Building Trust in Government

The Opportunities and Challenges of Adverse Shocks*

Avidit Acharya[†] Edoardo Grillo[‡] Harry Pei[§]

May 21, 2021

Abstract

How do adverse shocks that align the interests of state and society shape a government's ability to build and maintain the trust of its citizens? We focus on the case where citizens cannot tell whether an adverse shock that warrants government action has hit, unless the government mishandles the crisis. We show that: (i) increasing the frequency of shocks has no impact on the scope of trust if shocks are severe, but reduces the scope of trust if shocks are mild; (ii) increasing the severity of shocks lowers social welfare but increases the extent of trust; and (iii) the opportunity to build a reputation for being trustworthy enhances both the government's payoff and social welfare beyond what is achievable when citizens know that the government is strategic.

Key words: trust, reputation, crises, shocks

JEL Codes: C73, D72, O10, H12

*We are indebted to Steve Callander and Steve Haber for comments and discussions. We are also grateful to seminar participants at UVA. Pei thanks NSF Grant SES-1947021 for financial support.

[†]Associate Professor of Political Science, Stanford University, Encina Hall West Rm. 406, Stanford CA 94305-6044 (email: avidit@stanford.edu).

[‡]Assistant Professor of Economics, Collegio Carlo Alberto, Piazza Arbarello, 8, 10122, Torino, Italy (email: edoardo.grillo@carloalberto.org).

[§]Assistant Professor of Economics, Northwestern University, Office 3391, 2211 Campus Drive, Evanston, IL 60208-2600 (email: harrydp@northwestern.edu).

1 Introduction

Trust in government has taken a hit lately. Data compiled by the Pew Research Center shows that public trust in American government has declined in the last two decades, from roughly half of Americans saying that they "trust the government in Washington to do what is right" to now less than 20%.[1] Trust in government has also declined in other OECD countries.[2] Researchers and commentators have linked these declines to governments' handling of major challenges such as the Iraq War, the 2008 financial crisis, and the refugee crisis in Europe.[3]

Trust in government helps a society achieve common goals and solve common problems like national security threats, economic crises, and other calamities. These adverse shocks often require government action to contain their negative consequences, and governments typically cannot deal with these crises without some cooperation and trust from their citizens. For example, when facing a financial crisis, a government may need to use public funds to bail out industries whose bankruptcy could result in a devastating recession. Or, facing a national security threat, it may need to call on citizens to finance a costly war. Governments around the world have called on their citizens to make significant sacrifices to deal with such problems. Citizens are generally willing to support their government if they believe that the crisis is severe and the government can handle it effectively.

However, in the absence of full information, citizens may be suspicious of their government's true motivations. Before the attack against Iraq in 2003, for example, the American public was suspicious about the severity of the national security threat, specifically the existence of Iraqi weapons of mass destruction. During the subprime financial crisis of 2008, public opinion revealed skepticism as to whether bailing out the banks would be effective in preventing a deep recession, or if politicians were merely doing a favor to Wall Street. During the Covid-19 pandemic, citizens have expressed uncertainty as to whether stimulus payments to families and businesses are not just excuses for government to redistribute to its favored groups.

We study how the frequency and severity of adverse shocks affect the government's ability to build trust when it is privately informed about their occurrence. We embed these shocks into a repeated version of the peasant-dictator game that captures the government's time inconsistency problem described by Kydland and Prescott (1977). This game has been used to study trust in government by Acemoglu (2006), Phelan (2006), Besley and Ghatak (2010), and others.

In each period of our model, society is in one of two possible states: a normal state or a crisis. Citizens decide whether to trust their government, for example by

[1] See Pew Research Center "Public Trust in Government: 1958-2019."

[2] According to the Gallup World Poll public confidence in the British government declined from roughly half of respondents expressing confidence in 2006 to only just over a third in 2020.

[3] See for instance OECD (2013), Carlson et al. (2018) and Eichengreen et al. (2020).

making a costly investment that yields a positive return that could be taxed at a low or high rate. If citizens invest, the government then decides whether to force a sacrifice upon them by setting high taxes. Normal times correspond to the canonical case where citizens prefer low taxes, while the government prefers high taxes. In a crisis, however, both the citizens and the government prefer high taxes: citizens would like the government to be sufficiently resourced to deal with the crisis. The government privately observes whether a crisis has occurred, while citizens do not. Citizens only observe all the past actions and past citizens' payoffs, which means that they can learn that a crisis had occurred only if the government mishandled it.

Building trust in this setting is challenging. On the one hand, crises can expand the scope for cooperation since they align the interests of state and society. On the other hand, since the government needs to set high taxes during a crisis, citizens more frequently suspect that their leaders are acting rapaciously, which limits the government's ability to build trust.

Our first result shows that when crises become more severe, social welfare decreases but the frequency with which citizens trust the government increases. However, the effect of an increase in the frequency of crises depends on their severity. When crises carry severe consequences for the government, increasing their frequency has no impact on the possibility to build trust. When crises have only mild consequences for the government, increasing their frequency makes it harder to build trust.

We also show that when the government is sufficiently forward-looking, the opportunity to build a reputation for being trustworthy increases the highest payoff the government can attain, the extent of trust, and social welfare. This stands in contrast to the case where citizens can observe the state, in which Fudenberg, Kreps, and Maskin (1990)'s result implies that the ability to build a reputation for playing pure actions cannot alter the highest attainable payoff for the patient player.[4] Our methodological contribution is to establish a tight upper bound on the government's payoff based on the rate of reputation building and to use the optional stopping theorem to show that reputation increases trust in government and social welfare.

2 The Model

Time is discrete, indexed by $t = 0, 1, 2, ...$ A long-lived government with discount factor $\delta \in (0, 1)$ interacts with a sequence of short-lived citizens, one in each period.

[4]The only commitment type in our game plays a pure strategy and citizens cannot perfectly monitor the government's stage-game strategy due to the iid state. We are unaware of any existing results that characterize the government's highest equilibrium payoff in such games. We rule out mixed-strategy commitment types since their micro-foundations are debatable: Pei (2021) shows that there is no equilibrium where some rational type with generic payoffs mixes in every period.

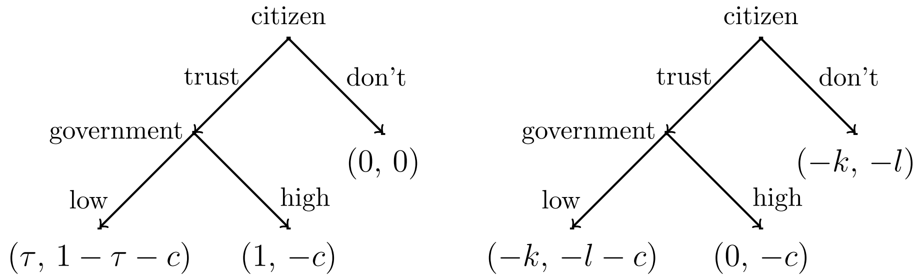

Figure 1: The left panel is for $\theta_t = G$ and the right panel for $\theta_t = B$. The first entry of the payoff vector stands for the government's payoff and the second entry stands for the citizens' payoff.

In each period t, a state $\theta_t \in \{G, B\}$ is realized where G is the normal state and B is the crisis state. We assume that $\{\theta_t\}_{t\in\mathbb{N}}$ are i.i.d. with $p \in (0, 1)$ being the probability that $\theta_t = B$. The citizen who is active in period t decides whether to trust the government, for example by making a costly investment that the government can later tax at a high or a low rate. As in Phelan (2006), the high tax is confiscatory. After the citizen moves, the government chooses the tax rate.

Players' moves and stage-game payoffs are shown in Figure 1. If $\theta_t = G$ and the citizen does not trust, then both players receive a payoff equal to 0. If the citizen invests, then he incurs a cost $c \in (0, 1)$ but makes a return of 1. If the government confiscates, then the citizen gets $-c$ while the government gets 1. If it does not confiscate, then the government gets $\tau \in (0, 1 - c)$ and the citizen gets $1 - \tau - c$.

When $\theta_t = B$, the interests of the government and of the citizens are aligned. If the citizen does not trust, he incurs a loss $l > c$ and the government incurs a loss $k > 0$. If the citizen trusts and the government does not choose high taxes, then the citizen incurs an additional investment cost c while the government's payoff remains $-k$. If the citizen trusts and the government chooses high taxes, then the citizen's payoff is $-c$ and the government's payoff is 0. The interpretation is that the government requires the additional finances to tackle the crisis.

We assume throughout the paper that

$$pl < c < (1 - p)(1 - \tau). \tag{1}$$

Thus, citizens prefer not to invest if they expect the government to choose high taxes in both states and they prefer to invest if they expect high taxes only in state B.

Payoffs and Welfare Measures. Let v_t and u_t respectively be the government's and the citizens' stage-game payoffs in period t. The government maximizes the sum of its discounted stage-game payoffs $(1 - \delta) \sum_{t=0}^{+\infty} \delta^t v_t$. We define social welfare to be the discounted sum of the government's and of the citizens' payoffs:

$$\mathcal{W} = (1-\delta)\sum_{t=0}^{+\infty}\delta^t(v_t + u_t). \tag{2}$$

An implicit assumption is that social welfare is computed according to the government's discount rate δ. We measure the extent of trust by the discounted frequency of investment:

$$\mathcal{I} \equiv \sum_{t=0}^{\infty}(1-\delta)\delta^t \mathbf{1}\{\text{citizen invests in period } t\}. \tag{3}$$

We say that trust is possible in an environment if there exist equilibria with $\mathcal{I} > 0$.

Benchmark with Observable States. Suppose that all players observe all the past actions as well as all the past and current states. This captures situations where negative shocks are observable and there is no uncertainty about the role of the government in offsetting their negative consequences. If

$$\delta \geq \underline{\delta} \equiv \frac{1-\tau}{1+p(k-\tau)},$$

then there exists an equilibrium that attains the government's highest equilibrium payoff $(1-p)\tau$ and the highest social welfare $1-p-c$. On the equilibrium path, citizens trust the government and the government sets high taxes only in state B. If the government ever deviates, future citizens never trust again.

Information Structure. The rest of this article examines the case where the current and past states are privately observed by the government, but not by the citizens. Citizens can infer past states only by observing players' past actions, and the payoffs of past citizens.

According to Figure 1, if the citizen invested in period t but the government confiscated, future citizens cannot infer θ_t. If the citizen invested and the government set low taxes, or if the citizen did not invest, then future citizens can infer θ_t. The interpretation is that citizens may not realize the severity of a crisis if the government handled it effectively, but they will know that a crisis occurred if the government was unprepared. See Li and Zhou (2020) for a similar informational assumption.

Solution Concept. We focus on perfect Bayesian equilibria (PBEs) where the government never chooses low taxes in the bad state. We refer to these as *equilibria.* In every PBE that satisfies this refinement, social welfare is

$$\mathcal{W}(\mathcal{I}) = (1-p-c)\mathcal{I} - (k+l)p(1-\mathcal{I}), \tag{4}$$

which is linear and strictly increasing in $\mathcal{I}$. Under our refinement, maximizing social welfare is thus equivalent to maximizing the frequency of trust. Hence, we only report the frequency of trust $\mathcal{I}$ in our results.

Our refinement is reminiscent of the renegotiation proofness refinement in Ely and Välimäki (2003). It implies that the long-run player will never take any action that reduces both his and his opponent's stage-game payoffs. In our application, it is unrealistic for a government to deliberately mishandle a crisis when it has the resources to deal with it and, given the payoff structure, future citizens will learn that this happened. Appendix A.4 provides conditions under which the optimal equilibrium under this refinement is optimal among all PBEs.

3 Building Trust under Asymmetric Information

Our first result provides a necessary and sufficient condition under which trust is possible when the government is sufficiently patient. Its proof (in the supplemental appendix) uses the standard arguments of Abreu, Pearce, and Stacchetti (1990).

Theorem 1. *Suppose that*

$$\delta \geq \bar{\delta} \equiv \frac{1-\tau}{1+p(k-1)}.$$

1. *If $\tau - p \leq -pk$, then in every equilibrium, the government's payoff is $-pk$, and the frequency of trust $\mathcal{I}$ is* 0.

2. *If $\tau - p > -pk$, then there exists an equilibrium that simultaneously attains the government's highest equilibrium payoff $\tau - p$, and the maximal frequency of trust*
$$\mathcal{I}^0 \equiv \frac{\tau - p + pk}{(1-p)\tau + pk}. \tag{5}$$

Theorem 1 characterizes the government's highest equilibrium payoff and the highest frequency of trust. The highest social welfare is obtained via equation (4). It also shows that these maximums can be attained in the same equilibrium.

Theorem 1 also shows that trust is possible and a patient government cab benefit from repeated interactions if and only if $\tau > p(1-k)$—i.e., when crises have severe enough consequences for the government ($k \geq 1$), or when they are not so severe ($k < 1$) but the frequency of crisis p is low enough relative to the government's payoff from low tax rates τ.

The government's highest equilibrium payoff and the maximal frequency of trust are both decreasing in p. Frequent crises thus lower the scope of trust. Intuitively, the government sets low taxes in the good state only if doing so leads to a higher continuation value. When the government sets high taxes, citizens cannot learn whether a crisis actually hit. Thus, they must punish the government after observing high taxes. When crises are more frequent, the government sets high taxes more frequently, which lowers its continuation value. This limits its incentives to build trust.

When trust is possible, the severity of crises for the government, k, does not affect its highest equilibrium payoff. But the maximum frequency of trust $\mathcal{I}^0$ is increasing in k while the maximum social welfare $\mathcal{W}(\mathcal{I}^0)$ is decreasing in k. For some intuition, as k increases, the government's loss when citizens do not invest goes up. This reduces the government's incentive to set high taxes, which increases the frequency of trust. The higher frequency of trust exactly offsets the negative effects of a larger k on the government's payoff. However, because investment comes at a cost c, the overall effect of an increase in k on social welfare is negative.

4 Building a Reputation for Trustworthiness

We now examine how the opportunity to build a reputation affects the government's payoff, the frequency of trust, and social welfare. We model reputation by assuming that with probability $\mu_0 \in (0,1)$, the government is a benevolent commitment type that sets high taxes if and only if the state is B. With complementary probability, it is a strategic type that maximizes its discounted average payoff $(1-\delta)\sum_{t=0}^{+\infty} \delta^t v_t$.

We focus on the case in which citizens *cannot* observe the current and past states, and we analyze the optimal equilibria for a patient government. This is because of two complementary reasons.

First, if a citizen does not invest, future citizens cannot observe what the government would have done if that citizen had instead invested. Due to this lack-of-identification problem, a small probability of a commitment type *cannot* rule out bad equilibria in which citizens never invest. Hence, the only plausible role of reputation is to increase players' highest attainable payoffs.

Second, when citizens observe the current and past states, reputation cannot improve a patient government's payoff, the frequency of trust, or social welfare.[5] When citizens cannot observe the i.i.d. state and the only commitment type plays a pure strategy, we are unaware of any existing result that characterizes the government's optimal equilibrium payoff. For instance, the results in Fudenberg and Levine (1992) and Gossner (2011) imply that the government's payoff is at least its minmax value,

[5]The argument in Fudenberg, Kreps, and Maskin (1990) implies that for every $\mu_0 \in (0,1)$ and $\varepsilon > 0$, there exists $\underline{\delta}_\varepsilon \in (0,1)$ such that when $\delta > \underline{\delta}_\varepsilon$, the government's payoff cannot exceed $(1-p)\tau + \varepsilon$ and social welfare cannot exceed $1-p-c+\varepsilon$.

$-kp$, and at most its optimal commitment payoff

$$v^* \equiv (1-p)\left[(1-\gamma^*)\tau + \gamma^*\right] = \tau - p + (1-\tau)\{p + (1-p)\gamma^*\}. \tag{6}$$

However, their payoff upper bound is *not tight* in our setting. We show later in equation (11) that the patient government's payoff is bounded below v^* in all equilibria.

4.1 Government's Payoff and Social Welfare

Let

$$\gamma^* \equiv 1 - \frac{c - pl}{(1-p)(1-\tau)}. \tag{7}$$

Assumption (1) implies that γ^* is strictly between 0 and 1. By definition, a citizen is indifferent between investing and not investing when the government sets high taxes for sure in state B and sets high taxes with probability γ^* in state G. Let $\alpha \in (0,1)$ be the solution to

$$\alpha \log \frac{p}{p + (1-p)\gamma^*} = (1-\alpha)\log(1-\gamma^*). \tag{8}$$

As we explain later, α measures the speed at which the government can build its reputation. Finally, let

$$\mathcal{I}^* \equiv \frac{\alpha}{p + (1-p)\gamma^*} + \Big(1 - \frac{\alpha}{p + (1-p)\gamma^*}\Big)\mathcal{I}^0.$$

Since $\mathcal{I}^*$ is a weighted average of 1 and $\mathcal{I}^0$, we have $\mathcal{I}^* > \mathcal{I}^0$ and $\mathcal{W}(\mathcal{I}^*) > \mathcal{W}(\mathcal{I}^0)$.

In the game without reputation, Theorem 1 shows that trust is possible if and only if $\tau - p > -pk$. Theorem 2 below states that under this condition, a small probability of the commitment type leads to a large increase in the government's payoff, in the frequency of trust, and in social welfare.

Theorem 2. *Suppose $\tau - p > -pk$. For every $\varepsilon > 0$, there exists $\delta^* \in (0,1)$ such that when $\delta > \delta^*$,*

1. *There exists no equilibrium in which the government's payoff is more than $\tau - p + \alpha(1-\tau) + \varepsilon$.*
2. *There exists an equilibrium in which the government's payoff is more than $\tau - p + \alpha(1-\tau) - \varepsilon$, the frequency of investment is more than $\mathcal{I}^* - \varepsilon$, and social welfare is more than $\mathcal{W}^* - \varepsilon$.*

Theorem 2 implies that a patient government's highest equilibrium payoff in the reputation game is $\tau - p + \alpha(1-\tau)$. This is strictly greater than its highest equilibrium payoff in the game without reputation $\tau - p$, and is strictly lower than its optimal commitment payoff when it can optimally commit to mixed strategies v^*.

The government's *gain from reputation* is $\alpha(1-\tau)$, which is the difference between its highest equilibrium payoff in the reputation game and its highest equilibrium payoff in the game without reputations. This gain is increasing in the government's gain from high taxes in the good state, $1 - \tau$, and the speed of reputation building, α.

We show in the supplemental appendix that $\partial\alpha/\partial l > 0$ and $\partial\alpha/\partial p > 0$. Thus, the gain from reputation is increasing in the frequency of crises and in their severity for the citizens, while it is independent of k, the severity of the crisis for the government. Intuitively, an increase in the frequency of crisis or the citizens' loss from the crisis increases the frequency with which the government is allowed to set high taxes while still gaining the citizens' trust. The increase in the frequency of high taxes reduces the extent of reputation deterioration when the citizens observe the government setting high taxes, which means that it takes longer for the government to milk and to deplete its good reputation thus increases its benefit from a good reputation.

Since $\mathcal{I}^* > \mathcal{I}^0$, the option to build reputation can improve the frequency of trust and social welfare. Although Theorem 2 characterizes the highest government payoff, it does not guarantee that $\mathcal{I}^*$ is the maximal equilibrium frequency of trust. Characterizing the highest attainable frequency of trust remains an open question.

Role of the Refinement. The existence of a gain from reputation does not rely on the fact that we restrict attention only to PBE in which the government always confiscates in the bad state. In the supplemental appendix, we examine the full set of PBE, and derive a necessary and sufficient condition for the government-optimal equilibrium under this refinement to coincide with the government-optimal PBE without it. Under this condition, Theorem 2 applies to the full set of PBE, and implies that the option to build a reputation can strictly increase the payoff of a patient government, the maximal frequency of trust, and social welfare in the full set of PBE.

4.2 Proof Sketch for Theorem 2

First, we show that the government's payoff cannot exceed $\tau - p + \alpha(1-\tau)$. Second, we construct an equilibrium in which the patient government attains this payoff upper bound. Third, we use the optional stopping theorem to show that the frequency of trust in this equilibrium is higher than in the optimal equilibrium without reputation. The full proof is in the supplemental appendix.

The Government's Payoff Upper Bound. Let the government's *reputation* be the log-likelihood ratio between the commitment type and the strategic type. Let γ_t be the *equilibrium probability* with which the strategic-type government confiscates if the state is good in period t. By Bayes rule, the government's reputation increases by $|\log(1-\gamma_t)|$ if the government sets low taxes and decreases by $|\log \frac{p}{p+(1-p)\gamma_t}|$ if the government confiscates. Define

$$\Lambda(\gamma_t) \equiv \frac{|\log(1-\gamma_t)|}{|\log \frac{p}{p+(1-p)\gamma_t}|}. \tag{9}$$

When $\Lambda(\gamma_t)$ is large, the government can restore its reputation fast after milking it, and can receive a higher payoff. Equation (8) implies that $\Lambda(\gamma^*) = \alpha/(1-\alpha)$. Hence, α measures the speed with which the government can rebuild its reputation when the citizens are indifferent between investing and not investing.

When $\gamma_t \in (0,1)$ for every $t \in \mathbb{N}$, it is optimal for the strategic government to set high taxes for sure in the bad state and to set high taxes with probability $\gamma_t' \in (0,1)$ in the good state. This deviation yields a stage-game payoff equals $\tau - p + (1-\tau)[p+(1-p)\gamma_t']$, which is increasing in γ_t'. The expected change in the government's reputation is

$$\Delta(\gamma_t', \gamma_t) \equiv [p+(1-p)\gamma_t'] \log \frac{p}{p+(1-p)\gamma_t} + (1-p)(1-\gamma_t') \log \frac{1}{1-\gamma_t}. \tag{10}$$

If $\Delta(\gamma_t', \gamma_t) < 0$ for every $t \in \mathbb{N}$, the government's reputation decreases in expectation when it plays according to $\{\gamma_t'\}_{t' \in \mathbb{N}}$. Hence, for every $\eta > 0$, there exists $T \in \mathbb{N}$ such that for every $t \geq T$, period t citizen will not invest unless $\gamma_t' \leq \gamma^* + \eta$.

Now let $\hat{\gamma}$ be such that $\Delta(\hat{\gamma}, \gamma^*) = 0$. By the definition of a best reply, an upper bound for the equilibrium payoff of the strategic government is

$$\begin{aligned}
\overline{v} &\equiv \sup_{\gamma_t \leq \gamma^*} \inf_{\{\gamma_t' \text{ s.t. } \Delta(\gamma_t', \gamma_t) < 0\}} \Big\{ \tau - p + (1-\tau)[p+(1-p)\gamma_t'] \Big\} \\
&= \tau - p + (1-\tau)\{p + (1-p)\hat{\gamma}\} \\
&= \tau - p + \alpha(1-\tau),
\end{aligned}$$

where the last equality follows from (8) and (10). In order to show that $\overline{v} < v^*$, let $\alpha^* \equiv p + (1-p)\gamma^* > p$, and note that the function $f(p) = \alpha^* \log \frac{p}{\alpha^*} + (1-\alpha^*) \log \frac{1-p}{1-\alpha^*}$ is maximized at $p = \alpha^*$ and its maximum is 0. Therefore,

$$[p+(1-p)\gamma^*] \log \frac{p}{p+(1-p)\gamma^*} + (1-p)(1-\gamma^*) \log \frac{1-p}{(1-\gamma^*)(1-p)} < 0. \tag{11}$$

Since $\Delta(\hat{\gamma}, \gamma^*) = 0$, equation (8) implies that $\hat{\gamma} < \gamma^*$. Hence, $\overline{v} < v^*$.

Attaining the Payoff Upper Bound. We now construct an equilibrium where a patient government's payoff is arbitrarily close to $\overline{v}$. Let V_t be the government's continuation value in period t. Define $V_0 \equiv \tau - p + \alpha(1-\tau) - \varepsilon$.

Play starts in a *learning phase* and it remains in this phase as long as the government's continuation value is above $\tau - p$. In the learning phase, citizens always invest. If $V_t \in [\tau - p, (1-\delta)(\tau - p) + \delta(1-p))$, the government confiscates in the bad state and confiscates with probability γ^* in the good state. Let $V_{t+1}(H)$ be the government's continuation value if it confiscated in period t, and $V_{t+1}(L)$ its continuation value if it set low taxes. These are defined recursively by:

$$\begin{aligned} V_t &= (1-\delta)(\tau - p) + \delta V_{t+1}(L) \\ V_t &= (1-\delta)(1-p) + \delta V_{t+1}(H). \end{aligned}$$

If $V_t \geq (1-\delta)(\tau - p) + \delta(1-p)$, the government confiscates in both states. Its continuation value is defined by $V_t = (1-\delta)(1-p) + \delta V_{t+1}(H)$. In this case, if it sets low taxes, it is believed to be trustworthy with probability 1 and its continuation value is $V_{t+1}(L) = 1 - p$.

In period t, play belongs to an *absorbing phase* if $V_s \leq \tau - p$ for some $s \leq t$. If $V_t \in [-(1-\delta)kp + \delta(\tau - p), \tau - p]$, then the citizen invests and the government confiscates in the bad state and sets low taxes in the good state. If $V_t < -(1-\delta)kp + \delta(\tau - p)$, then the citizen does not invest. Let $V_{t+1}(N)$ be the government's continuation value when the citizen did not invest in period t, the government's continuation value satisfies:

$$\begin{aligned} V_t &= (1-\delta)(-kp) + \delta V_{t+1}(N) \\ V_t &= (1-\delta)(\tau - p) + \delta V_{t+1}(L) \\ V_t &= (1-\delta)(1-p) + \delta V_{t+1}(H), \end{aligned}$$

By construction, once the government's continuation value falls below $\tau - p$, it never rises above this threshold thereafter.

The government's incentive constraints are implied by the construction of its continuation values. The citizen's incentive constraints in the absorbing phase as well as in the learning phase when $V_t < (1-\delta)(\tau - p) + \delta(1-p)$ are automatically satisfied. Finally, the citizen's incentive constraints when this inequality does not hold is also satisfied. If $V_t \geq (1-\delta)(\tau - p) + \delta(1-p)$, then the log of the likelihood ratio between the commitment type and the strategic type is at least $(1-\gamma^*)/\gamma^*$, which follows directly from Lemma A.2 in Pei (2021).

Improvement in the Frequency of Trust. In the equilibrium constructed above, the government's continuation payoff is close to $\tau - p$ when play first reaches the absorbing phase. Hence, the frequency of trust is close to $\mathcal{I}^0$.

We now provide a lower bound on the expected (discounted) number of periods after which the play reaches the absorbing phase. In the learning phase, the strategic type government confiscates with probability at least $\alpha^* = p + (1-p)\alpha^*$. The equilibrium constructed above implies that the government's continuation value in period t is lower than $\tau - p$ if and only if

$$\sum_{s=0}^{t-1}(1-\delta)\delta^s x_s + \delta^t(\tau - p) \geq \alpha(1-\tau) + \tau - p, \tag{12}$$

where $\{x_t\}_{t\in\mathbb{N}}$ is a sequence of i.i.d. random variables that equal $\tau - p$ with probability $1-\alpha^*$ and $1-p$ with probability α^*. The random variable $y_t \equiv x_t - (\tau - p) - \alpha^*(1-\tau)$ has zero mean. Let $Y_t \equiv \sum_{s=0}^{t}(1-\delta)\delta^s y_s$ for every t. The process $\{Y_t\}_{t\in\mathbb{N}}$ is a bounded martingale. Define T as the first time at which inequality (12) applies. By construction, $Y_T \geq \alpha(1-\tau) - (1-\delta^T)\alpha^*(1-\tau)$, so T is almost surely bounded. The optional stopping theorem implies that:

$$0 = \mathbb{E}[Y_0] = \mathbb{E}[Y_T] \geq \mathbb{E}\Big[\alpha(1-\tau) - (1-\delta^T)\alpha^*(1-\tau)\Big],$$

or equivalently $\mathbb{E}[1-\delta^T] \geq \alpha/\alpha^*$. By construction, citizens invest in every period before T, so the discounted frequency of trust is at least

$$\mathbb{E}[1-\delta^T] + \Big(1 - \mathbb{E}[1-\delta^T]\Big)\mathcal{I}^0 \geq \frac{\alpha}{\alpha^*} + \Big(1 - \frac{\alpha}{\alpha^*}\Big)\mathcal{I}^0 = \mathcal{I}^*.$$

Equation (4) implies that social welfare is at least $\mathcal{W}(\mathcal{I}^*)$, which is strictly greater than $\mathcal{W}(\mathcal{I}^0)$.

5 Extensions

5.1 Asymmetric Information with Observable Shocks

We have assumed that citizens cannot observe the occurrence of a crisis. Though some shocks like natural disasters or pandemics are in fact observable, the government's action may be ineffective in mitigating the negative spillover consequences of the calamity, and citizens usually face uncertainty about this effectiveness.

Consider the following extension. There are three states: a normal state G (with probability $1-p$) in which payoffs are the same as in the left panel of Figure 1, a mild crisis M (with probability $p(1-r)$) in which government intervention is not necessary, and a severe crisis B (with probability pr) in which a government intervention helps contain the effects of the crisis. In state M, payoffs are lower than they are in state G, by constants κ_C for the citizen and κ_G for the government. In state B, the costs

κ_C and κ_G are scaled up by a factor $\phi > 1$, unless the government chooses high taxes, in which case no additional costs beyond κ_C and κ_G are incurred. Citizens know if the state is G, but they cannot directly observe whether the state is B or M.

Although having three states—among which one is observable—changes the characterization of the optimal punishment strategy, our techniques and results extend: the possibility to build trust and its impact on players' payoffs depends on the frequency of the severe bad state, pr, and on the severity of the crisis in this state for the government, as captured by the multiplier, ϕ.

5.2 Government Competence and Turnover

We have assumed the government is always effective at handling crises. Yet, governments differ in their competence— their ability to deal with adverse shocks.

Our model can account for this through the following extension.[6] Suppose the government is a competent type with probability $\rho \in (0, 1)$ and an incompetent type with complementary probability. If a competent government is in office, players' payoffs are given by Figure 1. An incompetent government is unable to effectively handle a crisis: when the state is B, the citizen incurs an additional cost equal to $\kappa > 0$ no matter what the government and citizen choose that period.

A citizen can remove the government from office at the beginning of each period. If the citizen does this, a new government enters office and it is competent with probability ρ. As in other political career concerns models (Persson and Tabellini, 2000), the government and the citizens initially do not know the government's type, and they learn about it as soon as a crisis hits. Thus, following the first crisis, a government is replaced if it is incompetent and is retained if it is competent.

In the long run, a competent government will come into power almost surely, so the main insights of our theorems extend.

6 Summary of Contribution

Our paper contributes to the literature on reputation formation by characterizing tight bounds on the patient player's payoff in games with imperfect monitoring in which the commitment type plays pure strategies.[7] This stands in contrast to the bounds in Fudenberg and Levine (1992) and Gossner (2011) that are tight only when there exists a commitment type that plays the optimal mixed commitment action.

[6] This extension is also equivalent to one in which the citizens in every period decides whether to trust the government or replace it (with replacement being irreversible) and players have access to a public randomization device in the beginning of every period.

[7] The imperfect monitoring in our model comes from the iid state that citizens cannot directly observe. Bar-Isaac and Deb (2021) examine a related model where the long-run player's action are not observed in some periods.

Ekmekci (2011) examines a reputation model with a rating system that maps histories to ratings and each short-run player observes the current-period rating but not previous ratings. He shows that there exists a rating system under which the patient player receives his optimal commitment payoff in all equilibria. In our model, the citizens observe the entire history of actions and all previous citizens' payoffs. In this case, we show that the patient government's payoff is bounded below by his optimal commitment payoff in all equilibria.

Insofar we examine the *discounted* sum of the government and of the citizens' payoffs, our paper differs from Cripps, Mailath, and Samuelson (2004) that focuses on asymptotic outcomes, and Pei (2021) that considers the payoff of the patient player only. We are unaware of existing works on dynamic games that use the optimal stopping theorem to characterize players' payoffs and social welfare.

Finally, our paper contributes to the literature on trust in government and on its importance for economic development (see Algan and Cahuc, 2010, 2014, Tabellini, 2010, and the references therein.). In addition to the papers mentioned in the introduction, Stevenson and Wolfers (2011) attribute the declining trust in American government to the legacy of the Great Recession. Papaioannou (2013), Ehrmann et al. (2013), and Guiso et al. (2016) similarly document the relationship between financial downturns and trust in, respectively, European national governments, the ECB, and EU institutions broadly. Weng et al. (2015) argue that a lack of trust makes citizens less supportive of their governments by documenting a relationship between government trust in Hong Kong and citizens' willingness to support earthquake disaster relief. These contributions provide evidence both that crises affect trust, and trust in turn affects a society's ability to deal with crises. Our paper provides a theoretical understanding of these findings.

References

Abreu, D., D. Pearce, and E. Stacchetti (1990): "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, 58, 1041–1063.

Acemoglu, D. (2006): "Modeling inefficient institutions," Tech. rep., National Bureau of Economic Research.

Algan, Y. and P. Cahuc (2010): "Inherited Trust and Growth," *American Economic Review*, 100, 2060–92.

——— (2014): "Trust, growth, and well-being: New evidence and policy implications," in *Handbook of economic growth*, Elsevier, vol. 2, 49–120.

Bar-Isaac, H. and J. Deb (2021): "Reputation with opportunities for coasting," *Journal of the European Economic Association*, 19, 200–236.

Besley, T. and M. Ghatak (2010): "Property rights and economic development," in *Handbook of Development Economics*, Elsevier, vol. 5, 4525–4595.

Carlson, M., L. Jakli, and K. Linos (2018): "Rumors and Refugees: How Government-Created Information Vacuums Undermine Effective Crisis Management," *International Studies Quarterly*, 62, 671–685.

Cripps, M. W., G. J. Mailath, and L. Samuelson (2004): "Imperfect monitoring and impermanent reputations," *Econometrica*, 72, 407–432.

Ehrmann, M., M. Soudan, and L. Stracca (2013): "Explaining European Union citizens? Trust in the European Central Bank in normal and crisis times," *The Scandinavian Journal of Economics*, 115, 781–807.

Eichengreen, B., S. Orkun, and C. G. Aksoy (2020): "The Political Scar of Epidemics," .

Ekmekci, M. (2011): "Sustainable reputations with rating systems," *Journal of Economic Theory*, 146, 479–503.

Ely, J. C. and J. Välimäki (2003): "Bad reputation," *Quarterly Journal of Economics*, 118, 785–814.

Fudenberg, D., D. M. Kreps, and E. S. Maskin (1990): "Repeated games with long-run and short-run players," *Review of Economic Studies*, 57, 555–573.

Fudenberg, D. and D. K. Levine (1992): "Maintaining a Reputation When Strategies Are Imperfectly Observed," *Review of Economic Studies*, 59, 561–579.

GOSSNER, O. (2011): "Simple Bounds on the Value of a Reputation," *Econometrica*, 79, 1627–1641.

GUISO, L., P. SAPIENZA, AND L. ZINGALES (2016): "Monnet's error?" *Economic Policy*, 31, 247–297.

KYDLAND, F. E. AND E. C. PRESCOTT (1977): "Rules rather than discretion: The inconsistency of optimal plans," *Journal of Political Economy*, 85, 473–491.

LI, F. AND J. ZHOU (2020): "A Model of Crisis Management," *Cowles Foundation Discussion Paper, No 2266*.

OECD (2013): *Government at a Glance*, Organization For Economic Cooperation and Development.

PAPAIOANNOU, E. (2013): "Trust (ing) in Europe?: How Increased Social Capital Can Contribute to Economic Development," .

PEI, H. (2021): "Trust and Betrayals: Reputational Payoffs and Behaviors without Commitment," *Theoretical Economics*, 16, 449–475.

PERSSON, T. AND G. TABELLINI (2000): *Political Economics : explaining Economic Policy*, Cambridge (Mass.) London: The MIT Press.

PHELAN, C. (2006): "Public trust and government betrayal," *Journal of Economic Theory*, 130, 27–43.

STEVENSON, B. AND J. WOLFERS (2011): "Trust in public institutions over the business cycle," *American Economic Review*, 101, 281–87.

TABELLINI, G. (2010): "Culture and Institutions: Economic Development in the Regions of Europe," *Journal of the European Economic Association*, 8, 677–716.

WENG, W., C. K. WOO, Y. CHENG, T. HO, AND I. HOROWITZ (2015): "Public trust and corruption perception: disaster relief," *Applied Economics*, 47, 4967–4981.

A Appendix

A.1 Details for the case with observable states

Consider the following two-state automaton:

1. Play starts in the *reward phase*, in which the citizen invests, and the government sets low taxes in the good state and set high taxes in the bad state. Play remains in the reward phase in the next period if no deviation occurs. Otherwise, play goes to the punishment phase with probability 1.

2. The *punishment phase* is absorbing and it involves the citizen not investing and the government setting high taxes in both states.

Given the assumption in (1), citizen behavior is optimal in both phases. For the government, the discounted average sum of payoffs in the reward phase is $(1-p)\tau$, while in the punishment phase it is $-pk$. Obviously, the government has no incentive to deviate in the bad state. Instead, in the good state, the government sets low taxes if $(1-\delta)\tau + \delta(1-p)\tau \geq (1-\delta) - \delta pk$, which yields the threshold $\underline{\delta}$. The citizen's discounted average sum of payoffs is $(1-p)(1-\tau)-c$. If $\delta[\tau+p(k-\tau)] < (1-\delta)(1-\tau)$, it is not possible to incentivize the government to set low taxes in the good state. Hence, in equilibrium, the citizen's benefit from investing is bounded from above by $-c$, while the payoff from not investing is $-pl$. The assumption in (1) implies that in this case the citizen prefers not to invest. As a result, the payoffs of the citizen and of the government are respectively $-pl$ and $-pk$.

A.2 Proof of Theorem 1

The proof is in three steps. First, we derive upper bounds on the government's equilibrium payoff. Second, we construct equilibria that attain these bounds if δ is sufficiently large. Third, we show that total welfare is maximized in an equilibrium that maximizes the government's payoff.

Payoff Bounds. In equilibrium, the government's payoff cannot exceed $\tau - p$. Let $\overline{v}$ be the supremum of the government's payoff in the set of equilibria (which is well defined because stage-game payoffs are bounded and the discount factor is lower than 1). Take an increasing sequence $(v_n)_{n\geq 1}$ of government's equilibrium payoffs converging to $\overline{v}$. If $\overline{v} > -pk$, we can find an n^* such that for any $n > n^*$, the citizen must invest with strictly positive probability at any history where government's continuation payoff is v_n. If this were not the case, the government's stage-game payoff would be $-pk$ and $v_n = -(1-\delta)pk + \delta v_0$, where v_0 is the government continuation equilibrium payoff following non-investment by the citizen. Because, $(v_n)_{n\geq 1} \to \overline{v}$, there exists n^* such that for any $n > n^*$, $v_n = -(1-\delta)pk + \delta v_0 < \overline{v} < v_0$. This establishes a contradiction as the the continuation value would exceed $\overline{v}$.

Because citizens invest with positive probability, the government must set low taxes with positive probability in the good state (following the assumption in (1)). Pick any equilibrium payoff v for the government. By Abreu et al. (1990), there exists $q \in (0, 1]$ (a probability with which citizens do not invest) such that:

$$v = (1-q)(1-p)\left[(1-\delta)\tau + \delta v_{L,G}\right] + (1-q)p\delta v_H + q\left[-pk + \delta v_0\right],$$

where $v_{L,G}$ is the government's continuation value after low taxes in the good state and v_H is its continuation payoff when it sets high taxes.[8] The government sets low taxes with positive probability in the good state, hence: $(\tau - 1)(1-\delta) + \delta v_{L,G} \geq \delta v_H$. Plugging this inequality in the decomposition of v and using the definition of $\overline{v}$, we get:

$$\begin{aligned} v \leq (1-q)(1-p)\left[(1-\delta)\tau + \delta\overline{v}\right] + \\ + (1-q)p\left[(1-\delta)(\tau-1) + \delta\overline{v}\right] + q\left[-(1-\delta)pk + \delta\overline{v}\right] \end{aligned}$$

Because this inequality must hold for any v, we obtain $\overline{v} \leq (1-q)(\tau - p) - qpk$. Therefore, $\max\{\tau - p, -pk\}$ is an upper bound on the government's equilibrium payoff.

Equilibria. If $\tau - p \leq -pk$, it is immediate to construct an equilibrium in which the discounted average government's payoff is $-pk$: at any information set, the citizen never invests and the government always sets high taxes regardless of the state. Henceforth, we will thus focus on the $\tau - p > -pk$ case. Consider the following two-state automaton equilibrium:

1. Play starts in a *reward phase*: the citizen invests and the government sets low taxes in the good state and high taxes in the bad state. Play remains in the reward phase if the government sets low taxes and the state is good. Play transitions to the *punishment phase* (described below) with probability:

$$\frac{1-\delta}{\delta}\frac{1-\tau}{\tau - p + pk},$$

   if either the government sets high taxes or the government sets low taxes in the bad state. In this phase, the government's continuation value is equal to $\tau - p$.

[8] The continuation values following high taxation do not depend on the state because the citizens cannot identify the state.

2. In the *punishment phase*, the citizen does not invest and the government sets high taxes regardless of the state. This phase is absorbing and the government's payoff in this phase is equal to $-pk$.

Under the transition probability stated above, during the reward phase the government is indifferent between setting low or high taxes in the good state, and strictly prefers to set high taxes in the bad state. Furthermore, the government's discounted average equilibrium payoff is $\tau - p$. Finally, the transition probability is well defined if $\delta > (1-\tau)/[1-p(1-k)]$.

Joint maximization. If $\tau - p \leq -pk$, the government's payoff is $-pk$. Citizens never invest and the discounted average social welfare is $-p(k+l)$. Suppose $\tau - p > -pk$. In this case, a two state automaton equilibrium similar to the one described above can support any government's expected payoff $v \in [-pk, \overline{v}]$ by adjusting the transition probability to the punishment phase accordingly. If the government's equilibrium payoff is $v \in [-pk, \overline{v}]$, then discounted average number of investments $\mathcal{I}^0(v) \geq (v+pk)/[(1-p)\tau + pk]$. Indeed, under the two-stage automaton that attains continuation payoff $\tau - p$ for the government, we have:

$$\mathbb{E}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t \mathbf{1}_{\{a_t=1\}}\Big] = \frac{\tau - p + pk}{\tau - p + pk + p(1-\tau)} = \frac{\tau + p(k-1)}{\tau + p(k-\tau)} \in (0,1).$$

Next, we show that for every $v \in [-pk, \overline{v}]$, $\mathcal{I}^0(v) \leq (\overline{v} + pk)/[(1-p)\tau + pk]$. The equilibrium that attains v has the government setting a low tax in the good state and a high tax in the bad state. Hence, $v = \mathcal{I}^0(v)(1-p)\tau + (1-\mathcal{I}^0(v))(-pk)$. Since $v \leq \overline{v}$, we have $\overline{v} + pk \geq \mathcal{I}^0(v)[(1-p)\tau + pk]$, which yields the desired inequality. Finally, at each period in which the citizen invests, the citizen-maximal stage payoff is $(1-p)(1-\tau) - c$, and the maximal normalized social welfare is $(1-p) - c$. Instead, at each period in which the citizen does not invest, the citizen's payoff is $-pl$, while welfare in that period is $-p(l+k)$. Hence, the citizens' discounted average payoff is bounded above by $\mathcal{I}^0(\overline{v})[(1-p)(1-\tau) - c] - [1 - \mathcal{I}^0(\overline{v})]pl$ and the discounted average social welfare is bounded above by $\mathcal{I}^0(\overline{v})[(1-p) - c] - [1 - \mathcal{I}^0(\overline{v})]p(l+k)$. It is easy to verify that these values are attained in the two-stage automaton constructed above.

A.3 Comparative Statics of α

Substituting γ^* from (7) into (8) and rearranging, we have

$$\frac{\alpha}{1-\alpha} = \frac{\log(1-\gamma^*)}{\log(\frac{p}{p+(1-p)\gamma^*})} = \frac{\log\left(\frac{c-pl}{(1-p)(1-\tau)}\right)}{\log\left(\frac{p(1-\tau)}{1-\tau-c+pl}\right)} \tag{13}$$

We then note that for any parameter $\theta \in \{p, l, ...\}$,

$$\frac{\partial}{\partial \theta}\left(\frac{\alpha}{1-\alpha}\right) = \frac{\partial \alpha / \partial \theta}{(1-\alpha)^2}$$

so that $\partial \alpha / \partial \theta$ has the same sign as the partial derivative of the right side of (13) with respect to θ. The derivative of this ratio of logs with respect to p is

$$\frac{-\frac{(1-c-\tau)\log\left(\frac{c-pl}{(1-p)(1-\tau)}\right)}{p(1-\tau-c+pl)} - \frac{(l-c)\log\left(\frac{p(1-\tau)}{1-\tau-c+pl}\right)}{(1-p)(c-pl)}}{\left[\log\left(\frac{p(1-\tau)}{1-\tau-c+pl}\right)\right]^2}$$

which is positive, because $c - pl > 0$ by assumption and $1 - c - \tau + pl > 1 - c - \tau > 0$ as well. These parametric assumptions (along with the assumption that $l > c$) imply that the numerator is positive because the arguments of both the logs are less than 1. Therefore, we have $\partial \alpha / \partial p > 0$, unambiguously.

Next, the partial derivative of the right side of (13) with respect to l is

$$\frac{p\left(\frac{\log\left(\frac{c-pl}{(1-p)(1-\tau)}\right)}{1-\tau-c+lp} - \frac{\log\left(\frac{p(1-\tau)}{1-c-\tau+pl}\right)}{c-pl}\right)}{\left[\log\left(\frac{p(1-\tau)}{1-\tau-c+pl}\right)\right]^2}$$

Let Q denote the term in big parentheses in the numerator. Then note that the partial derivative of Q with respect to l is

$$-\frac{p\log\left(\frac{c-pl}{(1-p)(1-\tau)}\right)}{(1-c-\tau+pl)^2} - \frac{p\log\left(\frac{p(1-\tau)}{1-c-\tau+pl}\right)}{(c-pl)^2}$$

which is strictly positive. Therefore, Q is minimized at $l = 0$. (Note that even though we have assumed $c < l$, it is still true that Q is bounded below by the value it takes when we set $l = 0$. This bound is not tight in our parameter space, given the assumption that $c < l$; but it is a bound nonetheless.) When $l = 0$, Q takes value

$$-\frac{\log\left(\frac{c}{(1-p)(1-\tau)}\right)}{1-\tau-c} - \frac{\log\left(\frac{p(1-\tau)}{1-\tau-c}\right)}{c} \tag{14}$$

The derivative of this expression with respect to p is

$$-\frac{1}{cp} + \frac{1}{(p-1)(c+\tau-1)} \tag{15}$$

and its second derivative with respect to p is

$$\frac{1}{(p-1)^2(1-c-\tau)}+\frac{1}{cp^2}<0$$

Therefore, the value of p that minimizes (14) is the one that makes (15) equal 0 (so long as this value is contained in the interval $[0,1]$) which is

$$p=\frac{1-\tau-c}{1-\tau}$$

If we substitute this value of p into (14), then the expression in (14) reduces to 0. This means that 0 is a strict lower bound for Q. Therefore, $\partial\alpha/\partial l>0$, unambiguously.

A.4 The Full Set of PBE under Asymmetric Information

We first establish a tight upper bound for the government's equilibrium payoff. We then construct an equilibrium that achieves this bound. Finally, we show that if the following inequality fails, then the government's maximal equilibrium payoff and maximal social welfare exceed their maximum under the renegotiation proofness condition (provided δ is high enough), but when it holds then they correspond to the ones reported in Theorem 1:

$$\frac{l}{k}\geq\frac{1-\tau-c}{\tau}-\frac{p}{1-p}\frac{c}{\tau}. \tag{16}$$

Under this inequality, the results in Theorem 1 and Theorem 2 extend to all Perfect Bayesian equilibria. When inequality (16) does not hold, we compute the government's highest Perfect Bayesian equilibrium payoff as well as the highest level of trust and social welfare.

Payoff bounds. Following a similar argument to that of the first step in the proof of Theorem 1, we can show that if $\tilde{v}:=\max\{(1-p)\tau-pk,\tau-p\}>-pk$ then there is $\tilde{\delta}\in(0,1)$ such that for every $\delta>\tilde{\delta}$, the set of time averaged PBE payoffs for the government is $[-pk,\tilde{v}]$; otherwise, its unique PBE payoff is $-pk$. Since the arguments are nearly identical, we omit it here.

Equilibrium. If $\delta>\tilde{\delta}$ we can construct a PBE in which the government's payoff is $\tilde{v}$ via the following three-state automaton

1. Play starts in a *reward phase*. In this phase, the citizen invests and the government sets low taxes in the good state and high taxes in the bad. Play remains in the reward phase in the next period if the government sets low taxes and the

state is good. Play transits to the *punishment phase* (described below) with probability:

$$\frac{1-\delta}{\delta}\frac{1-\tau}{\tilde{v}+pk-(1-p)\tau},$$

if either the government sets high taxes, or sets low taxes in the bad state. The government's continuation value in this phase is $\tilde{v}$.

2. In the *punishment phase*, the citizen invests and the government sets low taxes in both states. Play remains in the punishment phase if the citizen invests and the government sets low taxes in either state. Otherwise, play moves to a *severe punishment phase* (described below). The government's continuation value in the punishment phase is $(1-p)\tau - pk$.

3. In the *severe punishment phase*, the citizen does not invest and the government sets high taxes regardless of the state. This phase is absorbing and the government's payoff in this phase is $-pk$.

Under the transition probability stated above, in the reward phase the government is indifferent between low and high taxes in the good state, and strictly prefers to set low taxes in the bad state. Furthermore, the government's discounted average payoff in the reward phase is $\tilde{v}$.

Inequality (16)**.** Note that the citizens' discounted average payoff is a convex combination of citizen payoffs in the five possible stage-game outcomes. These are: the citizen does not invest (N), the citizen invests and the government sets high taxes in both the good and bad states (HH), sets high taxes in the good but low taxes in the bad state (HL), sets low taxes in the good but high taxes in the bad state (LH), and sets low taxes in both states (LL). Let $y \in \{0, HH, HL, LH, LL\}$ denote a generic outcome.

Since the government's payoff cannot exceed $\tilde{v}$, the citizens' payoff is bounded from above by the following linear optimization problem:

$$\max_{\pi\in\Delta\{N,HH,HL,LH,LL\}} \sum \pi(y)u(y) \text{ subject to } \sum \pi(y)v(y) \leq \tilde{v}. \tag{17}$$

where $u(y)$ denotes the citizen's payoff and $v(y)$ the government's. When inequality (16) holds, the value of this linear program is given by

$$(1-p)(1-\tau-c) + p(-l-c) + l\frac{\tilde{v}-p\tau}{k},$$

and when inequality (16) fails, the value of this linear program is given by

$$\mathcal{I}^0(1-p-c) + (1-\mathcal{I}^0)p(-k-l) - \tilde{v}.$$

Given that these values are attainable in the equilibria that we have constructed above (both here and in the proof of Theorem 1), these values imply the maximal equilibrium social welfare and maximal equilibrium government payoff vary according to inequality (16), as asserted in the main text.

A.5 Complete Proof of Theorem 2

In the main text we described the construction of equilibria that approximately attains the highest equilibrium payoff when δ is close to 1 and we showed that the frequency of trust and the social welfare are strictly greater in those equilibria compared to the socially optimal equilibrium in the no-reputation benchmark. In what follows, we formalize the proof of the payoff upper bound, namely, the patient government's payoff cannot exceed $\tau - p + \alpha(1-\tau)$ in any equilibrium when $\delta \to 1$.

Suppose by way of contradiction that there exists $\overline{\alpha} \in (\alpha, 1)$ such that for every $\underline{\delta} \in (0,1)$, there exist $\delta > \underline{\delta}$ and an equilibrium where the government's payoff is at least $\tau - p + \overline{\alpha}(1-\tau)$.

Let V_t be the government's continuation value in period t. When the citizens put probability less than $1-\gamma^*$ on the commitment type, they invest only if the government chooses low taxes with positive probability in state G. Therefore, when the citizens do not invest or when the citizens invest and the government sets low taxes, the government's continuation value in period $t+1$ is at least

$$V_{t+1} \geq \frac{1}{\delta}(V_t - (1-\delta)(\tau - p)).$$

Let $\widehat{\alpha} \equiv \frac{1}{2}(\alpha + \overline{\alpha})$ and $\xi \in (0,1)$ be small enough such that

$$\frac{(1-2\xi)(1-\widehat{\alpha}-\varepsilon)}{(1-2\xi)(1-\widehat{\alpha}-\varepsilon)+(\widehat{\alpha}+\varepsilon)} \geq 1 - \frac{\widehat{\alpha}+\overline{\alpha}}{2} \tag{18}$$

When δ is close to 1, (18) implies that there exists $N \in \mathbb{N}$ such that $1-\delta^N \geq \xi$ and for every $\{y_i\}_{i=1}^N \in \{0,1\}^N$ that satisfies $\sum_{i=1}^N y_i \geq N(1-\widehat{\alpha}-\varepsilon)$, we have

$$\sum_{i=1}^N \delta^i y_i \geq \left(1 - \frac{\widehat{\alpha}+\overline{\alpha}}{2}\right) \sum_{i=1}^N \delta^i. \tag{19}$$

Let $M \in \mathbb{N}$ be large enough such that

$$\begin{aligned}\left(1-(1-\xi)^M\right)\left[\frac{\widehat{\alpha}+\overline{\alpha}}{2}(1-p) + \left(1-\frac{\widehat{\alpha}+\overline{\alpha}}{2}\right)(\tau-p)\right] + (1-\xi)^M(1-p) \\ < (1-\overline{\alpha})(\tau-p)+\overline{\alpha}(1-p)\end{aligned} \tag{20}$$

Since $\widehat{\alpha} < \overline{\alpha}$, such an M exists. By definition, M depends only on ξ, $\overline{\alpha}$ and $\widehat{\alpha}$, and does not depend on N and δ.

Let x_t be the probability that the government sets high taxes in period t when $\theta_t = G$. Let $l_t \equiv \mu_t/(1-\mu_t)$ be the log likelihood ratio between the commitment type and the rational type. If the citizen does not invest in period t, we have $l_{t+1} = l_t$. After the citizens observe low taxes in period t, their posterior log likelihood ratio is

$$l_{t+1} = l_t + \log \frac{1}{1-x_t}, \tag{21}$$

and after they observe high taxes in period t, their posterior log likelihood ratio is

$$l_{t+1} = l_t + \log \frac{p}{p+(1-p)x_t}. \tag{22}$$

Note that $\log \frac{1}{1-x_t} = \log \frac{p}{p+(1-p)x_t} = 0$ when $x_t = 0$, and

$$\log \frac{1}{1-x_t} \Big/ \log \frac{p}{p+(1-p)x_t} \tag{23}$$

is continuous and strictly increasing in x_t for all $x_t \in (0,1)$.

Recall that (23) equals $\alpha/(1-\alpha)$ when $x_t = \gamma^*$. Since $\widehat{\alpha} > \alpha$, there exists $\eta > 0$ such that

$$(1-\widehat{\alpha}) \log \frac{1}{1-x_t} + \widehat{\alpha} \log \frac{p}{p+(1-p)x_t} \leq 0, \quad \forall x_t \in [0, \gamma^* + \eta]. \tag{24}$$

Fixing η, there exists $\bar{l} > 0$ such that when $l_t < \bar{l}$, the citizens have an incentive to invest only if $x_t \leq \gamma^* + \eta$. We consider two cases, depending on whether $l_0 \leq \bar{l} - 1$.

Case 1: $l_0 \leq \bar{l} - 1$ Consider the payoff of the strategic government if it deviates and sets high taxes with probability $\widehat{\alpha}$ in every period where the citizen invests, and this probability is independent of its behavior in period t' for every $t \neq t'$. This is without loss of generality. First, since we look at histories where l_t is lower than $\frac{1-\gamma^*}{\gamma^*}$, the citizens invest only if $x_t < 1$. Second, when the citizens do not invest or when they invest and $x_t = 0$, there is no learning and the government's continuation value in period $t+1$ is at least $\frac{1}{\delta}(V_t - (1-\delta)(\tau - p))$, which is strictly greater than V_t as long as $V_t > \tau - p$.

Let $y_t = 1$ if the government sets low taxes and let $y_t = 0$ if the government sets high taxes. Since $\mathbb{E}[y] = 1 - \widehat{\alpha}$, Chernoff-Hoeffding inequality implies that for every

$N \in \mathbb{N}$ and $\varepsilon > 0$, we have

$$\Pr[\sum_{i=1}^{N} y_i > N(1 - \widehat{\alpha} - \varepsilon)] > 1 - \exp(-2N\varepsilon^2). \tag{25}$$

Inequality (24) implies that when $x_t \leq \gamma^* + \eta$ for all $t \in \{1, 2, ..., N\}$, $\{l_0, ..., l_N\}$ is a supermartingale bounded from below by $l_0 + N \log p$. The Doob's Upcrossing inequality leads to an upper bound on the probability that $\{l_t\}_{t=1}^{N}$ crosses $l_0 + \frac{1}{M}$ at least once

$$\begin{aligned} \Pr\left(l_t \leq l_0 + \frac{1}{M},\ \forall t \in \{1, 2, ..., N\}\right) &\geq \frac{1/M}{N \log p + 1/M} \\ &= \frac{1}{1 + MN \log p}. \end{aligned} \tag{26}$$

Since M is independent of N, there exists $\overline{N} \in \mathbb{N}$ such that when $N > \overline{N}$, we have

$$\frac{1}{1 + MN \log p} > 2 \exp(-2N\varepsilon^2). \tag{27}$$

This together with (25) implies that there exists an event measurable with respect to the σ-algebra induced by $\{l_0, ..., l_N\}$ such that

$$l_N \leq l_0 + \frac{1}{M} \text{ and } \sum_{i=1}^{N} \delta^i y_i \geq (1 - \frac{\widehat{\alpha} + \overline{\alpha}}{2}) \sum_{i=1}^{N} \delta^i$$

. Intuitively, the government's reputation after N periods increases by no more than $1/M$, yet its discounted average payoff in these N periods is no more than $\tau - p + \frac{\widehat{\alpha} + \overline{\alpha}}{2}(1 - \tau)$. Since its discounted average payoff is more than $\tau - p + \overline{\alpha}(1 - \tau)$, its continuation value increases after these N periods.

Let E_0 be such an event. Similarly, conditional on E_0, there exists an event E_1 that occurs with positive probability and is measurable with respect to the σ-algebra induced by $\{l_0, ..., l_{2N}\}$ such that

$$l_{2N} \leq l_N + \frac{1}{M} \text{ and } \sum_{i=N+1}^{2N} \delta^{i-N} y_i \geq (1 - \frac{\widehat{\alpha} + \overline{\alpha}}{2}) \sum_{i=1}^{N} \delta^i.$$

Iterating this procedure M times, we obtain a sequence of actions with length MN, $h^* \equiv \{a_0^*, ..., a_{MN-1}^*\}$ such that if the government plays according to h^* when the realized states from period 0 to $MN - 1$ are all G, then for all $m \in \{0, 1, .., M - 1\}$,

$l_{MN} \leq \bar{l}$ and

$$\sum_{i=mN+1}^{(m+1)N} \delta^{i-mN} y_i \geq (1 - \frac{\widehat{\alpha} + \overline{\alpha}}{2}) \sum_{i=1}^{N} \delta^i.$$

Let $\{x_t\}_{t\in\mathbb{N}}$ be such that for every $s \in \mathbb{N}$ with $x_s = 0$, x_t does not depend on the realization of y_s for every $t > s$. Conditional on $\theta_t = G$ for all $t \in \{0, ..., MN-1\}$, suppose the government plays a_t^* in period t whenever $x_t \neq 0$ and sets low taxes in period t if $x_t = 0$, then its discounted average payoff in each of the first M N-period block is less than

$$\frac{\widehat{\alpha} + \overline{\alpha}}{2}(1-p) + (1 - \frac{\widehat{\alpha} + \overline{\alpha}}{2})(\tau - p)$$

and the log likelihood ratio in period MN is no more than $\bar{l}$. However, (20) implies that if the government's equilibrium payoff is $(\tau - p) + \overline{\alpha}(1-\tau)$, then its continuation value after NM periods is more than $1-p$. This leads to a contradiction since $1-p$ is its highest feasible payoff.

Case 2: $l_0 > \bar{l} - 1$ Let

$$L \equiv \left| \log \frac{p}{p + (1-p)(\gamma^* + \eta)} \right| \quad \text{and} \quad K \equiv \left\lceil \frac{l_0 - (\bar{l} - 1)}{L/(2\widehat{\alpha})} \right\rceil. \tag{28}$$

Choose ξ small enough that satisfies (18) as well as

$$\Big(1 - (1-\xi)^K\Big)(1-p) + (1-\xi)^K \Big(\frac{2\overline{\alpha} + \widehat{\alpha}}{3}(1-p) + \big(1 - \frac{2\overline{\alpha} + \widehat{\alpha}}{3}\big)(\tau - p) \Big) < (1-\overline{\alpha})(\tau - p) + \overline{\alpha}(1-p). \tag{29}$$

Inequality (29) implies that if $1 - \delta^N = \xi$ and the government sets high taxes for KN consecutive periods, then its continuation value exceeds

$$\frac{2\overline{\alpha} + \widehat{\alpha}}{3}(1-p) + \big(1 - \frac{2\overline{\alpha} + \widehat{\alpha}}{3}\big)(\tau - p).$$

Let $\widetilde{y}_t$ be a random variable such that $\widetilde{y}_t = y_t$ if $x_t \leq \gamma^* + \eta$ and $\widetilde{y}_t = 0$ if $x_t > \gamma^* + \eta$. Let $\{\widetilde{Z}_t\}_{t\in\mathbb{N}}$ be a sequence of random variables such that $\widetilde{Z}_t = \log \frac{1}{1-x_t}$ when $\widetilde{y}_t = 1$ and $\widetilde{Z}_t = \log \frac{p}{p+(1-p)x_t}$ when $\widetilde{y}_t = 0$. Let $l_t \equiv l_{t-1} + \widetilde{Z}_t$. By construction, $\{l_t\}_{t=0}^N$ is a supermartingale bounded from below by $l_0 + N \log p$, and therefore, (26) still applies. Next, we bound the probability of event $\sum_{i=1}^N \widetilde{y}_t \geq N(1 - \widehat{\alpha} - \varepsilon)$ from below. Let $h^t \equiv \{y_0, ..., y_{t-1}\}$ and let $x_t(h^t)$ be the value of x_t conditional on $(\theta_0, .., \theta_t) = (G, ..., G)$ and the history of actions is y_t. Let

$$\mathcal{E}_t \equiv \{h^t | x_t(h^t) > \gamma^* + \eta \text{ and } x_s(h^s) \leq \gamma^* + \eta \text{ for every } h^s \prec h^t\} \tag{30}$$

and let $P \equiv \sum_{t=1}^{N} \Pr(\mathcal{E}_t)$. Inequality (25) implies that

$$\Pr\Big(\sum_{i=1}^{N} \widetilde{y}_t \geq N(1-\widehat{\alpha}-\varepsilon)\Big) \geq \frac{1-\exp(-2N\varepsilon^2)-\widehat{\alpha}P}{1-\widehat{\alpha}P}. \tag{31}$$

If $h^t \in \mathcal{E}_t$, then $l_t - l_{t-1} = \log \frac{p}{p+(1-p)x_t} \leq \log \frac{p}{p+(1-p)(\gamma^*+\eta)} = |L|$. Since the expected value of $\widetilde{Z}_t$ is non-positive when $x_t \leq \gamma^* + \eta$, we have

$$\mathbb{E}[l_N - l_0] \leq -LP. \tag{32}$$

For every N-period block, consider two cases,

1. If $P \leq \frac{1}{2\widehat{\alpha}}$, then

   $$\Pr\Big(\sum_{i=1}^{N} \widetilde{y}_t \geq N(1-\widehat{\alpha}-\varepsilon)\Big) + \Pr\Big(l_t \leq l_0 + \frac{1}{M}, \forall t \in \{1,2,...,N\}\Big) > 1, \tag{33}$$

   which implies the existence of an action sequence $(a_0^*, ..., a_{N-1}^*)$ such that if the government plays according to this when $(\theta_0, ..., \theta_{N-1}) = (G, ..., G)$, we have

   $$\sum_{i=1}^{N} \delta^i y_i \geq (1 - \frac{\widehat{\alpha}+\overline{\alpha}}{2}) \sum_{i=1}^{N} \delta^i \text{ and } l_N \leq l_0 + \frac{1}{M}.$$

2. If $P > \frac{1}{2\widehat{\alpha}}$, then since $\mathbb{E}[l_N - l_0] \leq -LP \leq -\frac{L}{2\widehat{\alpha}}$, there is $(a_0^*, ..., a_{N-1}^*)$ such that when the government plays according to this and $(\theta_0, ..., \theta_{N-1}) = (G, ..., G)$, we have $l_N \leq l_0 - \frac{L}{2\widehat{\alpha}}$. The government's continuation value in period N, denoted by $V_N(a_0^*, ..., a_{N-1}^*)$, satisfies:

   $$(1-\delta^N)(1-p) + \delta^N V_N(a_0^*, ..., a_{N-1}^*) \geq (1-\overline{\alpha})(\tau - p) + \overline{\alpha}(1-p). \tag{34}$$

   Equation (28) implies that there exists at most K such blocks.

According to (29), after at most $K + M$ blocks, there exists an on-path history under which the government's continuation value exceeds $1 - p$, from which we obtain a contradiction.