxation and  $w$  is the wage, both in final good terms. Finally, the worker lives, works,

and consumes only in period  $t = 2$  and has utility  $v = wN - \frac{1}{1+\phi}N^{1+\phi}$ , where  $\phi > 0$  parameterizes the Frisch elasticity.

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

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

This implies that the revenue for the entrepreneur has the following form:

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

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

Labor supply has the following form:

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

Labor demand is set by the final-goods firm:

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

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

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

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

The first-order condition re-arranges to

$$x_i^\rho = \eta(1-\rho)\mathbb{E}_i [(1-r)X^{\eta+\rho-1}N^{1-\eta}] \quad (35)$$

Investment solves this fixed-point equation.

**REE benchmark.** Assume rational expectations. In equilibrium, the agent will conjecture that  $x_{-i} = x_i \equiv X$ . Since everything is now known, we can pull  $X$  out of the expectation and solve to get

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

It is immediate that output is linear in labor:

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

Setting labor supply to labor demand gives

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

and plugging that back into the equation for output gives

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

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

### B.2.2 Mapping to the abstract model

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 35 can no longer be solved without expectations. To make progress, we will take log-linear approximations around  $r = 0$ . Let  $(\bar{Q}, \bar{N}, \bar{X})$  denote output, labor, and investment evaluated at this point. Let  $Y = \log Q - \log \bar{Q}$  and  $n = \log N - \log \bar{N}$  be log deviations of the first two quantities. Further, define  $k_i = \frac{1+\eta\phi}{1+\phi}(\log x_i - \log \bar{X})$  and  $\tau = \frac{1+\eta\phi}{\phi(1-\eta)} \log(1-r)$  be convenient monotonic transformations of investment and the tax, respectively, and  $K = \int_i k_i di$  be the aggregate (log deviation) rescaled investment.

Aggregate production is log-linear:

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

Equilibrium labor is

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

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

$$Y = (1-\alpha)\tau + \alpha K \quad (36)$$

with

$$\alpha \equiv \frac{\eta(1+\phi)^2}{(\eta+\phi)(1+\eta\phi)} \quad (37)$$

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

Let us now turn to the investment decision (35). To a log-linear approximation, it is

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

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

$$k_i = (1-\gamma) \mathbb{E}_i[\tau] + \gamma \mathbb{E}_i[Y] \quad (38)$$

for feedback parameter

$$\gamma \equiv \frac{(1+\eta\phi)(\rho(\eta+\phi) - \phi(1-\eta))}{\eta\rho(1+\phi)^2} \quad (39)$$

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

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

The feedback parameter is positive if and only if

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

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

### B.2.3 On the policymaker's objective

In the previous model that the first-best is implemented if  $Y = \tau = 0$ , or no distortionary taxes are imposed upon the economy. To make progress and allow some reason for government intervention, we might imagine  $\theta$  as a random shifter of the (negative) value of tax revenues for the government. This could encompass political considerations and/or the marginal value of tax revenue for other applications in the economy. The policy objective would then encode the trade-off between raising revenues and distorting production and labor supply decisions at  $t = 2$ .

## C Miscellanea

In this appendix we fill in various details. First, we explain how our main specification can be recast as a “smooth” version of Level-k Thinking. Second, we explain why two other forms of communication, announcing  $\theta$  or committing to a value for  $K$ , are ill posed. Third, we elaborate on the role played by the assumption  $\alpha < \frac{1}{2-\gamma}$ . Fourth, we provide a suggestive, back-of-the-envelope calculation of the critical threshold for the MPC that justifies target communication in our stylized version of the New Keynesian model. Finally, we describe an extension that combines our logic with that of [Poole \(1970\)](#).

### C.1 Level-k Thinking

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

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

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

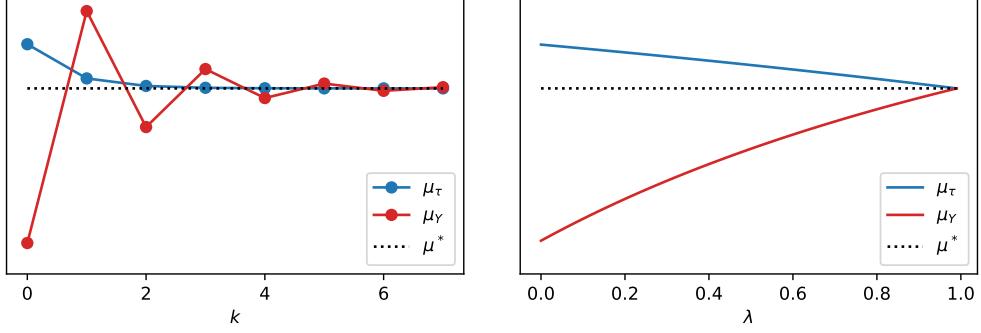


Figure 2: The implementability coefficients  $\mu_\tau$  and  $\mu_Y$  under Level- $k$  Thinking (left) and rigid beliefs (right).

$$\tilde{\mathbb{E}}[K] = (1 - \delta_X) \sum_{h=0}^{k-1} (\delta_X)^h \hat{X} = (1 - (\delta_X)^k) \hat{X} \quad (40)$$

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

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

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

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

The same goal can be achieved with a “smooth” version of Level- $k$  Thinking along the lines of [Garcia-Schmidt and Woodford \(2019\)](#). The concept of “cognitive discounting” introduced in [Gabaix \(2019\)](#) works in a similar manner, too, because it directly postulates that the subjective expectations of endogenous variables such as  $K$  move less than the rational expectations of it.

## C.2 Communicating other objects

Our focus on communicating  $\tau$  or  $Y$  seemed natural for applications. But, for completeness, we discuss here the possibilities of committing to a target for the aggregate action  $K$  or communicating the realized value of  $\theta$  along with (or perhaps instead of) a policy plan.

### C.2.1 Communicating a target for $K$

Consider the scenario in which the policymaker commits to a target for  $K$ , instead of a value for  $\tau$  or  $Y$ . 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 is not well-posed in our model.

Consider in particular the specification studied in Section 4 and let the policymaker announce and commit to a value  $\hat{K}$  for aggregate investment. Assume that first-order beliefs about investment are correct ( $\bar{\mathbb{E}}[K] = \hat{K}$ ) and higher-order beliefs are sticky around 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  $\tau$  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  $\tau$ :

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

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  $\tau$ .

At the same time, agent  $i$  knows that, like himself, all attentive agents expect  $\tau$  to coincide with  $\hat{K}$ . And since agent  $i$  believes that the fraction of attentive agents is  $\lambda$ , 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 it is feasible to announce and commit to any  $\hat{K}$  other than 0 (the default point).

In a nutshell, the problem with communicating  $K$  is that the policymaker has no direct control over it. From this perspective, output communication worked precisely because the policymaker had some plausible commitment. Agents could rationalize  $Y = \hat{Y}$  regardless of their beliefs about  $K$  because there always existed some level of  $\tau$  that implemented  $\hat{Y}$ . We alluded to the failure of this mechanism as  $\alpha \rightarrow 1$ , and the direct effect of policy vanished, in our baseline model (Section 5.4).

### C.2.2 Expanding the message space

Return to the case in which the policymaker commits to a value for  $\tau$  or  $Y$  (conditional on  $\theta$ ), but allow her to provide an additional message of the form

$$m = a\theta + b\varepsilon$$

where  $a, b \in \mathbb{R}$  and  $\varepsilon$  is an arbitrary random variable. This could capture a perfect or imperfect signal of the fundamental, a “justification for the policy choice,” or some other arbitrary message. Let Assumption 3 apply to the vector  $(\theta, X)$ .

It is obvious that the additional message plays no role in the best response (9) or the expansion (11); does not enter the expression for  $K$ ; and thus does not affect the implementability constraint. Hence the implementable sets are the same as the ones given for instrument and target communication in Proposition 3. The messages provide no extra flexibility.

### C.2.3 Communicating only $\theta$

What about communicating a message *without* a policy plan? In particular, what if the policymaker communicates *only* the value of  $\theta$ ? In general, agents may have no idea what  $\theta$  means, or how to map its announced value to an expectation for  $\tau$  and  $Y$ . So, unless additional assumptions are made, this scenario is ill-posed.

One way to close this scenario is to assume that the agents have knowledge of the policymaker’s *entire* problem, namely her objective as given in (3), her set of options (pick a value for  $\tau$  or one for  $Y$ ), and her beliefs about the structure of the economy. The agents could then use this knowledge along with the announcement of  $\theta$  to figure out the policymaker’s choices. This would only replicate the outcomes of our baseline analysis, in a indirect and uninteresting way.

This is is true, of course, insofar as the policymaker’s problem remains the same as in our baseline analysis: the policymaker is still committing to a value for  $\tau$  or  $Y$ , although “secretly” so. If, instead, the policymaker lacks commitment, they will expect her to play a different strategy. This takes us to the territory of Section 8, where we explain why commitment is essential for regulating the bite of bounded rationality.

## C.3 The restriction $\alpha < \frac{1}{2-\gamma}$ , or $|\delta_X| < 1$

Most of our analysis restricts  $\alpha < \frac{1}{2-\gamma}$  so as to guarantee that  $-1 < \delta_X < 1$  for both form of forward guidance. This allows the characterization of beliefs and behavior by repeated iteration of the best responses. In

particular, in Section 3 it guarantees that the joint of Assumptions 1 and 2 replicates the REE benchmark; in Appendix C.1, it guarantees that the Level-k outcome converges to the REE outcome as agents become “infinitely rational” ( $k \rightarrow \infty$ ); and in Sections 4 and 6.2, it guarantees that Assumptions 3 and 4 yield the corresponding PBE outcomes.

We think it is most reasonable to restrict to  $|\delta_X| < 1$ . But for completeness we discuss here what happens otherwise. Consider first an “adversarial” selection of outcome in this case. This will only strengthen the case for the main results. For  $-1/\alpha < \gamma < 2 - 1/\alpha$ , we have  $\delta_Y < -1$  and  $\delta_\tau \in (-1, 1)$ . Instrument communication would clearly be preferred to prevent arbitrarily poor outcomes under target communication. To use an analogy which applies directly in our extension that considers policy rules (Section 8), this is like picking a policy that obeys the Taylor principle over one that does not. For  $\gamma < -1/\alpha$ , we have both  $\delta_Y < -1$  and  $\delta_\tau < -1$ , so theory lacks a clear prediction under either communication strategy.

#### C.4 Back-of-the-envelope calculation

While our version of the New Keynesian model in Appendix B.1 is admittedly very simple, it can offer a hint at how steep the Keynesian cross needs to be to justify target communication. Let us focus on the vanishing friction limit for the critical GE feedback, given in Corollary 5 as  $\hat{\gamma}_{\text{vf}} \equiv \lim_{\lambda \rightarrow 1} \hat{\gamma} = \frac{1}{2-\alpha}$ . After substituting in the micro-founded values of  $\gamma$  and  $\alpha$ , we get the optimality condition for target communication,  $\gamma > \hat{\gamma}_{\text{vf}}$ , reduces to

$$1 - \beta > 0.5$$

To interpret this condition, note first that  $1 - \beta$  corresponds with the marginal elasticity to consume out of income earned during in the liquidity trap, which in a model with no capital (i.e.,  $C = Y$ ) also equals the local marginal propensity to consume over the same period. Note next that, for model calibration, this value depends on the *length* of the period (i.e., the length of the liquidity trap).

Assume, as an example, that one period in our model corresponds to 3 years. Let  $m$  denote the annual MPC. The corresponding three-year MPC is given by

$$1 - \beta = (m)(1 + (1 - m) + (1 - m)^2) = 1 - (1 - m)^3.$$

It follows that the optimality condition for target communication can be re-expressed in terms of the annual MPC as follows:

$$m > m^* \approx 0.21.$$

An annual MPC, for the average household, in this range is quite reasonable according to the available micro-data studies (see, e.g., the review in [Jappelli and Pistaferri, 2010](#)).

#### C.5 Connection to [Poole \(1970\)](#)

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 the quantities  $(\tau, Y)$  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  $\tau$  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.

**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  $\sigma_u^2$ , is orthogonal to  $\theta$ , 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  $\tau$  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  $\tau$ , which is in turn the case whenever  $\chi$  is high enough.<sup>42</sup>

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 the form of forward guidance: Lemmas 1 and 2 remain intact. By the same token, when  $\lambda = 1$ , the sets of the implementable  $(\tau, Y)$  pairs remain invariant to  $\gamma$ , even though they now depend on the realization of  $u$ . It then also follows that, as long as  $\lambda = 1$ , the optimal communication strategy does not depend on  $\gamma$ .

As soon as  $\lambda < 1$ , the implementability sets and the optimal communication strategy start depending on  $\gamma$ , for exactly the same reasons as those explained before. To make this more clear, note that the implementable set for instrument communication is

$$\{(\tau, Y) : Y = \mu_\tau^{-1}\tau + u\}$$

---

<sup>42</sup>The above scenario has maintained that the ideal level of output is  $Y^{\text{fb}} = \theta$ . What if instead  $Y^{\text{fb}} = \theta + u$ ? 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 instrument communication always dominates target communication with rational expectations. 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 \(1970\)](#).

which means the policymaker, free to choose announcement  $\tau = r_\tau \theta$ , can implement  $(\tau, Y)$  pairs of the form  $(r_\tau \mu_\tau^{-1} \theta, r_\tau \theta + u)$ .

Consider a policymaker who must commit ex ante, before the realization of  $\theta$ , to either instrument or target communication and a mapping from  $\theta$  to their announcement  $\hat{X}$ . This is a slightly different assumption than our main analysis, but an appropriate translation of the classic Poole problem. We could just as easily have assumed contingency on  $\theta$  but *not*  $u$ , with the minor change that optimal policy now depends on the realization of  $\theta$  in place of its ex ante variance.

The appropriate translation of the loss function is

$$\mathcal{L}_\tau \equiv \min_{r_\tau \in \mathbb{R}} \left[ \sigma_\theta^2 [(1 - \chi)(r_\tau - 1)^2 + \chi(r_\tau / \mu_\tau - 1)^2] + \chi \sigma_u^2 \right]$$

where  $(\sigma_u^2, \sigma_\theta^2)$  are the respective variances of  $u$  and  $\theta$ . To re-iterate, were policy contingent on realized  $\theta$ , the same would apply with  $\theta^2$  in place of  $\sigma_\theta^2$ .

For target communication, the implementable set is

$$\left\{ (\tau, Y) : \tau = \mu_Y Y - \frac{u}{1 - \alpha} \right\}$$

which means that, for announcement  $Y = r_Y \theta$ , the policymaker can implement  $(\tau, Y)$  pairs of the form  $(r_Y \mu_Y \theta - u / (1 - \alpha), r_Y \theta)$ . The appropriate translation of the loss function is

$$\mathcal{L}_Y \equiv \min_{r_Y \in \mathbb{R}} \left[ \sigma_\theta^2 [(1 - \chi)(r_Y - 1)^2 + \chi(r_Y / \mu_Y - 1)^2] + \frac{1 - \chi}{(1 - \alpha)^2} \sigma_u^2 \right]$$

Note that the extra terms that appear in the loss functions for  $\sigma_u^2 > 0$  have no dependence on  $\gamma$ . To map to the loss functions plotted in Figure 1 as a function of  $\gamma$ , each loss function is shifted above, but neither “twists” or loses its monotonicity in  $\gamma$ .

Finally, note that in expectation both implementable sets are the same as the ones that are presented in Proposition 3. In particular, when  $\lambda = 1$  and  $\mu_\tau = \mu_Y = 1$ , the implementable sets are the same as the rational-expectations ones in Proposition 1. This demonstrates Proposition 8.

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

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

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

By combining the above with condition (1), we infer that, under both form of forward guidance, the following restriction has to hold:

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

where

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

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

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

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

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

## D Inattention vs. Distorted Reasoning

Our main analysis allows people to imperfectly reason about equilibrium, which is the friction of interest, but abstracts from the possibility that people are inattentive to forward guidance. In this appendix, we accommodate this possibility and study how it matters, or does not matter, for our paper's lessons. In particular, we show that our main result (Theorem 1) remains intact if we let people be *rationally* inattentive and maintain our working hypothesis that the policymaker aims at getting the economy as close as possible to the rational-expectations outcome. But we also explore what happens away from this case.

### D.1 Implementability and distortions

We start with a reduced-form specification that let us flexibly incorporate both inattention and imperfect equilibrium reasoning. A specific micro-foundation in terms of information and priors will be provided in the next subsection.

Maintain that higher-order beliefs have the structure from Section 4, or

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

for some  $\lambda \in (0, 1]$  and all  $h \geq 2$ . But now let first-order beliefs satisfy

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

for some parameter  $q \in (0, 1]$ . Our analysis so far is nested by  $q = 1$ . Inattention, rational or not, is introduced by letting  $q < 1$ . Alternatively,  $q < 1$  can be interpreted as the main specification of “sparsity” employed in [Gabaix \(2014\)](#) and [Gabaix \(2019\)](#).

Behavior is still determined by the solution to the following game:

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

with  $X \in \{\tau, Y\}$  depending on the form of forward guidance. Aggregating this and replacing  $\tilde{\mathbb{E}}[X] = q$ , we get

$$K = (1 - \delta_X)qX + \delta_X\tilde{\mathbb{E}}[K],$$

which makes clear that aggregate behavior depends, not only on the average beliefs of  $K$ , but also on the average belief of  $X$ , which now moves less than to one-to-one with  $X$  insofar as  $q < 1$ . That said, the following property still holds:

$$\tilde{\mathbb{E}}[K] = \lambda K.$$

This makes clear that  $\lambda$  alone pins down the perceived responsiveness of others relative the truth.

Proposition 3 readily extends modulo the following change in the slopes of the implementability constraints:

$$\mu_\tau = \left(1 - \alpha + \frac{1 - \alpha\gamma}{1 - \alpha\gamma\lambda} \alpha q\right)^{-1} \quad \text{and} \quad \mu_Y = \frac{1 - \alpha + \alpha(1 - \gamma)\lambda - \alpha(1 - \alpha\gamma)q}{(1 - \alpha)(1 - \alpha + \alpha(1 - \gamma)\lambda)}. \quad (41)$$

Instrument communication necessarily produces attenuation, or  $\mu_\tau > 1$ , because both frictions ( $q < 1$  and  $\lambda < 1$ ) work in the same direction. By contrast, the case for target communication is ambiguous ( $\mu_Y \leq 1$ ), because the amplification induced by rigid higher-order beliefs ( $\lambda < 1$ ) opposes the attenuation induced by inattention ( $q < 1$ ). Which effect dominates depends on the belief parameters ( $q, \lambda$ ) and the GE feedback  $\gamma$ , because the last interacts with rigid higher-order beliefs as explained in our main analysis.<sup>43</sup>

Finally, note that  $\mu_\tau = \mu_Y$  if and only if  $q = \lambda$ . In this knife-edge case, agents' perception of *all* variables, communicated directly or not, is uniformly damped by a single parameter and, as a result, we recover irrelevance of the instrument-versus-target choice. We allude to this fact in our discussion of clarity and confidence in Section 7.2—a model with “plain” inattention does *not* capture our desired feature of relaxing the most essential property of (full information) rational expectations equilibrium, which is the interchangeability of different objects that appear in the equilibrium allocation.

## D.2 A signal extraction model

To provide some more specific structure for what  $q$  and  $\lambda$  mean as independent parameters, consider the following model of inattention with a behavioral twist. Let the announcement  $X$  be Gaussian with mean 0 and known variance  $\sigma_X^2$ .<sup>44</sup> Each agent, because of their inattention, observes in effect a noisy signal  $s_i = X + u_i$ , where  $u_i$  is idiosyncratic Gaussian noise with mean 0. Agent  $i$  *perceives*  $u_i$  to have variance  $\omega^2$ ; the noise actually has variance  $\xi^2$ , where  $\xi$  may or may not be the same as  $\omega$  depending on whether the agent has the correct prior about his cognitive capacities. It follows, from simple signal-extraction math, that the agent's own expectation of  $X$  is

$$\mathbb{E}_i[X] = \frac{\sigma_X^2}{\sigma_X^2 + \omega^2}(X + u_i).$$

<sup>43</sup>Indeed, attenuation is obtained with target communication (i.e.,  $\mu_Y > 1$ ) if and only if  $q < \tilde{q}(\lambda, \gamma) \equiv \frac{1 - \alpha(1 - (1 - \gamma)\lambda)}{1 - \alpha\gamma}$ . The threshold  $\tilde{q}$  is increasing in both  $\lambda$  and  $\gamma$ , always exceeds  $\lambda$ , and reaches 1 when either  $\lambda = 1$  or  $\gamma = 1$ .

<sup>44</sup>This property will be maintained in the policy problem we consider if the underlying shock  $\theta$  is Gaussian with known variance, because  $X$  is itself proportional to  $\theta$  in equilibrium

This expression, averaged and then mapped to the reduced-form model for average expectations introduced in the previous subsection, gives  $\bar{\mathbb{E}}[X] = qX$  with

$$q \equiv \frac{\sigma_X^2}{\sigma_X^2 + \omega^2}.$$

When agents perceive their internal representations to have more noise (i.e.,  $\omega^2$  is higher),  $q$  becomes smaller and first-order beliefs are more attenuated. Note that there is no direct role for the actual noise variance  $\xi^2$  in determining the mean belief, which is sufficient for characterizing implementable allocations. Nonetheless, we can also define a “rational” signal-to-noise ratio, or

$$q^* \equiv \frac{\sigma_X^2}{\sigma_X^2 + \xi^2},$$

which is a benchmark to which we can compare the actual outcome whenever subjective perceptions diverge from reality. When agents over-estimate their cognitive capacities or the precision of their information,  $\omega^2 < \xi^2$  and  $q > q^*$ . When they make the opposite mistake,  $\omega^2 > \xi^2$  and  $q < q^*$ .

The above completes the description of how agents think about *themselves*. Let us now turn to how they think about *others*. Agent  $i$  perceives any other agent  $j$  to receive a signal of the form  $X + u_j$ , where  $u_j$  has mean 0 and variance  $\tilde{\omega}^2$ , which again may not equal the true variance  $\xi^2$ . Agent  $i$  believes further that agent  $j$  will associate variance  $\tilde{\omega}^2$  with the signals of agents  $k \neq j$ , and so forth. It is simple to show that second-order beliefs thus satisfy

$$\mathbb{E}_i[\mathbb{E}_j[X]] = \mathbb{E}_i\left[\frac{\nu^2}{\nu^2 + \tilde{\omega}^2}(X + u_j)\right] = \left(\frac{\nu^2}{\nu^2 + \tilde{\omega}^2}\right)\mathbb{E}_i[X]$$

Averaging and iterating this argument, we get  $\bar{\mathbb{E}}^h[X] = \lambda^{h-1}\bar{\mathbb{E}}[X]$  with

$$\lambda \equiv \frac{\nu^2}{\nu^2 + \tilde{\omega}^2}.$$

This scalar therefore depends exclusively on what each agent perceives to be the quality of *others*’ information.

Note now that  $q \in (0, 1)$  and  $\lambda \in (0, 1)$  are guaranteed respectively by  $\omega^2 > 0$  and  $\tilde{\omega}^2 > 0$ , or positive perceived variances. The case  $q > \lambda$  is guaranteed by  $\omega^2 > \tilde{\omega}^2$ , or a given agent believing he is more informed and/or attentive than the average other agent. The opposite case,  $q < \lambda$ , is associated with the opposite, or a given agent’s belief that others are more likely to be paying attention.

The canonical noisy rational expectations case is nested for  $\tilde{\omega}^2 = \omega^2 = \xi^2$ , or  $\lambda = q = q^*$ . But  $\omega^2 = \tilde{\omega}^2$ , or  $\lambda = q$ , alone is necessary for a model that, in terms of the equivalence between instrument and target communication irrelevance outcomes, is isomorphic to the noisy rational-expectations model.

Going back to the analysis of the previous subsection, recall that the implementability constraints depend only on  $q$  and  $\lambda$ , not on  $q^*$ . This is because, in a linear model such as ours, the actual level, or the value of  $q^*$ , does not matter at all for the *positive* properties of aggregate behavior; what matters is only people’s subjective view of the world. But as we explain below,  $q^*$  matters for judging the *normative* implications of any given behavior.

### D.3 Rational inattention or “one-distortion case”

Our baseline analysis and the loss function (3) compared all allocations to the full-information, rational-expectations allocation. This may not be appropriate in an environment with *rational* inattention, as in Sims (2003) and large follow up literature. Angeletos and Sastry (2019) show that the introduction of such inattention alone does not upset the Welfare Theorems: there is no policy that can improve upon market outcomes. Angeletos and La’O (2018) find a related result in a business-cycle model that allows dispersed private information be the source of both nominal and real friction.

The basic intuition is that there is no good reason for the policymaker to try to correct people’s behavior if any “friction” in it is merely the product of the agents’ optimal use of limited information or limited cognitive capacity. To capture this idea in reduced form, we now consider an altered policy problem that is “re-centered” around the rational expectations equilibrium (i.e., the one with *correct* perceptions of the noise variance).

Let us first consider the simplest such case, in which mis-perception of others’ variance is the only behavioral distortion. This means  $q = q^*$ , or agents correctly perceive their own variance, but  $\lambda \neq q = q^*$ , or agents mis-perceive others’ variance. Let  $\mu_{\text{in}}$  be the slope of the implementability constraint in a counterfactual world in which  $\lambda = q = q^*$ . Re-centering the policymaker’s objective around this reference point amounts to the following modification of the loss function:

$$L(\tau, Y, \theta) \equiv (1 - \chi)(\tau - \theta)^2 + \chi(Y - \theta/\mu_{\text{in}})^2. \quad (42)$$

This problem features only one distortion, relative over or under confidence, and thus resembles our baseline policy problem with a new “center point.” Given this adjustment in the relevant benchmark for optimality, we can prove that our main result is once again generic for  $\lambda \neq q$ :

**Proposition 15.** *Assume the combination of first and higher-order uncertainty described above and a policy objective that treats the noisy rational expectations equilibrium as the first-best. For  $q \leq 1$  and  $\lambda \neq q$ , there exists some critical threshold  $\hat{\gamma} \in [0, 1)$  such that target communication is strictly preferred for  $\gamma > \hat{\gamma}$ .*

This result, and all others in this section, are proved in a final subsection. The result intuitively “re-isolates” our main friction of interest as the *only* source of distortion.

### D.4 Irrational inattention or “two-distortion case”

Let us now consider a situation in which there is a second *competing* distortion induced by irrational inattention or some other “wedge” in first-order beliefs.

A first path forward for evaluating optimal policy is to treat inattention and the behavioral bias as joint sources of inefficiency. This is tantamount to evaluating  $\mu_{\text{in}}$  in (42) with  $q^* = 1$ , or continuing to use the original objective (3). Provided  $\lambda < q$ , the paternalistic planner again uses a threshold strategy:

**Proposition 16.** *Let  $c(\gamma) \equiv \mathbb{I}\{\mathcal{A}^* = \mathcal{A}_Y\}$  be a 0 or 1 indicator for using target communication. For  $\lambda < q \leq q^* = 1$ ,  $c(\gamma)$  weakly increases on the domain  $[0, 1]$ .*

In this case, target communication may be preferred on the entire domain. This has the opposite intuition from the previous result: some over-reaction in GE reasoning helps offset the attenuation from incomplete information.

Note that the previous two propositions do not cover the case of  $q < \lambda < q^* = 1$ , with agents believing they are “worse than average.” In such a case, the considerations of canceling out the friction in higher-order reasoning and “fighting” the wedge in first-order beliefs do not stack with one another. Instead, there is now room for the familiar second-best logic of using one distortion to fight another.

Next, consider the case of  $q^* < 1$  but  $q > q^*$ . There is optimally some inattention, but agents over-perceive the precision of their own signals. As discussed later, the empirical evidence in [Kohlhas and Broer \(2018\)](#) and [Bordalo et al. \(2018\)](#) supports such a case in the data. We now consider a policy problem with the objective [42](#), but with  $\mu_{in}$  evaluated at  $q = q^*$  and  $\lambda = q^*$ . In such a case, we can show that if  $\lambda < q$  our result extends in the following sense:

**Proposition 17.** *Let  $c(\gamma) \equiv \mathbb{I}\{\mathcal{A}^* = \mathcal{A}_Y\}$  be a 0 or 1 indicator for using target communication. For  $\lambda < q \leq 1$  and  $q^* < q$ ,  $c(\gamma)$  weakly increases on the domain  $[0, 1]$ .*

**Empirical (and psychological) context.** The combination of the evidence provided in [Bordalo et al. \(2018\)](#), [Coibion and Gorodnichenko \(2012, 2015\)](#), [Coibion et al. \(2018\)](#), [Kohlhas and Broer \(2018\)](#), and [Kohlhas and Walther \(2018\)](#) from various surveys of macroeconomic forecasts rejects the representative-agent, rational-expectations benchmark. Much of this evidence concentrates on professional forecasters, but some of it covers firms and consumers as well. Notwithstanding the difficulty of extrapolating from such broad-scope evidence to the specific counterfactual studied in our paper, we now explain why this evidence points towards the following combination of parameters, which (per Proposition [17](#)) suffices for our main result to survive even when inattention is irrational:

- $q < 1$ , meaning that people are inattentive or imperfectly informed;
- $q^* < q$ , meaning that people over-estimate the precision of their information relative to the truth; and
- $\lambda < q$ , meaning that people under-estimate the precision of others’ information relative to their own.

The first property is documented in [Coibion and Gorodnichenko \(2012, 2015\)](#) by showing that average forecasts under-react to news. These papers also offer a structural interpretation of this fact in terms of models with dispersed noisy information and rational expectations, along the lines of [Morris and Shin \(2002\)](#) and [Woodford \(2003\)](#). But they do not contain any evidence that would help support this hypothesis against the richer alternative. That is, they *presume*  $q = \lambda = q^* < 1$ , but the provided evidence actually only proves  $q < 1$  and  $\lambda < 1$ , leaving the  $q - \lambda$  gap and the value of  $q^*$  free. Accordingly, [Gabaix \(2019\)](#) interprets the same fact as evidence of a certain form of *irrational* inattention, or in terms of a model where  $q < 1$  but  $q^* = 1$ .

This ambiguity is resolved by the combination of [Bordalo et al. \(2018\)](#) and [Kohlhas and Broer \(2018\)](#). These papers provide evidence that, whereas forecast errors are positively related to past forecast revisions

at the *aggregate* level (as originally shown in [Coibion and Gorodnichenko, 2015](#)), they are negatively related at the *individual* level.

The second fact, by itself, rejects rational expectations: with rational expectations, an individual's forecast error cannot be forecastable by his *own* past information. Furthermore, the sign of the documented bias points towards individual over-reaction to own information. [Kohlhas and Broer \(2018\)](#) attribute such over-reaction to the tendency of an individual to think that his information is more precisely than it actually is ("absolute over-confidence"). In the language of the simple model presented above, this means  $q > q^*$ . [Bordalo et al. \(2018\)](#) propose a variant explanation, based on "representativeness bias," which though works in essentially the same way and, for our purposes, can also be captured by  $q > q^*$ .

To match the first fact, or the under-reaction of the average forecasts, it is then necessary to have dispersed noisy information. To understand why, recall that this fact alone could be explained either by dispersed noisy information, as originally shown by [Coibion and Gorodnichenko \(2012, 2015\)](#) themselves, or by a bias that causes individual beliefs to under-react, as suggested by [Gabaix \(2019\)](#). But we just argued that the bias in individual beliefs, as evidenced in the second fact, is of the opposite kind. The two facts *together* therefore point towards the combination of over-confidence and dispersed noisy information, which in the language of the model presented above means  $q^* < q < 1$ .

Both [Bordalo et al. \(2018\)](#) and [Kohlhas and Broer \(2018\)](#) reach basically the same conclusion. [Angeletos and Huo \(2018\)](#) further clarify why information has to be not only noisy but also dispersed: the aforementioned facts together imply one agent's forecast error is predictable by the *another* agent's information. [Angeletos and Huo \(2018\)](#) also develop the precise mapping between these facts and a model that has a similar formal structure as our framework—and that adds various extra features that are needed for quantitative purposes, including richer micro-foundations, long horizons and learning dynamics, but are of course beyond the scope of our paper.

More importantly for the present purposes, [Kohlhas and Broer \(2018\)](#) provide a *third* fact, which points towards  $\lambda < q$ : individual forecasts over-react to consensus forecasts. This is consistent with the hypothesis that the typical individual under-estimates the information of others and is thus "relatively over-confident" in their own assessment. As mentioned in the main text, such a perception in being "better than average" is described by psychologists in various contexts (see, for instance, [Alicke and Govorun, 2005](#)). In our context, it translates into a lack of confidence in other agents' attentiveness to forward guidance, or  $\lambda < q$ .

Of course, the literature reviewed here may not be the final word on what the best structural interpretation of the available evidence on expectations is. Also, this evidence need not be directly importable to the context of interest. In particular, [Garcia-Schmidt and Woodford \(2019\)](#) and [Farhi and Werning \(2019\)](#) argue that, because this was the first time the United States had hit the ZLB context and nobody could draw from past data to infer the GE effects of the various unconventional policies the Fed had to experiment with, people may have naturally resorted to introspection and deductive (iterative) reasoning, of the kind seen in experiments. If this argument is valid, it offers a more direct justification for our baseline analysis. Still, the evidence discussed above is complementary: not only it rejects the representative-agent, rational-expectations benchmark but also favors, within the extension presented in this appendix, the par-

ticular scenario of  $\lambda < q$  and  $q^* < q$ , which in turn suffices for our main policy prescription to continue to hold (Proposition 17) despite the presence of confounding distortions.

## D.5 Proofs

### Proof of Proposition 15

Note first the following properties of  $(\mu_{\text{in}}, \mu_\tau, \mu_Y)$ , which can be verified by direct calculation:

1.  $\mu_\tau = \mu_{\text{in}}$  when  $\gamma = 0$ , and  $\mu_\tau > \mu_{\text{in}}$  when  $\gamma \in (0, 1]$ .
2.  $\mu_Y = \mu_{\text{in}}$  when  $\gamma = 1$ , and  $\mu_Y < \mu_{\text{in}}$  when  $\gamma \in [0, 1)$ .

Note finally that  $\mu_Y > 0$  if and only if  $\lambda > 1 + q - 1/\alpha$ , which by the same argument provided in Lemma 5 is always true if we have specified  $|\delta_Y| < 1$  for all  $\gamma \in [0, 1]$ . As with the main result, we will focus on such a case in the proof.

Assume that the policymaker's objective function is given by (42), where  $\mu_{\text{in}}$  defines the slope of the implementability constraint in the noisy rational expectations case of a given model (i.e., in which  $\lambda$  is set equal to  $q$ ).

The objective in terms of the message slope  $r$  and the implementability slope  $\mu$  is

$$(1 - \chi)(r - 1)^2 + \chi(r/\mu - 1/\mu_{\text{in}})^2$$

The optimal  $r$  in closed-form, as a function of other parameters, is

$$r(\mu) = \frac{(1 - \chi)\mu^2 + \chi\frac{\mu}{\mu_{\text{in}}}}{(1 - \chi)\mu^2 + \chi}$$

and the new objective function, in terms of  $(\mu, \mu_{\text{in}})$ , is a function  $\ell_0$ :

$$\mathcal{L} = \ell(\mu, \mu_{\text{in}}) \equiv \chi(1 - \chi) \frac{(\mu - \mu_{\text{in}})^2}{\mu_{\text{in}}^2(\mu^2(1 - \chi) + \chi)} \quad (43)$$

Note that the derivative of the loss function  $\mathcal{L}$  with respect to  $\gamma$  comes through two components, and is

$$\frac{\partial \ell}{\partial \gamma} = \frac{\partial \ell}{\partial \mu} \frac{\partial \mu}{\partial \gamma} + \frac{\partial \ell}{\partial \mu_{\text{in}}} \frac{\partial \mu_{\text{in}}}{\partial \gamma} \quad (44)$$

The two partial derivatives of  $\ell$  are

$$\frac{\partial \ell}{\partial \mu} = 2(1 - \chi)(\chi) \frac{(\mu - \mu_{\text{in}})(\mu\mu_{\text{in}}(1 - \chi) + \chi)}{\mu^2(\mu^2(1 - \chi) + \chi)^2}$$

which is positive if and only if  $\mu > \mu_{\text{in}}$ , and

$$\frac{\partial \ell}{\partial \mu_{\text{in}}} = -2(1 - \chi)(\chi) \frac{\mu}{\mu_{\text{in}}} \cdot \frac{(\mu - \mu_{\text{in}})}{\mu_{\text{in}}^2(\mu^2(1 - \chi) + \chi)}$$

which is positive if  $\mu < \mu_{\text{in}}$ .

Plugging the previous expressions into (44), we have that  $\partial\ell/\partial\gamma > 0$  is positive if  $\mu > \mu_{\text{in}}$  and

$$\frac{\partial\mu}{\partial\gamma} > \frac{\mu}{\mu_{\text{in}}} \frac{\mu^2(1-\chi) + \chi}{\mu\mu_{\text{in}}(1-\chi) + \chi} \frac{\partial\mu_{\text{in}}}{\partial\gamma} \quad (45)$$

or if  $\mu < \mu_{\text{in}}$  and

$$\frac{\partial\mu}{\partial\gamma} < \frac{\mu}{\mu_{\text{in}}} \frac{\mu^2(1-\chi) + \chi}{\mu\mu_{\text{in}}(1-\chi) + \chi} \frac{\partial\mu_{\text{in}}}{\partial\gamma} \quad (46)$$

Finally, note that the partial derivative of  $\mu_{\text{in}}$  with respect to  $\gamma$  is

$$\frac{\partial\mu_{\text{in}}}{\partial\gamma} = \frac{\alpha^2(1-q)\lambda}{(1-\alpha+\alpha q(1-q))^2} > 0$$

**Monotonicity of loss with instrument communication.** Note that the derivative of  $\mu_{\tau}$  in  $\gamma$  is given by

$$\frac{\partial\mu_{\tau}}{\partial\gamma} = \frac{q\alpha^2(1-\lambda)}{(1-\alpha+q\alpha(1-\alpha\gamma)-\alpha\gamma\lambda(1-\alpha))^2} > 0$$

Consider first this case  $q > \lambda$  which entails  $\mu_{\tau} > \mu_{\text{in}}$ . A looser version of (45) is

$$\frac{\partial\mu_{\tau}}{\partial\gamma} > \left(\frac{\mu_{\tau}}{\mu_{\text{in}}}\right)^2 \frac{\partial\mu_{\text{in}}}{\partial\gamma}$$

and this can be verified by “brute force”: the previous expression is

$$\frac{(1-\lambda)}{(1-q)} > \frac{(1-\lambda\alpha\gamma)^2}{(1-q\alpha\gamma)^2} \quad (47)$$

Note that an upper bound for the right-hand-side is given for  $\gamma = 1$ , or

$$\frac{(1-\lambda)}{(1-q)} > \frac{(1-\lambda\alpha)^2}{(1-q\alpha)^2}$$

But this is guaranteed if we impose  $\alpha < 1/2$ , which was consistent with  $|\delta_Y| > -1$  on the entire domain of study.

Now consider  $q < \lambda$ . The loose version of (46) is

$$\frac{\partial\mu_{\tau}}{\partial\gamma} < \left(\frac{\mu_{\tau}}{\mu_{\text{in}}}\right)^2 \frac{\partial\mu_{\text{in}}}{\partial\gamma}$$

because for  $\mu_{\tau} < \mu_{\text{in}}$  the right-hand-side is a lower bound. From the exact same math of (47), the key condition is now

$$\frac{(1-\lambda)}{(1-q)} < \frac{(1-\lambda\alpha\gamma)^2}{(1-q\alpha\gamma)^2} \quad (48)$$

which is satisfied for the exact same reason.

Together, these arguments suffice to show that in any case,  $\ell(\mu_{\tau}, \mu_{\text{in}})$  increases in  $\gamma$ . Note finally that this loss function is 0 at  $\gamma = 0$ , where  $\mu_{\tau} = \mu_{\text{in}}$ , and strictly positive at  $\gamma = 1$ , where  $\mu_{\tau} \neq \mu_{\text{in}}$ .

**Monotonicity of loss with target communication.** The derivative of  $\mu_Y$  in  $\gamma$  is given by

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

First consider  $q > \lambda$ , which entails  $\mu_Y < \mu_{\text{in}}$ . It is simple to show that (46) is never satisfied because  $\frac{\partial \mu_Y}{\partial \gamma} > \frac{\partial \mu_{\text{in}}}{\partial \gamma}$ , since

$$\frac{\partial \mu_Y}{\partial \gamma} = \frac{(1-\lambda)}{(1-q)} \frac{(1-\alpha+\alpha q(1-\gamma))^2}{(1-\alpha+\alpha\lambda(1-\gamma))^2} \frac{\partial \mu_{\text{in}}}{\partial \gamma} > \frac{\partial \mu_{\text{in}}}{\partial \gamma}$$

Next consider the case  $q < \lambda$ , which entails  $\mu_Y > \mu_{\text{in}}$ . Note that condition (45) is violated because

$$\frac{\partial \mu_Y}{\partial \gamma} = \frac{(1-\lambda)}{(1-q)} \frac{(1-\alpha+\alpha q(1-\gamma))^2}{(1-\alpha+\alpha\lambda(1-\gamma))^2} \frac{\partial \mu_{\text{in}}}{\partial \gamma} < \frac{\partial \mu_{\text{in}}}{\partial \gamma} < \left(\frac{\mu_Y}{\mu_{\text{in}}}\right)^2 \frac{\partial \mu_{\text{in}}}{\partial \gamma}$$

Together, these arguments suffice to show that  $\ell(\mu_Y, \mu_{\text{in}})$  decreases in  $\gamma$ . Note finally that this loss function is 0 at  $\gamma = 1$ , where  $\mu_Y = \mu_{\text{in}}$ , and strictly positive at  $\gamma = 0$ , where  $\mu_Y \neq \mu_{\text{in}}$ .

**Proving the threshold strategy.** Given the monotonicities established above, proving the sought-after result—that target communication is optimal if and only if  $\gamma > \hat{\gamma}$ , for some  $\hat{\gamma} \in (0, 1)$ —requires only using continuity arguments like in the proof of Theorem 1.

### Proof of Proposition 16

First, we note the monotonicity of  $(\mu_\tau, \mu_Y)$  in  $\gamma$ . The derivative of  $\mu_\tau$  with respect to  $\gamma$  is

$$\frac{\partial \mu_\tau}{\partial \gamma} = \frac{1}{\mu_\tau^2} \frac{\alpha q(1-\lambda)}{(1-\lambda\gamma)^2} > 0$$

and the derivative of  $\mu_Y$  is

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

Next, we want to show that  $\mu_\tau > \mu_Y$ . The correct condition in terms of parameters is

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

Given that  $\mu_Y > 0$ , which is guaranteed like just as in Lemma 5, the left denominator is positive. The other three terms are necessarily positive. Thus an equivalent statement, after cross-multiplying, is the following:

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

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

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

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

Finally, let us return to the proof of optimality. It is straightforward to solve the expression for  $\mu_Y$  for some  $\tilde{\gamma}_Y(\alpha, \lambda, q) \in (0, 1]$  such that  $\mu_Y|_{\gamma=\tilde{\gamma}_Y} = 1$ . One can apply the argument in the proof of Theorem 1 to the loss functions  $\mathcal{L}_\tau(\gamma)$  and  $\mathcal{L}_Y(\gamma)$  on the domain  $[0, \tilde{\gamma}_Y]$ . There is some  $\hat{\gamma} \in [0, \tilde{\gamma}_Y)$  where the functions cross.

For  $\gamma \in (\tilde{\gamma}_Y, 1)$ , we know that (i)  $\mu_Y$  and  $\mu_\tau$  both increase in  $\gamma$  and (ii)  $\mu_\tau > \mu_Y$ . It is straightforward to deduce that  $\mu_\tau > \mu_Y > 1$  for  $\gamma > \tilde{\gamma}_Y$  (and hence  $\mathcal{L}_Y < \mathcal{L}_\tau$ ), which shows the optimality of target communication and completes the proof.

### Proof of Proposition 17

We proceed with the same parameter restriction assumed in the proof of Proposition 15. Note also that the same expressions for the loss functions, the partial derivatives thereof, and sufficient conditions for monotonicity of the loss function in  $\gamma$  still apply.

Applying arguments from the proof of Proposition 15, it is simple also to show that  $\mu_\tau > \mu_Y$  and  $\mu_{\text{in}} > \mu_Y$  on this domain.

**Case 1:**  $q^* \leq \lambda < q$ . In this case,  $\mu_\tau \leq \mu_{\text{in}}$  with equality only for  $q^* = \lambda$  and  $\gamma = 1$ , verified by the direct calculation

$$\frac{1 - \alpha \lambda \gamma}{(1 - \alpha)(1 - \alpha \lambda \gamma) - q \alpha(1 - \alpha \gamma)} \leq \frac{1 - \alpha q^* \gamma}{1 - \alpha - q^* \alpha(1 - \gamma)} \quad (49)$$

In this case, we have  $\mu_Y < \mu_\tau \leq \mu_{\text{in}}$ , and all three increasing in  $\gamma$ . It follows that instrument communication always produces less loss and is preferred on the entire domain  $\gamma \in [0, 1]$ . To see this, note that for  $\mu < \mu_{\text{in}}$ , the loss function is decreasing in  $\mu$ .

**Case 2:**  $\lambda < q^* \leq \frac{q}{1 + \alpha(q - \lambda)} < q$ . Re-arrangement of (49), with this condition, again shows  $\mu_\tau \leq \mu_{\text{in}}$  with equality only at  $\gamma = 1$  and  $q^* = \frac{q}{1 + \alpha(q - \lambda)}$ . Again, instrument communication is preferred on the entire domain.

**Case 3:**  $\frac{q}{1 + \alpha(q - \lambda)} < q^* < q$  and  $\lambda < q^*$ . In this final case, there exists a  $\check{\gamma}$  such that  $\mu_\tau > \mu_{\text{in}}$  for  $\gamma > \check{\gamma}$  and  $\mu_\tau \leq \mu_{\text{in}}$  for  $\gamma \leq \check{\gamma}$ . The previous argument applies to show the optimality for instrument communication for  $\gamma \leq \check{\gamma}$ . For  $\gamma > \check{\gamma}$ , we want to show that the loss for instrument communication strictly increases and the loss from target communication strictly decreases.

From the proof of Proposition (15), a sufficient condition for the first is that

$$\frac{\partial \mu_\tau}{\partial \gamma} > \left( \frac{\mu_\tau}{\mu_{\text{in}}} \right)^2 \frac{\partial \mu_{\text{in}}}{\partial \gamma}$$

This condition simplifies to

$$\frac{q}{q^*} \frac{(1 - \lambda)}{(1 - q^*)} > \frac{(1 - \lambda \alpha \gamma)^2}{(1 - q^* \alpha \gamma)^2}$$

Taking a lower bound on the left (with  $q/q^* \geq 1$ ) and an upper bound on the right (evaluating at  $\gamma = 1$ ) gives

$$\frac{(1-\lambda)}{(1-q^*)} > \frac{(1-\lambda\alpha)^2}{(1-q^*\alpha)^2}$$

which, as used in the proof of Proposition 15, will always hold for  $\lambda < q^*$  and  $\alpha < 1/2$ . Thus we have shown that  $\mathcal{L}_\tau(\gamma)$ , the loss function associated with instrument communication, strictly increases for  $\gamma > \check{\gamma}$ .

Next, a sufficient condition for  $\mathcal{L}_Y(\gamma)$ , the loss function from target communication, to decrease for  $\gamma > \check{\gamma}$  is  $\frac{\partial \mu_Y}{\partial \gamma} > \frac{\partial \mu_{in}}{\partial \gamma}$ . By direct calculation,

$$\frac{\partial \mu_Y}{\partial \gamma} = \frac{q}{q^*} \frac{(1-\lambda)}{(1-q^*)} \frac{(1-\alpha + \alpha q^*(1-\gamma))^2}{(1-\alpha + \alpha \lambda(1-\gamma))^2} \frac{\partial \mu_{in}}{\partial \gamma} > \frac{\partial \mu_{in}}{\partial \gamma}$$

so this is always true.

We have thus established that the difference in loss between target and instrument communication, or  $\Delta \equiv \mathcal{L}_Y(\tau) - \mathcal{L}_\tau(\tau)$ , decreases in  $\gamma$  for  $\gamma > \check{\gamma}$ .

Let the choice of target communication be a 0 or 1 indicator variable,  $c \equiv \mathbb{I}\{\mathcal{A}^* = \mathcal{A}_Y\} = \mathbb{I}\{\Delta < 0\}$ .  $c$  weakly decreases in  $\Delta$ , so the choice of target communication weakly increases in  $\gamma$  for  $\gamma \in (\check{\gamma}, 1]$ . Because  $c = 0$  for any  $\gamma \in [0, \check{\gamma}]$ , this completes the proof that  $c$  weakly increases in  $\gamma$  in  $[0, 1]$ .