Available online at www.sciencedirect.com

ScienceDirect

Journal of Economic Theory 204 (2022) 105508

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

A reputation for honesty [☆]

Drew Fudenberg [a], Ying Gao [a], Harry Pei [b,*]

[a] *Department of Economics, Massachusetts Institute of Technology, United States of America*
[b] *Department of Economics, Northwestern University, United States of America*

Received 24 May 2021; final version received 25 April 2022; accepted 13 June 2022
Available online 16 June 2022

Abstract

We analyze situations where players build reputations for honesty rather than for playing particular actions. A patient player faces a sequence of short-run opponents. Before players act, the patient player announces their intended action after observing both a private payoff shock and a signal of what actions will be feasible that period. The patient player is either an honest type who keeps their word whenever their announced action is feasible, or an opportunistic type who freely chooses announcements and feasible actions. Short-run players only observe the current-period announcement and whether the patient player has kept their word in the past. We provide sufficient conditions under which the patient player can secure their optimal commitment payoff by building a reputation for honesty. Our proof introduces a novel technique based on concentration inequalities.

JEL classification: C73; D82; D83

Keywords: Reputation; Honesty; Learning; Concentration inequalities

1. Introduction

Many economic actors have reputations for keeping or breaking their promises. For example, firms make non-binding promises to their employees about bonuses and promotions, with the op-

☆ We thank Mehmet Ekmekci, Faruk Gul, Navin Kartik, Giacomo Lanzani, David Levine, Wojciech Olszewski, Larry Samuelson, an associate editor, and two referees for helpful comments. We thank the National Science Foundation grants SES-1947021 and SES-1951056 for financial support.

* Corresponding author.
E-mail addresses: drewf@mit.edu (D. Fudenberg), gaoy@mit.edu (Y. Gao), harrydp@northwestern.edu (H. Pei).

https://doi.org/10.1016/j.jet.2022.105508
0022-0531/

tion to renege on them. However, failing to honor promises may make employees feel aggrieved and undermine workplace morale. Similarly, advertising can set customers' expectations, and if those expectations are not aligned with the actual customer experience, the firm's brand and business will suffer.

This paper examines when and whether a reputation for honesty might facilitate commitment. Compared to reputations for taking specific actions, a reputation for honesty allows decisions to better adapt to current circumstances, which is valuable when the environment changes over time. We focus on reputations for honestly announcing intended actions rather than reputations for honestly announcing payoff-relevant states. This is because in some applications, states could be difficult to verify ex post, and it seems unrealistic to make commitments based on future contingencies that are hard to describe in advance. Making promises about their intended actions may be simpler.

However, building a reputation for honesty can be challenging when some actions might turn out to be infeasible and the reputation-building player has imperfect information about the actions they will be able to take. As a result, players may renege on promises they had intended to keep, as happened to Lincoln Electric in 1992. It promised to share its domestic profits with its workers, but by the end of the year, its surplus in domestic business was unexpectedly wiped out by losses in recently acquired foreign operations, making it hard to pay high bonuses to its workers.

In our model, a patient player (e.g., a firm) faces a sequence of myopic opponents (e.g., consumers). Each period, the patient player observes a private payoff shock (e.g., their production cost), as well as some information about which of their actions are currently feasible. Then the patient player announces the action they intend to play, after which players act. The myopic players cannot observe the patient player's feasible actions or the payoff shocks, but can observe the patient player's announcement in the current period and whether the patient player has kept their word in the past.[1]

The patient player is either an *honest type*, who strategically chooses their announcements but keeps their word whenever their announced action is feasible, or an *opportunistic type*, who strategically chooses both the announcements and the actions. Note that the honest type is not a "pure commitment" type, as it is not committed to any particular action-announcement pair. Instead, the honest type optimizes its announcements, but unlike the opportunistic type it is constrained to implement them whenever that is possible.

Our main result shows that if (1) the distribution over feasible action sets has full support, and (2) the patient player knows their feasible action set with high probability when they announce their intended action, then each type of the patient player receives at least their optimal commitment payoff in every equilibrium. This does not follow from Fudenberg and Levine (1989), since the honest type may not announce the opportunistic type's optimal commitment action. As a result, the opportunistic type cannot guarantee their optimal commitment payoff by playing the honest type's equilibrium strategy.

To explain this result, first consider situations where the patient player knows which of their actions are feasible at the announcement stage. In this case, every type can guarantee their optimal commitment payoff by announcing their optimal commitment action in every period and keeping their word. Intuitively, when a myopic player fails to best respond to any announced

[1] In Section 5, we show that Theorem 1 generalizes to situations where player 2 observes a bounded history of player 1's past actions and announcements in addition to whether player 1 has kept their word. We also provide an example showing that when player 2 can observe the entire history of player 1's past actions and announcements, there is an equilibrium in which player 1's payoff is strictly less than his commitment payoff.

action, their belief puts a significant probability on the event that the patient player is opportunistic and will break their word. Hence, observing the patient player keep their word increases the posterior probability of the honest type. Thus, the event that the patient player is opportunistic and breaks their word with high probability can happen in at most a bounded number of periods, regardless of the equilibrium.

When the patient player makes their announcements without being certain about which actions will be feasible, their reputation may deteriorate in expectation even if they announce their optimal commitment action and keep their promise when their announced action is feasible, because the probability they will be forced to renege may differ under the honest type's announcement strategy and under the optimal commitment announcement. This feature is not present in Fudenberg and Levine (1989, 1992) and subsequent work that provides lower bounds for the value of a reputation. This is because in those models, when the long-run player plays a fixed commitment action, the probability that they are the corresponding commitment type weakly increases on average in every period.

Our methodological contribution is to analyze reputation with concentration inequalities. Consider either type of patient player's payoff from a deviation that (1) announces their optimal commitment action in "good" periods where the short-run player best replies to any announcement, and plays the announced action whenever it is feasible, and (2) in the other "bad" periods, plays the honest type's equilibrium strategy. Since the short-run players do not best reply to all announced actions in bad periods, and the distribution of feasible action sets has full support, the opportunistic type must be breaking their word with significant probability in equilibrium. This yields a lower bound on the expected increase in the log likelihood ratio between the honest type and the opportunistic type. When the patient player knows their feasible action set with high probability, the honest type keeps their word with high probability in every good period. Although the log likelihood ratio may decrease in expectation, the magnitude of this decrease is bounded from above.

Based on these observations, we establish an upper bound on the undiscounted frequency of bad periods using the Azuma-Hoeffding inequality (Azuma, 1967) and then derive an upper bound on the discounted frequency of bad periods using summation by parts. The expected number of bad periods is unbounded, unlike in Fudenberg and Levine (1992), but the discounted frequency of the bad periods goes to zero as the patient player's information about feasible actions becomes arbitrarily precise. This yields the lower bound on the patient player's payoff from the deviation we proposed. Since such a deviation is feasible for both types, both the opportunistic and the honest type can obtain their respective optimal commitment payoffs in every equilibrium.

Section 2 sets up the baseline model and presents an example motivating the study of reputation for honesty and issues related to action feasibility. Sections 3 and 4 state and prove the main result. Section 5 presents extensions and discusses our assumptions on the monitoring structure. Section 6 explains our contributions to the reputation literature and Section 7 concludes.

2. Baseline model

Time is discrete, indexed by $t = 0, 1\dots$. A long-lived player 1 (e.g., a seller) with discount factor $\delta \in (0, 1)$ interacts with an infinite sequence of short-lived player 2s (e.g., consumers), with 2_t denoting the short-lived player in period t.

Let A be the potentially feasible set of actions for player 1, with $\mathcal{A} \equiv 2^A \backslash \{\varnothing\}$ the collection of all non-empty subsets of A. Each player 2's action set is B. In period t, $(\theta_t, A_t) \in \Theta \times \mathcal{A}$ is drawn according to $p \in \Delta(\Theta \times \mathcal{A})$, where $\theta_t \in \Theta$ affects player 1's stage-game payoff (e.g., their cost of supplying high quality), and $A_t \subset A$ is the set of *feasible actions* for player 1. We assume that A, B, and Θ are finite sets. We also assume that for every $s \neq t$, θ_t is independent of θ_s, A_t is independent of A_s, and θ_t is independent of A_s.[2]

Each period consists of an *announcement stage* and an *action stage*. In the announcement stage, player 1 privately observes θ_t and a signal $\widetilde{A}_t$ of their feasible actions, where $A_t \subseteq \widetilde{A}_t \subseteq A$ and $\widetilde{A}_t$ is drawn according to $G(\cdot|A_t)$. Then player 1 announces to player 2_t that they intend to play action $m_t \in A$. In the action stage, player 1 observes A_t, and then players simultaneously choose their actions $a_t \in A_t$ and $b_t \in B$. Intuitively, player 1 learns at the announcement stage that some of their actions are infeasible, but at the action stage they may learn that other actions are infeasible as well.

Player 1 has persistent private information about their type $\gamma \in \{\gamma_h, \gamma_o\}$, where γ_h stands for an *honest type* and γ_o stands for an *opportunistic type*. The honest type is restricted (i) to announce an action that might be feasible, i.e., $m_t \in \widetilde{A}_t$, and (ii) to take an action that matches their announced action if it is feasible, i.e., $a_t = m_t$ if $m_t \in A_t$. The opportunistic type can announce any action, including ones that do not belong to $\widetilde{A}_t$, and can take any action in A_t regardless of their announcement. Our baseline model focuses on the case of two types in order to simplify the exposition. We generalize our main result (Theorem 1) to any finite number of honest and opportunistic types in Section 5.

Let $\pi_0 \in (0, 1)$ be the common prior probability the short-run players assign to the honest type. Player 2_t observes $\{y_0, ..., y_{t-1}\}$ in addition to m_t before choosing b_t, where the *record* $y_s \equiv \mathbf{1}\{a_s = m_s\}$ tracks whether player 1 has kept their word but does not track their actions played or their announcements.

For every $t \in \mathbb{N}$, player 2_t's history is $h_2^t \equiv \{y_0, y_1, ..., y_{t-1}, m_t\}$. Let $\mathcal{H}_2^t \equiv \{0, 1\}^t \times A$ be the set of player 2_t's histories. Player 2_t's strategy is $\sigma_2^t : \mathcal{H}_2^t \to \Delta(B)$, with $\sigma_2 \equiv (\sigma_2^t)_{t\in\mathbb{N}}$. Player 1's history in the announcement stage of period t is[3]

$$\widetilde{h}_1^t \equiv \{\gamma, \theta_0, ..., \theta_t, \widetilde{A}_0, ..., \widetilde{A}_t, A_0, ..., A_{t-1}, m_0, ..., m_{t-1}, a_0, ..., a_{t-1}, b_0, ..., b_{t-1}\}.$$

For every $t \in \mathbb{N}$, let $\widetilde{\mathcal{H}}_1^t$ be the set of $\widetilde{h}_1^t$. Let $\widetilde{\mathcal{H}}_1 \equiv \bigcup_{t=0}^{\infty} \widetilde{\mathcal{H}}_1^t$ be the set of player 1's histories at the announcement stage. Player 1's history in the action stage of period t is

$$h_1^t \equiv \{\gamma, \theta_0, ..., \theta_t, \widetilde{A}_0, ..., \widetilde{A}_t, A_0, ..., A_t, m_0, ..., m_t, a_0, ..., a_{t-1}, b_0, ..., b_{t-1}\}.$$

For every $t \in \mathbb{N}$, let $\mathcal{H}_1^t$ be the set of h_1^t. Let $\mathcal{H}_1 \equiv \bigcup_{t=0}^{\infty} \mathcal{H}_1^t$ be the set of player 1's histories at the action stage. Type γ's strategy is $(\widetilde{\sigma}_\gamma, \sigma_\gamma)$, with $\widetilde{\sigma}_\gamma : \widetilde{\mathcal{H}}_1 \to \Delta(A)$ and $\sigma_\gamma : \mathcal{H}_1 \to \Delta(A)$. At each t, type γ_o can only take actions that belong to A_t. Type γ_h faces this constraint as well as the requirements that the support of $\widetilde{\sigma}_{\gamma_h}(\widetilde{h}_1^t)$ is a subset of $\widetilde{A}_t$, and that $\sigma_{\gamma_h}(h_1^t) = m_t$ whenever $m_t \in A_t$.

Type γ's stage-game payoff is $u_1(\gamma, \theta_t, a_t, b_t)$ and player 2_t's is $u_2(a_t, b_t)$. Note that the stage-game payoffs of type γ_o and type γ_h can differ, and that player 2_t's payoff does not depend on γ and θ_t.

---

2 Theorem 1 extends when the distribution of (θ_t, A_t) varies over time, although the statement of the result becomes more involved. We do not know whether the result extends when (θ_t, A_t) is correlated over time.

3 As in most of the literature on repeated games, we allow a history to be both an element of a set of histories and a random variable whose realization belongs to the set of histories. We will distinguish between the two in our proof, and in particular, highlight places where "history" is a random variable.

A Nash equilibrium is a strategy profile $(\{\widetilde{\sigma}_\gamma, \sigma_\gamma\}_{\gamma\in\Gamma}, \sigma_2)$, in which σ_2^t maximizes player 2_t's stage-game payoff, and for each $\gamma \in \{\gamma_h, \gamma_o\}$, $(\widetilde{\sigma}_\gamma, \sigma_\gamma)$ maximizes type γ's discounted average payoff $\sum_{t=0}^{\infty}(1-\delta)\delta^t u_1(\gamma, \theta_t, a_t, b_t)$ subject to the constraints that type γ faces. Because the stage game and set of types are both finite and payoffs are discounted, the game is continuous at infinity in the sense of Fudenberg and Levine (1983), and it is straightforward to adapt their arguments to show that a Nash equilibrium exists.[4]

Example: product choice game Player 1 is a firm and player 2s are consumers. Every period, the firm privately observes their cost of production $\theta_t \in \{\theta_g, \theta_b\}$. We assume that $\theta_t = \theta_g$ with probability $\frac{1}{2}$ for every $t \in \mathbb{N}$, and θ_t is independent of θ_s for every $s \neq t$. In this example, the honest and opportunistic types have the same payoff function, given by the matrices below.

$\theta_t = \theta_g$	T	N
H	$1, 2$	$-1, 0$
L	$2, -3$	$0, 0$

$\theta_t = \theta_b$	T	N
H	$-1, 2$	$-3, 0$
L	$2, -3$	$0, 0$

Since the firm's optimal (pure) commitment action is H when $\theta_t = \theta_g$ and is L when $\theta_t = \theta_b$, they have an incentive to build a reputation for keeping their word by announcing their intended action in every period. This is better than a reputation for always playing H, since it lets the firm avoid the cost of choosing H when $\theta_t = \theta_b$.

Suppose that the firm can make an announcement m_t about its intended action a_t to the period-t consumer after observing θ_t but before taking actions. The period-t consumer observes m_t, as well as whether the firm's announcement matched its action in previous periods. With positive probability, the firm is an *honest type* who strategically chooses their announcements but commits to keep their word. With complementary probability, the firm is an *opportunistic type* who can freely choose their announcements and actions. The period-t consumer observes m_t, as well as whether the firm's announcement matched its action in previous periods.

In this model, there are equilibria in which both types of the firm receive their minmax value 0.[5] This low-payoff equilibrium hinges on the assumption that all of the firm's actions are feasible in every period. It does not fit a number of applications of interest, where with positive probability some of the firm's actions might be infeasible. For example, when the firm is an individual contractor, they can occasionally be sick, and so unable to provide high-quality service. The firm may also face occasional regulatory inspections during which playing L can lead to a risk of fines and being shut down. In this situation, the firm will always choose to supply high quality.[6]

Another practical concern is that firms may not know their feasible action set when making announcements, and might be forced to renege on promises they intended to keep. For example, the patient player might be a contractor who believes that they can provide high-quality service with high probability, and promises to do that, but later realize that they cannot deliver on their promise due to technical difficulties or other priorities on their schedule.

[4] Specifically the set of strategies is compact in the product topology and payoff functions are continuous in that topology, so any sequence of Nash equilibria of the finite-horizon truncation of the game has an accumulation point, and any accumulation point is a Nash equilibrium of the infinite horizon game.

[5] See the working paper version Fudenberg et al. (2020) for an explicit construction.

[6] Formally, if the firm chooses L in periods where they are inspected, they face a probability $q \in (0,1)$ of shutting down and a fine $f > 0$. One can show that there exist $\underline{f} > 0$ and $\underline{q} \in (0,1)$ such that when $f > \underline{f}$ and $q > \underline{q}$, it is a dominant strategy for both types of the firm to choose $a_t = H$ at the action stage.

Motivated by these observations, we assume that the firm's feasible action set A_t is drawn according to a *full support distribution* over $\{\{H\}, \{L\}, \{H, L\}\}$. At the announcement stage, the firm observes $\widetilde{A}_t \supset A_t$. That is, the firm knows that action H is infeasible when $\widetilde{A}_t = \{L\}$ and vice versa, and when $\widetilde{A}_t = \{H, L\}$, the firm recognizes the possibility that either H or L may turn out to be infeasible. The honest type announces an action in $\widetilde{A}_t$ and keeps their word whenever feasible. The opportunistic type only faces a feasibility constraint that the action they take belongs to A_t.

Our result shows that each type of the patient firm can secure their optimal pure-strategy commitment payoff in every equilibrium when (1) at the announcement stage, the firm knows their feasible action set with sufficiently high probability, and (2) the ratio between the probability with which H is infeasible when $\widetilde{A}_t = \{H, L\}$ and the probability with which L is infeasible when $\widetilde{A}_t = \{H, L\}$ is neither 0 nor infinity.[7]

Remark. In our model, the patient player privately observes the realization of an i.i.d. state before they announce their intended action. Our main result applies in environments where Θ is a singleton, but in this case the patient player can also secure their commitment payoff by establishing a reputation for playing their optimal commitment action; the i.i.d. state makes a reputation for honesty more interesting. Alternatively, one could consider settings in which the patient player announces the realized state, but a reputation for truthfully announcing the state on its own does not guarantee the commitment payoff in all equilibria. For example, in the game we considered earlier, there exists an equilibrium in which both types of the patient player play L and the short-run players play N regardless of the announcement. Moreover, as we noted in the introduction, states may be hard to monitor ex-post. In Section 5, we show that our result extends to the case where the patient player announces the realized state *in addition* to announcing his intended action, i.e., our main result applies when the patient player can make additional announcements.

3. Main result

Let $\text{BR}_2 : \Delta(A) \to 2^B \backslash \{\varnothing\}$ be player 2's *best reply correspondence*. Type γ's *optimal commitment payoff* in state θ_t when the set of feasible actions is A_t is

$$U^*(\gamma, \theta_t, A_t) \equiv \max_{a \in A_t} \left\{ \min_{b \in \text{BR}_2(a)} u_1(\gamma, \theta_t, a, b) \right\}, \tag{3.1}$$

and type γ's *optimal commitment actions* in A_t under state θ_t are the actions that attain this maximum. Type γ's (expected) *optimal commitment payoff* is

$$\overline{U}^*(\gamma) \equiv \sum_{(\theta_t, A_t) \in \Theta \times \mathcal{A}} p(\theta_t, A_t) U^*(\gamma, \theta_t, A_t). \tag{3.2}$$

Let $\Pr(\cdot|\widetilde{A}_t) \in \Delta(\mathcal{A})$ be player 1's belief about A_t after observing $\widetilde{A}_t$. Since A_t is distributed according to p and $\widetilde{A}_t$ is distributed according to $G(\cdot|A_t)$ conditional on A_t, $\Pr(\cdot|\widetilde{A}_t)$ is derived from Bayes rule for every $\widetilde{A}_t$ that occurs with positive probability under some on-path A_t.

Assumption 1.

[7] A pure-strategy commitment is a map from (θ, A_t) to A_t. Our results still hold if there are honest types that can announce mixed actions, as long as there is an honest type that can only make pure-strategy commitments.

(i) Every $\widetilde{A}_t \in \mathcal{A}$ occurs with positive probability.
(ii) For every $\widetilde{A}_t \in \mathcal{A}$, either $\Pr(A_t = \widetilde{A}_t | \widetilde{A}_t) = 1$, or $\Pr(a \notin A_t | \widetilde{A}_t) > 0$ for every $a \in \widetilde{A}_t$.

Assumption 1(i) requires that every non-empty subset of A is the set of feasible actions with positive probability; it rules out situations in which all of player 1's actions are feasible with probability 1. When that is the case, our example in Section 2 implies that the patient player can receive their minmax value in some equilibria. Assumption 1(ii) requires that upon observing $\widetilde{A}_t$, either player 1 knows that $A_t = \widetilde{A}_t$, or every action in $\widetilde{A}_t$ is infeasible with positive probability. A sufficient condition for Assumption 1(ii) is that there exists $M > 0$ such that $G(\widetilde{A}_t | A'_t) \geq M \cdot G(\widetilde{A}_t | A''_t)$ for every $A'_t \subsetneqq \widetilde{A}_t$ and $A''_t \subsetneqq \widetilde{A}_t$.

Theorem 1. *Fix an $\varepsilon > 0$. Then there exist $\underline{\delta} \in (0, 1)$ and $\eta > 0$ such that if p and G satisfy Assumption 1, $\delta > \underline{\delta}$, and $G(\widetilde{A}_t = A_t | A_t) \geq 1 - \eta$ for every $A_t \in \mathcal{A}$, then each type $\gamma \in \{\gamma_o, \gamma_h\}$ receives payoff at least $\overline{U}^*(\gamma) - \varepsilon$ in every equilibrium.*

Theorem 1 shows that each type of patient player can secure (approximately) their optimal commitment payoff when they know their feasible action set with probability above some cutoff at the announcement stage. Section 5 discusses several extensions, including situations where player 1 has more than two types, or when player 2_t observes $\{y_0, ..., y_{t-1}\}$ with noise, or when player 2_t can also observe a bounded number of signals of player 1's past actions and announcements in addition to $\{y_0, ..., y_{t-1}\}$ and m_t, or when $A_t \nsubseteq \widetilde{A}_t$ with small but positive probability, and more generally, the role of our modeling assumptions.

Section 5 also presents an example where Theorem 1 fails when player 2's observe the entire history of player 1's actions and announcements. Here the opportunistic type cannot build a reputation for honesty by simply keeping their word, they also need to announce actions that the honest type announces with high probability. Since the honest type and the opportunistic type can have different stage-game payoff functions, the opportunistic type may face a tradeoff between announcing actions that lead to a high payoff and announcing actions that lead to a better reputation.

For a snapshot of our argument, fix any equilibrium, and consider type γ's payoff when they announce an optimal commitment action for $\widetilde{A}_t$ under state θ_t, that is

$$a^*(\gamma, \theta_t, \widetilde{A}_t) \in \arg\max_{a \in \widetilde{A}_t} \left\{ \min_{b \in \mathrm{BR}_2(a)} u_1(\gamma, \theta_t, a, b) \right\}, \tag{3.3}$$

and keep their word whenever feasible. We call this the "naive commitment strategy".

Let us start from the case in which player 1 has perfect information about their feasible action set A_t at the announcement stage, i.e. $G(\widetilde{A}_t = A_t | A_t) = 1$ for every A_t. The naive commitment strategy generates the same distribution over $\{y_0, y_1, ..., y_{t-1}\}$ as the honest type's equilibrium strategy since the honest type always keeps their word.[8] Because every $\widetilde{A}_t$ occurs with positive probability, the honest type announces each of their actions with positive probability, so player 2_t cannot rule out the honest type regardless of player 1's period-t announcement. Thus if player 2_t fails to best reply to the announced action, their belief must assign a significant probability to the event that $a_t \neq m_t$, which implies that observing $a_t = m_t$ increases the posterior probability with which player 1 is honest. As a result, the expected number of "bad" periods where player 2 does

[8] In this sentence, $\{y_0, y_1, ..., y_{t-1}\}$ is a random variable rather than an element of a set.

not best reply to all announcements is bounded from above uniformly in δ, which is why each type of patient player receives at least their optimal commitment payoff in every equilibrium.

Next, suppose the patient player has imperfect information about which of their actions are feasible, i.e., $G(\widetilde{A}_t = A_t | A_t) < 1$, and consider type γ's payoff from the "naive commitment strategy": announce the optimal commitment action $a^*(\gamma, \theta_t, \widetilde{A}_t)$, play the announced action whenever $a^*(\gamma, \theta_t, \widetilde{A}_t) \in A_t$, and otherwise play an arbitrary action in A_t. Because the announced action may not be feasible, the following issues arise:

1. When type γ deviates to the naive commitment strategy, it can induce a different distribution of y than that induced by player 1's equilibrium strategy,[9] and the induced distribution may not be absolutely continuous with respect to the equilibrium distribution. This precludes the direct application of the results in Fudenberg and Levine (1989, 1992) and Sorin (1999). Moreover, the short-run players observe the long-run player's announcement before moving, and announcements and actions are imperfectly recalled, which makes the payoff lower bounds provided by Gossner (2011) and Ekmekci et al. (2012) not directly applicable.[10]
2. The opportunistic type's reputation can deteriorate in expectation under the naive commitment strategy.[11] This is the case when announcing the optimal commitment action induces a distribution that is closer to the distribution induced by the opportunistic type's equilibrium strategy than to the honest type's equilibrium strategy. This is not the case when the deviation is to imitate the play of a positive-probability commitment type.

Our proof bounds the patient player's payoff by examining an auxiliary misspecified learning problem faced by the short-run players where the true data generating process is the one induced by the patient player's deviating strategy, and the data generating processes in the support of their prior belief are the ones induced by the honest type's and the opportunistic type's equilibrium strategies.[12]

We develop a novel argument using concentration inequalities. For any given equilibrium, consider type γ's payoff from the following deviation: (1) In "good periods" where player 2_t best replies to any announcement, announce $a^*(\gamma, \theta_t, \widetilde{A}_t)$ and play the announced action whenever it is in A_t; (2) in other "bad periods", imitate the honest type's equilibrium strategy. The opportunistic type's probability of playing honestly is bounded away from 1 in every bad period, which leads to a lower bound on the Kullback-Leibler divergence between its induced distribution over outcomes and that under the honest type's equilibrium strategy.[13] In every good period,

[9] This occurs when type γ's optimal commitment action is less likely to be feasible than the action announced by the honest type in equilibrium. In the example in Section 2, suppose $\Pr(A_t = \{H\} | \widetilde{A}_t = \{H, L\}) = \varepsilon$ and $\Pr(A_t = \{L\} | \widetilde{A}_t = \{H, L\}) = 2\varepsilon$. Type γ_h announces L in equilibrium when $\theta_t = \theta_g$ and $\widetilde{A}_t = \{H, L\}$ while type γ's optimal commitment action when $\theta_t = \theta_g$ and $\widetilde{A}_t = \{H, L\}$ is H.

[10] Ekmekci et al. (2012) studies a reputation model in which the patient player's type changes over time. The key step in their proof is to bound payoffs by a function of the discounted sum of divergences between the equilibrium histories seen by player 2_n and the histories when player 1 imitates the commitment type. In the limit when the probability of type change in every period goes to 0, the sum goes to 0 and the bound approaches the optimal commitment payoff. Appendix B explains why an analogous argument does not work here.

[11] The expectation here is taken at the ex-ante stage before player 1 observes θ_t and $\widetilde{A}_t$.

[12] Past work on misspecified learning, e.g. Esponda and Pouzo (2016), Bohren and Hauser (2021), Esponda and Pouzo (2020), Fudenberg et al. (2021), Esponda et al. (2021), Bohren and Hauser (2021), and He (2021) either assumes i.i.d. signals or does not bound the limit frequencies of actions or beliefs.

[13] For two distributions $P, Q \in \Delta(X)$, the Kullback-Leibler divergence between P and Q is $\sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)}$.

the Kullback-Leibler divergence between observed outcomes under type γ's deviation and those under the honest type's equilibrium strategy is bounded from above by a strictly positive function of η that converges to zero as $\eta \to 0$.

Applying the Azuma-Hoeffding inequality to the log likelihood ratio between the honest type and the opportunistic type,[14] we provide an upper bound on the fraction of bad periods from period 0 to T for every large enough T. Intuitively, this is because when η is small, the honest type keeps their word with high probability conditional on every announcement, which implies that the myopic players have a strict incentive to best reply to any announcement when the log likelihood ratio is above a certain cutoff. We then translate the upper bound on the undiscounted frequency of bad periods to an upper bound on the discounted average frequency of bad periods that converges to 0 as η becomes arbitrarily small. For a sufficiently patient player of type γ, this deviation gives a payoff arbitrarily close to their optimal commitment payoff.

4. Proof of Theorem 1

Fix a Nash equilibrium $(\{\widetilde{\sigma}_\gamma, \sigma_\gamma\}_{\gamma \in \Gamma}, \sigma_2)$, and define random variable ν_t taking values in $\{0, 1\}$ by

1. $\nu_t = 1$ if for every $a \in A$, when player 1 announces a in period t, player 2 strictly prefers one of the actions in $\mathrm{BR}_2(a)$ to all actions that do not belong to $\mathrm{BR}_2(a)$,
2. $\nu_t = 0$ otherwise.

Now define a strategy for type γ, $(\widetilde{\sigma}'_\gamma, \sigma'_\gamma)$, by:

1. At histories where $\nu_t = 1$, type γ announces $a^*(\gamma, \theta_t, \widetilde{A}_t)$ defined in (3.3) upon observing $(\theta_t, \widetilde{A}_t)$ and keeps their word if $a^*(\gamma, \theta_t, \widetilde{A}_t) \in A_t$, and uniformly mixes between all actions in A_t otherwise.
2. At histories where $\nu_t = 0$, type γ plays the honest type's equilibrium strategy, that is, $\widetilde{\sigma}'_\gamma = \widetilde{\sigma}_{\gamma_h}$ and $\sigma'_\gamma = \sigma_{\gamma_h}$ at every such history.

We will bound type γ's payoff from $(\widetilde{\sigma}'_\gamma, \sigma'_\gamma)$ when player 2s uses their equilibrium strategy. As a first step, note that there exists $\xi \in (0, 1)$ that depends only on u_2 such that for every $a \in A$, all actions outside of $\mathrm{BR}_2(a)$ are strictly inferior for player 2 when they believe that player 1 plays a with probability more than ξ. Let $\underline{p} \equiv \min_{a \in A} p(\widetilde{A}_t = \{a\})$, which is strictly positive by Assumption 1. Markov's inequality implies that in every period where $\nu_t = 0$, the probability that $m_t = a_t$ is less than $1 - \underline{p}(1 - \xi)$ conditional on player 1 being opportunistic. Since $G(\widetilde{A}_t = A_t | A_t) \geq 1 - \eta$ for every $A_t \in \mathcal{A}$, the probability that $m_t = a_t$ is at least $1 - \eta$ under $(\widetilde{\sigma}'_\gamma, \sigma'_\gamma)$.

Let $\widetilde{\pi}_t \in (0, 1)$ be the probability player 2_t's belief assigns to the honest type after observes $\{y_0, ..., y_{t-1}\}$ but not m_t. Let $\pi_t \in (0, 1)$ be the probability of honest type after player 2_t observes $\{y_0, ..., y_{t-1}\}$ and m_t. Let $\widetilde{l}_t \equiv \log \frac{\widetilde{\pi}_t}{1 - \widetilde{\pi}_t}$ and $l_t \equiv \log \frac{\pi_t}{1 - \pi_t}$, where we view $\widetilde{\pi}_t$, π_t, $\widetilde{l}_t$, and l_t as random variables. If $\nu_t = 0$, then according to Bayes rule, we have

$$\mathbb{E}\left[\widetilde{l}_{t+1} - \widetilde{l}_t \middle| \widetilde{\sigma}'_\gamma, \sigma'_\gamma\right] \geq D(1 - \eta || 1 - \underline{p}(1 - \xi)) \equiv \alpha,$$

[14] In order to apply the Azuma-Hoeffding inequality, we subtract the expected increment of this process in every period and construct a martingale process under the proposed deviation for type γ.

where $D(x_1||x_2)$ stands for the Kullback-Leibler divergence between a distribution that attaches probability x_1 to $y_t = 1$ and one that attaches probability x_2 to $y_t = 1$.

In every period where $\nu_t = 1$, Assumption 1(ii) and the fact that p has full support together imply the existence of $\rho \in (0, 1)$ that is independent of η such that the probability that $m_t = a_t$ is less than $1 - \rho\eta$ under the equilibrium strategy of the honest type. Since the probability that $m_t = a_t$ is at least $1 - \eta$ under $(\widetilde{\sigma}'_\gamma, \sigma'_\gamma)$,

$$\mathbb{E}\Big[\widetilde{l}_{t+1} - \widetilde{l}_t \Big| \widetilde{\sigma}'_\gamma, \sigma'_\gamma\Big] \geq -D(1-\eta||1-\rho\eta) \equiv -\beta.$$

Therefore, for every $L \in \mathbb{R}_+$, there exists $\overline{\eta} > 0$ such that $\alpha/\beta > L$ when $\eta \in [0, \overline{\eta}]$.

Recall that $\underline{p} \equiv \min_{a \in A} p(\widetilde{A}_t = \{a\}) > 0$. When $\eta \in [0, \overline{\eta}]$ the honest type announces a and takes action a with probability at least $\underline{p}(1-\overline{\eta})$ for every $a \in A$. Therefore, $l_t - \widetilde{l}_t \geq \log \underline{p}(1-\overline{\eta})$. As a result, there exists $l^* \in \mathbb{R}_+$ such that $\widetilde{l}_t \geq l^*$ implies that $\nu_t = 1$.

We establish a lower bound for the expected value of $\sum_{t=0}^{\infty}(1-\delta)\delta^t \nu_t$ when δ is close to 1. Let Z_t be a random variable such that for every $y_t \in \{0, 1\}$,

$$Z_t = \log \frac{\Pr(y_t|\widetilde{\sigma}_{\gamma_h}, \sigma_{\gamma_h})}{\Pr(y_t|\widetilde{\sigma}_{\gamma_o}, \sigma_{\gamma_o})} \text{ with probability } \Pr(y_t|\widetilde{\sigma}'_\gamma, \sigma'_\gamma). \tag{4.1}$$

By definition, $\mathbb{E}[Z_t|\widetilde{\sigma}'_\gamma, \sigma'_\gamma] \geq -\beta$ when $\nu_t = 1$ and $\mathbb{E}[Z_t|\widetilde{\sigma}'_\gamma, \sigma'_\gamma] \geq \alpha$ when $\nu_t = 0$. By construction, $\widetilde{l}_{t+1} = \widetilde{l}_t + Z_t$ for every $t \in \mathbb{N}$.

Lemma 1. *For every $\varepsilon > 0$, there exists $T \in \mathbb{N}$ such that for every $t \geq T$,*

$$\sum_{s=0}^{t-1} \mathbb{E}\Big[\nu_s \Big| \widetilde{\sigma}'_\gamma, \sigma'_\gamma\Big] \geq t\Big(\frac{\alpha}{\alpha+\beta} - \varepsilon\Big). \tag{4.2}$$

Our proof uses the Azuma-Hoeffding inequality.

Azuma-Hoeffding Inequality. *Let $\{X_0, X_1, \cdots\}$ be a martingale such that $|X_k - X_{k-1}| \leq c_k$. For every $n \in \mathbb{N}$ and $\bar{\varepsilon} > 0$,*

$$\Pr\Big(X_n - X_0 \geq \bar{\varepsilon}\Big) \leq \exp\left(\frac{-\bar{\varepsilon}^2}{2\sum_{k=1}^n c_k^2}\right).$$

Proof of Lemma 1. Construct a martingale process $\{\widehat{l}_t\}_{t\in\mathbb{N}}$ recursively from $\{\widetilde{l}_t\}_{t\in\mathbb{N}}$. Let $\widehat{l}_0 \equiv \widetilde{l}_0$, and for every $t \in \mathbb{N}$, let $\widehat{l}_{t+1} \equiv \widehat{l}_t + Z_t - \mathbb{E}[Z_t|\widetilde{\sigma}'_\gamma, \sigma'_\gamma]$. The process $\{\widehat{l}_t\}_{t\in\mathbb{N}}$ is a martingale. Since $\widetilde{l}_{t+1} = \widetilde{l}_t + Z_t$, we have

$$\widehat{l}_t = \widetilde{l}_t - \sum_{s=0}^{t-1} \mathbb{E}\Big[Z_s \Big| \widetilde{\sigma}'_\gamma, \sigma'_\gamma\Big].$$

If $\frac{1}{t}\sum_{s=0}^{t-1} \nu_s \leq \frac{\alpha}{\alpha+\beta} - \varepsilon_1$ for some $\varepsilon_1 > 0$, then

$$\sum_{s=0}^{t-1} \mathbb{E}\Big[Z_s \Big| \widetilde{\sigma}'_\gamma, \sigma'_\gamma\Big] \geq t\varepsilon_1(\alpha+\beta).$$

This is because $\mathbb{E}[Z_t|\widetilde{\sigma}'_\gamma, \sigma'_\gamma] \geq -\beta$ when $\nu_t = 1$ and $\mathbb{E}[Z_t|\widetilde{\sigma}'_\gamma, \sigma'_\gamma] \geq \alpha$ when $\nu_t = 0$.

Applying the Azuma-Hoeffding inequality, we obtain:

$$\Pr(\widetilde{l}_t \le l^* | \widetilde{\sigma}'_\gamma, \sigma'_\gamma) = \Pr\Big(\widehat{l}_t - \widehat{l}_0 \le l^* - \widetilde{l}_0 - t\varepsilon_1(\alpha+\beta) \Big| \widetilde{\sigma}'_\gamma, \sigma'_\gamma\Big) \le \exp\Big(-\frac{(l^* - \widetilde{l}_0 - t\varepsilon_1(\alpha+\beta))^2}{2tC^2}\Big), \tag{4.3}$$

where $C > 0$ is the difference between the largest realization of Z_t and the smallest realization of Z_t. The right-hand-side of (4.3) vanishes to zero exponentially as $t \to +\infty$.

Since $\nu_t = 1$ when $\widetilde{l}_t \ge l^*$, we know that for every $\varepsilon_0 > 0$, there exists $T_0 \in \mathbb{N}$, such that for every $t \ge T_0$, if $\sum_{s=0}^{t-1} \nu_s \le t\Big(\frac{\alpha}{\alpha+\beta} - \varepsilon_1\Big)$, then $\nu_t = 1$ with probability at least $1 - \varepsilon_0$ under $(\widetilde{\sigma}'_\gamma, \sigma'_\gamma)$. By setting $\varepsilon_0 < \frac{\beta}{\alpha+\beta}$, we have $\mathbb{E}[\nu_t|\widetilde{\sigma}'_\gamma, \sigma'_\gamma] \ge 1 - \varepsilon_0 > \frac{\alpha}{\alpha+\beta}$. Then for every $t > T_0$, we have $\mathbb{E}[\sum_{s=0}^{t-1} \nu_s|\widetilde{\sigma}'_\gamma, \sigma'_\gamma] \ge (t - T_0 - 1)(\frac{\alpha}{\alpha+\beta} - \varepsilon_1)$. The conclusion of Lemma 1 follows by choosing any $\varepsilon > \varepsilon_1 > 0$, and $T \ge \frac{\alpha(T_0+1)}{(\varepsilon-\varepsilon_1)(\alpha+\beta)}$. □

Since ν_t is either 0 or 1, we can use summation by parts to obtain:

$$\mathbb{E}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t \nu_t \Big| \widetilde{\sigma}'_\gamma, \sigma'_\gamma\Big] = (1-\delta)^2 \sum_{t=0}^{+\infty} \delta^t \sum_{s=0}^{t} \mathbb{E}\Big[\nu_s \Big| \widetilde{\sigma}'_\gamma, \sigma'_\gamma\Big]. \tag{4.4}$$

Lemma 1 implies that (4.2) applies to every large enough t. Plugging (4.2) into (4.4), we know that for every $\widehat{\varepsilon} > 0$, there exists $\underline{\delta} \in (0, 1)$, such that for every $\delta \in (\underline{\delta}, 1)$, we have

$$\mathbb{E}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t \nu_t \Big| \widetilde{\sigma}'_\gamma, \sigma'_\gamma\Big] \ge \frac{\alpha}{\alpha+\beta} - \widehat{\varepsilon}. \tag{4.5}$$

Since $\frac{\alpha}{\beta} \to +\infty$ as $\eta \to 0$, and type γ's stage-game payoff is at least $\overline{U}^*(\gamma)$ in every period where $\nu_t = 1$, we can find η small enough that each type's payoff from strategy $(\widetilde{\sigma}'_\gamma, \sigma'_\gamma)$ can be an arbitrarily large fraction of $\overline{U}^*(\gamma)$.

5. Extensions & discussion

Multiple types: Theorem 1 extends to any finite number of honest types and opportunistic types, who can potentially have different stage-game payoffs. Let Γ_h be the set of honest types and let Γ_o be the set of opportunistic types, with Γ_h and Γ_o being finite. Let $\Gamma \equiv \Gamma_h \cup \Gamma_o$. For every $\gamma_h \in \Gamma_h$, type γ_h is restricted to choose $m_t \in \widetilde{A}_t$ and to choose $a_t = m_t$ if $m_t \in A_t$. For every $\gamma_o \in \Gamma_o$, type γ_o only faces the restriction that $a_t \in A_t$. Type $\gamma \in \Gamma$'s stage-game payoff is $u_1(\gamma, \theta_t, a_t, b_t)$. Player 2's payoff does not depend on γ and θ.

If we define $U^*(\gamma, \theta_t, A_t)$ and $\overline{U}^*(\gamma)$ in the same way as in (3.1) and (3.2), we can show that under the conditions in Theorem 1, type γ's payoff in every equilibrium is at least $\overline{U}^*(\gamma) - \varepsilon$ for every $\gamma \in \Gamma$. The proof uses a similar argument as that of Theorem 1, except for the construction of $(\widetilde{\sigma}'_\gamma, \sigma'_\gamma)$ at histories where $\nu_t = 0$, which is given by

$$\widetilde{\sigma}'_\gamma \equiv \sum_{\gamma_h \in \Gamma_h} \frac{\widetilde{\pi}_t(\gamma_h)}{\sum_{\gamma'_h \in \Gamma_h} \widetilde{\pi}_t(\gamma'_h)} \widetilde{\sigma}_{\gamma_h}$$

and

$$\sigma'_\gamma \equiv \sum_{\gamma_h \in \Gamma_h} \frac{\pi_t(\gamma_h)}{\sum_{\gamma'_h \in \Gamma_h} \pi_t(\gamma'_h)} \sigma_{\gamma_h}$$

where $\widetilde{\pi}_t \in \Delta(\Gamma)$ is player 2_t's belief after they observe $\{y_0, ..., y_{t-1}\}$ but before they observe m_t, and $\pi_t \in \Delta(\Gamma)$ is player 2_t's belief after they observe both $\{y_0, ..., y_{t-1}\}$ and m_t. The rest of the proof follows from the same step as that of Theorem 1.

Noisy signals of feasible actions: In our baseline model, the patient player knows that actions in $A \backslash \widetilde{A}_t$ will not be feasible in period t. This assumption is not necessary for Theorem 1, which extends to settings where actions in $A \backslash \widetilde{A}_t$ can be feasible at the action stage with small but positive probability (i.e., $G(\widetilde{A}_t = A_t | A_t)$ is close to 1 for every A_t), provided that the honest type continues to only announce actions in $\widetilde{A}_t$.

Noisy observation of past honesty: In many applications of interest, such as retail markets, information about the seller's honesty is passed on to future consumers via word-of-mouth communication, and errors are likely in the process of information transmission.

For this reason it is interesting to note that Theorem 1 extends to settings where y_t is observed with noise. In particular, let $x_t \in X$ be a noisy signal of y_t, distributed according to $F(\cdot|y_t) \in \Delta(X)$. We assume that X is finite and moreover, $F(\cdot|y = 1) \neq F(\cdot|y = 0)$, that is, x_t can statistically identify y_t. Corollary 1 generalizes Theorem 1 when player 2_t observes $\{x_0, ..., x_{t-1}\}$ instead of $\{y_0, ..., y_{t-1}\}$

Corollary 1. *Suppose the distributions p and G satisfy Assumption 1 and x_t can statistically identify y_t. For every $\varepsilon > 0$, there exist $\underline{\delta} \in (0, 1)$ and $\eta > 0$ such that when $\delta > \underline{\delta}$ and $G(\widetilde{A}_t = A_t | A_t) \geq 1 - \eta$ for every $A_t \in \mathcal{A}$, then each type γ receives payoff at least $\overline{U}^*(\gamma) - \varepsilon$ in every equilibrium.*

The proof resembles that of Theorem 1, so it is omitted. Intuitively, observing a noisy signal of y_t reduces the responsiveness of player 2's posterior belief with respect to their observations. In the proof of Theorem 1, the monitoring noise reduces the absolute value of $\mathbb{E}[\widetilde{l}_{t+1} - \widetilde{l}_t]$ both in good periods (i.e., periods where $\nu_t = 1$) and in bad periods (i.e., periods where $\nu_t = 0$). Nevertheless, when η is small enough, the ratio between the expected increase in $\widetilde{l}_t$ during bad periods and the expected decrease in $\widetilde{l}_t$ during good periods remains large. The same argument implies that the fraction of bad periods is small and the patient player can secure their optimal commitment payoff in every equilibrium.

Announcing the state & observability of the state In some applications, the patient player also announces the state θ_t in addition to their intended action a_t, and the state can be observed by the future short-run players. The honest type announces the state truthfully and plays their announced action whenever it is available.

Our main result applies in this setting. Intuitively, suppose type γ uses the deviation we constructed in the proof of Theorem 1 to announce and to take actions, and to truthfully report the state in every period. Since truthfully announcing the state can never decrease the log likelihood ratio between the honest type and the opportunistic type, one can use the same argument as the proof of Theorem 1 to show that every type of patient player receives at least their optimal commitment payoff.

Announcing the state is different from announcing an intended action, since the true state is fixed when player 1 makes their announcement while player 1's action is not fixed. As a result, there are multiple ways in which player 1's announced action matches their realized action (e.g., announcing any available action and playing it afterward), while there is only one way in which they can announce the true state. As a result, truthfully announcing the state cannot lead to a decrease in the patient player's reputation, while announcing an intended action and playing the announced action may lead to a decrease in reputation, since in equilibrium, the honest type may announce that action with lower probability than the opportunistic type does.

Bounded observation of past actions and announcements: Our baseline model excludes the possibility that player 2 observes player 1's past actions and announcements in addition to whether they coincide. We extend Theorem 1 so that the player 2s can also observe a noisy signal z_t about a_t and m_t, in addition to y_t, as long as each of them can only observe the realizations of z_t in a bounded number periods.

Formally, let $z_t \in Z$, where z_t is distributed according to $H(\cdot|m_t, a_t) \in \Delta(Z)$, where Z is a finite set. Suppose for every $t \in \mathbb{N}$, player 2_t can observe player 1's announcement m_t, the history of whether player 1 has kept their word $\{y_0, ..., y_{t-1}\}$ and a (possibly stochastic) subset of $\{z_0, ..., z_{t-1}\}$ that has at most K elements, with $K \in \mathbb{N}$ an exogenous parameter.

Our assumption on the asymmetry between player 2s' observations of y_t and z_t is motivated by retail markets in developing economies, or more generally, markets without well developed recording-keeping institutions. In those markets, detailed information about sellers' actions and announcements (e.g., the quality of their services, various attributes of their products, the content of their advertisements, and so on, which correspond to z_t) is likely to get lost over time. By contrast, simple coarse information about sellers' records, such as whether they have kept their word (which corresponds to y_t), is likely to be more persistent.[15]

Corollary 2. *Suppose Assumption 1 is satisfied, each player 2 observes at most K realizations of z, and at least one of the following two conditions is satisfied:*

1. $H(\cdot|a, m)$ *has full support for every* $(a, m) \in A \times A$*,*
2. $G(\widetilde{A}_t = A_t|A_t) = 1$ *for every* $A_t \in \mathcal{A}$*,*

then for every $\varepsilon > 0$, there exist $\underline{\delta} \in (0, 1)$ and $\eta > 0$ such that when $\delta > \underline{\delta}$ and $G(\widetilde{A}_t = A_t|A_t) \geq 1 - \eta$ for every $A_t \in \mathcal{A}$, then each type γ receives payoff at least $\overline{U}^*(\gamma) - \varepsilon$ in every equilibrium.

Intuitively, the honest type announces every $a \in A$ with positive probability. Therefore, if any type of the patient player adopts the deviation in the proof of Theorem 1, no realization of z can rule out the honest type, so the decrease in the log likelihood ratio between the honest type and

[15] This justification only applies to situations in which the patient player's optimal commitment action is different in different states.

the opportunistic type is bounded from above for every $z \in Z$.[16] If player 2 observes at most K realizations of z, then the decrease in the log likelihood ratio must also be bounded from above.

Let $\widetilde{l}_t$ be the log likelihood ratio between the honest type and the opportunistic type after player 2_t observes $\{y_0, ..., y_{t-1}\}$ but before they observe m_t and the past realizations of z. Both m and z being boundedly informative implies the existence of $l^* > 0$ such that player 2_t has a strict incentive to best reply to any announcement when $\widetilde{l}_t > l^*$. The rest of the proof follows from that of Theorem 1.

Unbounded observation of past actions and announcements: We show that there are equilibria where the opportunistic type's payoff is bounded below their commitment payoff when player 2's can observe the entire history of player 1's past actions and announcements. The intuition is that the honest type and the opportunistic type can have different stage-game payoff functions, so the opportunistic type may receive a low payoff when they play the honest type's equilibrium strategy. To illustrate, consider an example where Θ is a singleton and players' stage-game payoffs are given by:

$\gamma = \gamma_o$	T	N
H	2, 1	0, 0
L	3, 0	1, 1

$\gamma = \gamma_h$	T	N
H	0, 1	0, 0
L	0, 0	0, 1

These payoff functions satisfy the assumptions in our paper since player 2's payoff does not depend on player 1's type. We assume that $G(\widetilde{A}_t = A_t|A_t) = 1$ for every $A_t \subset \{H, L\}$, that is, player 1 perfectly observes the set of feasible actions before making their announcement. The distribution of A_t is such that $A_t = \{H, L\}$ with probability $1 - \varepsilon$, $A_t = \{H\}$ with probability $\frac{\varepsilon}{2}$, and $A_t = \{L\}$ with probability $\frac{\varepsilon}{2}$. Throughout, we fix an $\varepsilon \in (0, 2/5)$.

Such a distribution of $(A_t, \widetilde{A}_t)$ satisfies Assumption 1 in our paper. The opportunistic-type player 1's commitment payoff equals

$$\frac{\varepsilon}{2} + (1 - \frac{\varepsilon}{2})2 = 2 - \frac{\varepsilon}{2}.$$

This payoff can be obtained if player 1 commits to play H when H is feasible, and commits to play L when H is not feasible.

Theorem 2. *Suppose player 2s' prior belief attaches probability no more than $\frac{1}{2}$ to the honest type. There exists $\underline{\delta} \in (0, 1)$ such that for every $\delta > \underline{\delta}$, there exists an equilibrium where the opportunistic type's payoff equals $\frac{7}{4}$.*

We provide a constructive proof of Theorem 2 in Appendix A. The comparison between Theorem 2 and Theorem 1 implies that allowing the short-run players to receive more information can hurt the patient player's incentives to build reputations for honesty.

[16] This requires either the distribution of z to have full support, or player 1 to perfectly observe the set of available actions when making announcements. Intuitively, if $G(\widetilde{A}_t = A_t|A_t) \neq 1$ and the distribution of z does not have full support, e.g., $z_t = (a_t, m_t)$, then suppose the honest type announces a^* if and only if $\widetilde{A}_t = \{a^*\}$, and the opportunistic type's optimal commitment action in $\widetilde{A}_t = \{a^*, a'\}$ is a^*, then under the deviation we construct, the opportunistic type announces a^* after observing $\widetilde{A}_t = \{a^*, a'\}$ and plays a' when $A_t = \{a'\}$. If the distribution of z_t does not have full support, then some realizations of z_t may occur with positive probability if and only if $(a_t, m_t) = (a', a^*)$, which does not occur under the honest type's equilibrium strategy so the opportunistic type fully reveals their type with positive probability through z under the deviation we construct.

Short-run players' payoffs depending on the state: Our baseline model assumes that the short-run players' payoffs do not depend on θ. Our reputation result fails when u_2 is a function of θ. For example, suppose players' payoffs are given by:

θ_1	T	N
H	$2, 2$	$0, 0$
L	$3, -3$	$0, 0$

θ_2	T	N
H	$3, -3$	$0, 0$
L	$2, 2$	$0, 0$

The two states θ_1 and θ_2 are equally likely. In state θ_1, the Stackelberg action is H and the Stackelberg payoff is 2 for both types of the long-run player. In state θ_2, the Stackelberg action is L and the Stackelberg payoff is 2 for both types of the long-run player. The expected Stackelberg payoff is thus 2.

Suppose $A_t = \widetilde{A}_t$ with probability 1, i.e., player 1 always knows the set of feasible actions at the announcement stage. In each period, with probability $1 - \varepsilon$, player 1 chooses from the set $\{H, L\}$, with probability $\varepsilon/2$ they must choose H, and with probability $\varepsilon/2$ they must choose L. This environment satisfies our other assumptions except that player 2's payoff depends on θ. Now consider the following strategy profile:

- Short-run players play N at every history and never revise beliefs about the long-run player's type.
- Both types of the long-run player announce and play L whenever it is feasible, and announce and play H otherwise.

The short-run players are always playing their myopic best responses, since the long-run player's announcements reveal no information about the state, so N is a best response to both H and L. Since the actions of the two types of the long-run player coincide, the short-run players never learn anything about the long-run player's type. Both types of the long-run player are playing best responses since short-run players always play N. Hence, we have an equilibrium in which the payoffs of both types of the long-run player are bounded below their Stackelberg payoffs even when $\varepsilon \to 0$ and $\delta \to 1$.

6. Related literature

Our paper contributes to the reputation literature by showing that the patient player can secure their optimal commitment payoff in all Nash equilibria when every type's behavior is endogenous. This contrasts to reputation models in which at least one type is committed to an exogenous strategy (Sobel, 1985; Fudenberg and Levine, 1989, 1992; Benabou and Laroque, 1992; Mathevet et al., 2019; Gossner, 2011; Ekmekci et al., 2012), as well as other reputation models without commitment types that focus on Markov equilibria (Schmidt, 1993; Daley and Green, 2012; Board and Meyer-ter-Vehn, 2013),[17] or establishing folk theorems (Pei, 2020, 2021). Our model also differs from those that posit that an exogenous value for a reputation, e.g. Olszewski (2004) and Ottaviani and Sørensen (2006).

Our work is related to the literature on repeated communication games e.g. Sobel (1985), Benabou and Laroque (1992), Best and Quigley (2020), Mathevet et al. (2019), and Pei (2020).

[17] Section 3 of Ely and Valimaki (2003) studies reputation models without commitment types and shows that reputation concerns can generate perverse incentives that lead to low-payoff equilibria for the long-run player.

In those papers, a sender communicates with a receiver about a payoff-relevant state, which stands in contrast to our model where the sender communicates their intended action. As we have commented in Section 6, communicating the payoff-relevant state and allowing the short-run players to observe the state ex post do not affect our conclusions.

Our model is also related to the literature on reputational bargaining, e.g. Kambe (1999), Abreu and Gul (2000), Abreu and Pearce (2007), Bagwell (2018), Kim (2009), and Sanktjohanser (2020), in which players announce their bargaining postures in the beginning of the game and decide when to concede to their opponents' offers. In contrast to those papers, the honest type's announcement in our model is only valid for only one period; they are free to make any announcement in the future.

The fact that many people prefer to be honest has been established experimentally by e.g. Gneezy (2005), Charness and Dufwenberg (2006) and Gneezy et al. (2018). Kartik et al. (2007) and Kartik (2009) show how costs of lying change the equilibrium outcomes of strategic communication games. Instead of positing that some players have a cost of lying, we follow Chen et al. (2008) and Chen (2011) and assume that the patient player is either an honest type who never lies, or an opportunistic type who faces no cost of lying. Since Theorem 1 allows for different types to have different preferences as well as any finite number of types, it extends to cases with strictly positive and possibly heterogeneous lying costs.

Jullien and Park (2020) studies repeated buyer-seller games in which a seller privately observes their product quality, which is a noisy signal of their effort, and shows that communication about quality improves the maximum social welfare if and only if the seller's cost of effort is intermediate.[18] Our paper examines whether a patient player can guarantee high payoffs in *all* equilibria by building reputations for honesty. Successful reputation building in our model hinges on the patient player's knowledge about their feasible action set when making announcements, but does not depend on the players' payoff functions. In a working paper version (Fudenberg et al., 2020), we extend our analysis to Jullien and Park (2020)'s setting where the seller announces a private signal of their effort before buyers act. We establish a reputation result when the signal has full support and show that reputation fails when the signal perfectly reveals the agent's effort.

Our requirement that the feasible action set is stochastic is related to Celentani et al. (1996), and Atakan and Ekmekci (2015), which show that full support monitoring can help reputation building when the uninformed player is long-lived. Their results, unlike ours, require that the informed player cannot perfectly observe the uninformed player's actions.

7. Conclusion

This paper provides sufficient conditions under which a patient player can obtain a high payoff by building a reputation for honestly announcing their intended actions, rather than for playing particular actions. We establish a reputation result when the uninformed players can observe whether the reputation-building player has kept their word in the past, and face uncertainty about which of the reputation-building player's actions are feasible.

[18] Jullien and Park (2014) shows that communication accelerates consumer learning when product quality is determined by the seller's type, and the high type seller is non-strategic and always tells the truth. Awaya and Krishna (2016) identify a class of games in which players can achieve perfectly collusive payoffs with communication, but not without it.

Appendix A. Proof of Theorem 2

We construct a class of equilibria where the opportunistic type's payoff equals $v^* \equiv \frac{7}{4}$, which is strictly lower than the opportunistic type's optimal commitment payoff $2 - \frac{\varepsilon}{2}$ when ε is small enough. As a convention, we say that type γ plays (a, a') at history h^t if he announces action a and takes action a' at h^t.

State variables: Our construction keeps track of two state variables:

1. Let $l(h^t)$ be the *log likelihood ratio* between the honest type and the opportunistic type under player 2's belief at h^t after they observe $\{a_0, ..., a_{t-1}, m_0, ..., m_{t-1}\}$ but before they observe player 1's period t announcement. Let $L(h^t) \equiv e^{l(h^t)}$ be the likelihood ratio at h^t.
2. Let $v(h^t)$ denote the opportunistic type's continuation value at h^t.

The initial values of these state variables are $v(h^0) = \frac{7}{4}$ and $l(h^0) = \log \frac{\pi_0}{1-\pi_0} \leq 0$.

Let

$$\underline{v} \equiv \frac{2+\delta}{1+\delta}, \tag{A.1}$$

which is strictly between 1 and $\frac{7}{4}$ when δ is close to 1.

In equilibrium, learning takes place if and only if $l(h^t) \notin \{-\infty, +\infty\}$, and learning stops if $l(h^t) \in \{-\infty, +\infty\}$. The set of histories where learning takes place is partitioned into two classes: histories where $v(h^t) > \underline{v}$ and histories where $v(h^t) \leq \underline{v}$. We will verify later that at every history where $l(h^t) \notin \{-\infty, +\infty\}$ and $v(h^t) \leq \underline{v}$, it must be the case that $l(h^t) \leq 0$.

We claim that by our construction, at every history at which active learning takes place,

$$\frac{\underline{v} - 2(1-\delta)}{\delta} \leq v(h^t) \leq 2 - \frac{\varepsilon}{2}, \tag{A.2}$$

which implies, when $\delta > 1 - \frac{\varepsilon}{7}$, that the opportunistic type's continuation value is between $\underline{v} - \frac{\varepsilon}{2}$ and $2 - \frac{\varepsilon}{2}$ once learning stops. The upper bound holds if, at every history where the opportunistic type plays (H, L) with positive probability, they are indifferent between playing (H, L) and playing either (H, H) or (L, L), at least one of which yields a stage-game payoff less than $2 - \frac{\varepsilon}{2}$. Since $v^* > \underline{v}$ for large enough δ, the lower bound holds if in every active learning period, either $v(h^t) \geq v(h^{t-1})$, or $v(h^{t-1}) \geq \underline{v}$ and $v(h^t) \geq \frac{v(h^{t-1})-2(1-\delta)}{\delta}$.

We first describe players' strategies when $l_t = -\infty$ by showing that every payoff $v \in [1 - \frac{\varepsilon}{2}, 2 - \frac{\varepsilon}{2}]$ can be attained in some equilibrium of the game where player knows that player 1 is the opportunistic type. Then we describe players' strategies when active learning takes place.

No reputation phase: We show that there exists $\underline{\delta} \in (0, 1)$ such that for every $\delta > \underline{\delta}$ and $v \in [1 - \varepsilon, 2 - \frac{\varepsilon}{2}]$, there exists an equilibrium where the opportunistic type's continuation value is v in the repeated complete information game where player 1 is known to be the opportunistic type and their discount factor is δ.

According to Fudenberg and Maskin (1991), for every $\eta > 0$, there exists $\underline{\delta} \in (0, 1)$ such that when $\delta > \underline{\delta}$, there exists a sequence $v_0^*, v_1^*, ..., v_n^* .... \in \{1 - \frac{\varepsilon}{2}, 2 - \frac{\varepsilon}{2}\}$ such that

$$v = (1-\delta) \sum_{t=0}^{+\infty} \delta^t v_t^*,$$

and for every $s \in \mathbb{N}$,

$$(1-\delta)\sum_{t=s}^{+\infty}\delta^{t-s}v_t^* \in (v-\eta, v+\eta).$$

Next, the opportunistic type's payoff is $1-\frac{\varepsilon}{2}$ in the following equilibrium: Player 1 announces L regardless of the realization of A_t, player 1 plays L as long as $L \in A_t$ and player 2 plays N. Hence, for every $v \in [1-\varepsilon, 2-\frac{\varepsilon}{2}]$, player 1's continuation value is v in an equilibrium where:

1. For every t such that $v_t^* = 2-\frac{\varepsilon}{2}$, player 1 announces H when $H \in A_t$ and announces L otherwise, and takes an action that coincides with their announcement. Player 2 best replies to player 1's announcement. If player 1 plays an action different from their announcement, then play the continuation equilibrium where player 1's payoff is $1-\frac{\varepsilon}{2}$.
2. For every t such that $v_t^* = 1-\frac{\varepsilon}{2}$, player 1 announces L no matter what and plays L no matter what. Player 2 plays N regardless of player 1's announcement.

In what follows, we assume that $\delta > \max\{\underline{\delta}, 1-\varepsilon/7\}$.

Active learning phase: high continuation values If h^t is such that $v(h^t) > \underline{v}$, then there are two subcases. If $v(h^t)$ is such that

$$\frac{v(h^t)-(1-\delta)}{\delta} > 2-\frac{\varepsilon}{2}, \tag{A.3}$$

The honest type plays (L, L) with probability $\frac{3}{4}$ and plays (H, H) with probability $\frac{1}{4}$. The opportunistic type plays (H, L) with probability $\frac{1}{2}$ and plays (H, H) with probability $\frac{1}{2}$. Player 2 best replies to player 1's announced action.

1. After observing (L, L) at h^t, the posterior log likelihood ratio is $l_{t+1} = +\infty$, and starting from period $t+1$, player 2 best replies to player 1's announced action in every future period until they have observed player 1 not keeping their word (after which they play N in all future periods). Hence, the opportunistic type's continuation value after playing (L, L) is $2-\frac{\varepsilon}{2}$. Inequality (A.3) implies that the opportunistic type has no incentive to play (L, L) at h^t.
2. After observing (H, L) at h^t, the posterior log likelihood ratio is $l_{t+1} = -\infty$ and the opportunistic type's continuation value in period $t+1$ equals $v(h^{t+1}) \equiv \frac{v(h^t)-3(1-\delta)}{\delta}$. Since $v(h^t) < 2-\frac{\varepsilon}{2}$, we have $1 < v(h^{t+1}) < v(h^t) < 2-\frac{\varepsilon}{2}$, which means that $v(h^{t+1})$ can be delivered by some continuation equilibrium after player 2 knows that player 1 is opportunistic.
3. After observing (H, H) at h^t, player 1's continuation value in period $t+1$ is $v(h^{t+1}) \equiv \frac{v(h^t)-2(1-\delta)}{\delta}$, which is strictly less than $v(h^t)$ since $v(h^t) < 2-\frac{\varepsilon}{2}$. The posterior likelihood ratio after observing (H, H) is $l(h^t) - \log 2$.

If $v(h^t)$ is such that

$$\frac{v(h^t)-(1-\delta)}{\delta} \leq 2-\frac{\varepsilon}{2},$$

then player 2 best replies to player 1's announced action. The honest type plays (L, L) with probability $1-\frac{\varepsilon}{2}$ and plays (H, H) with probability $\frac{\varepsilon}{2}$. The opportunistic type plays (H, H) with probability ε and plays (L, L) with probability $1-\varepsilon$. Player 1's continuation value in period $t+1$ is $\frac{v(h^t)-(1-\delta)}{\delta}$ after playing (L, L) at h^t and is $\frac{v(h^t)-2(1-\delta)}{\delta}$ after playing (H, H) at h^t. The log likelihood ratio is $l(h^t) + \log \frac{1-\frac{\varepsilon}{2}}{1-\varepsilon}$ after playing (L, L) and is $l(h^t) - \log 2$ after playing (H, H).

Active learning phase: low continuation values If h^t is such that $v(h^t) \leq \underline{v}$, then player 2 plays N when player 1 announces L and plays T with probability $\frac{v(h^t)}{2}$ when player 1 announces H. The honest type plays (L, L) with probability $1 - \frac{\varepsilon}{2}$ and plays (H, H) with probability $\frac{\varepsilon}{2}$. The opportunistic type plays (H, H) with probability $\frac{\varepsilon}{2}$, plays (H, L) with probability $(L(h^t) + 1)\frac{\varepsilon}{2}$, and plays (L, L) with complementary probability. One can verify that as long as $l(h^t) \leq 0$, i.e., $L(h^t) \leq 1$, the above mixed strategy is feasible and player 2 is indifferent between T and N when player 1 announces H.

By construction, if (H, H) is played at h^t, then $l(h^{t+1}) = l(h^t)$ and the opportunistic type stage-game payoff equals their continuation value at h^t, and let their continuation value in period $t+1$ to equal $v(h^t)$. After playing (L, L),

$$l(h^{t+1}) = l(h^t) + \log \frac{1-\varepsilon/2}{1-(L(h^t)+2)\varepsilon/2}$$

and the opportunistic type's continuation value is $\frac{v(h^t)-(1-\delta)}{\delta}$.

After playing (H, L), $l(h^{t+1}) = -\infty$ and the opportunistic type's continuation value is

$$\frac{v(h^t) - (1-\delta)(2v(h^t)+1)}{\delta}.$$

The above payoff is between $\frac{3}{2} - \varepsilon$ and $2 - \varepsilon$ when $\delta > 1 - \frac{\varepsilon}{7}$, which means that it can be delivered by an equilibrium when player 2 knows that player 1 is opportunistic. With these continuation values, the opportunistic type is indifferent between playing (L, L) and playing (H, L).

Verifying feasibility of opportunistic type's equilibrium strategy: By construction (A.2) holds at every history where active learning takes place, and since $\delta > 1 - \frac{\varepsilon}{7}$, the opportunistic type's continuation value is between $\underline{v} - \frac{\varepsilon}{2}$ and $2 - \frac{\varepsilon}{2}$ once learning stops.

In order to show that player 1's mixed strategy in the low continuation value phase is well-defined, we show that $l(h^t)$ is less than 0 at every history where $l(h^t)$ is finite and $v(h^t) \leq \underline{v}$. This is sufficient for the existence of a mixed strategy for the opportunistic type that makes player 2 indifferent between N and T after player 1 announces H.

Learning takes place in the next period only if player 1 plays (H, H) or (L, L) so it is sufficient to keep track of histories in which either (H, H) or (L, L) is played. In particular, player 2's beliefs and the opportunistic type's continuation value are the same when (H, H) is played at a low continuation value history, so every history in which (H, H) or (L, L) is played is equivalent to a shortened history in which all instances of (H, H) played in a low continuation value period are removed. In such a shortened history, whenever (L, L) is played, the opportunistic type gets a within-period payoff of 1, while when (H, H) is played, the payoff is 2.

Let $\{\xi_t\}_{t \in \mathbb{N}}$ be such that $\xi_t \in \{H, L\}$. Let

$$u(\xi_t) \equiv \begin{cases} 2 & \text{if } \xi_t = H \\ 1 & \text{if } \xi_t = L \end{cases} \tag{A.4}$$

be the stage-game payoffs in periods where active learning takes place and H or L are announced. For a given finite sequence $\boldsymbol{\xi} \equiv \{\xi_0, ..., \xi_s\}$, let $N_L(\boldsymbol{\xi})$ be the number of L in this sequence and let $N_H(\boldsymbol{\xi})$ be the number of H in this sequence.

Recall that the opportunistic type's discounted average payoff is $v^* > \underline{v} \equiv \frac{2+\delta}{1+\delta}$. If player 1 receives stage-game payoff $u(\xi_t)$ in period t for every $t \in \{0, 1, ..., s\}$, then player 1's continuation value in period $s+1$, denoted by v_{s+1} satisfies

$$v^* = \sum_{t=0}^{s}(1-\delta)\delta^t u(\xi_t) + \delta^{s+1} v_{s+1},$$

Lemma 2. *Suppose δ is large enough such that $(1-\delta)2 + \delta\underline{v} < v^*$. For every $s \in \mathbb{N}$ and every finite sequence $\{\xi_0, ..., \xi_s\}$. If*

$$v^* - \sum_{t=0}^{s}(1-\delta)\delta^t u(\xi_t) \leq \delta^{s+1}\underline{v} \tag{A.5}$$

$$v^* - \sum_{t=0}^{k}(1-\delta)\delta^t u(\xi_t) > \delta^{k+1}\frac{\underline{v} - 2(1-\delta)}{\delta} \quad \text{for every} \quad k \leq s, \tag{A.6}$$

then

$$N_L(\boldsymbol{\xi}) < N_H(\boldsymbol{\xi}). \tag{A.7}$$

The payoff bound in (A.2) implies that (A.6) must hold at any s at which active learning is still ongoing, and (A.5) holds at every low continuation value history.

Proof. We show that for every finite sequence $\boldsymbol{\xi}$ satisfying $N_L(\boldsymbol{\xi}) \geq N_H(\boldsymbol{\xi})$, $\boldsymbol{\xi}$ cannot satisfy both (A.5) and (A.6). We show by strong induction on $N_L(\boldsymbol{\xi})$. For every $\boldsymbol{\xi}$ such that $N_L(\boldsymbol{\xi}) = 1$ and $N_L(\boldsymbol{\xi}) \geq N_H(\boldsymbol{\xi})$, inequality (A.5) is violated since $(1-\delta)2 + \delta\underline{v} < v^*$ and there can be at most one H in such a sequence given that $N_L(\boldsymbol{\xi}) \geq N_H(\boldsymbol{\xi})$.

Suppose for every $\boldsymbol{\xi}$ such that $N_L(\boldsymbol{\xi}) \leq N$ and $N_L(\boldsymbol{\xi}) \geq N_H(\boldsymbol{\xi})$, $\boldsymbol{\xi}$ violates either (A.5) or (A.6). Suppose by way of contradiction that there exists $\hat{\boldsymbol{\xi}} \equiv \{\hat{\xi}_0, ..., \hat{\xi}_s\}$ such that $N_L(\hat{\boldsymbol{\xi}}) = N+1$, $N_L(\hat{\boldsymbol{\xi}}) \geq N_H(\hat{\boldsymbol{\xi}})$, and $\hat{\boldsymbol{\xi}}$ satisfies both (A.5) and (A.6). We consider two cases, depending on the largest $t \in \{0, 1, ..., s\}$ such that $\hat{\xi}_t = L$, which we denote by $\hat{t}$.

1. Suppose $\hat{t} = s$, i.e., L appears in the last period of sequence $\hat{\boldsymbol{\xi}}$. Then the presumption that $\hat{\boldsymbol{\xi}}$ satisfies (A.6) implies that the continuation value in period $s-1$ is strictly more than $\frac{\underline{v}-2(1-\delta)}{\delta}$, and $\hat{\boldsymbol{\xi}}$ satisfies (A.5) implies that the continuation value in period s is weakly less than $\underline{v}$. Hence
$$\frac{\underline{v} - 2(1-\delta)}{\delta} < \hat{v}_t = (1-\delta)u(\hat{\xi}_s) + \delta\hat{v}_{t+1} \leq (1-\delta) + \delta\underline{v}. \tag{A.8}$$
Plugging in the expression that $\underline{v} \equiv \frac{2+\delta}{1+\delta}$, we have
$$\frac{\underline{v} - 2(1-\delta)}{\delta} = (1-\delta) + \delta\underline{v}.$$
This leads to a contradiction.
2. Suppose $\hat{t} < s$, i.e., L does not appear in the last period of sequence $\hat{\boldsymbol{\xi}}$. By definition, $\hat{\xi}_{\hat{t}+1} = H$. Consider another sequence $\boldsymbol{\xi}' \equiv \{\xi'_0, ..., \xi'_{s-2}\}$ defined as $\xi'_t \equiv \hat{\xi}_t$ for every $t \leq \hat{t} - 1$, and

$\xi'_t \equiv \hat{\xi}_{t+2}$ for every $t \geq \hat{t}$. Since $N_L(\hat{\boldsymbol{\xi}}) = N+1$ and $N_L(\hat{\boldsymbol{\xi}}) \geq N_H(\hat{\boldsymbol{\xi}})$, we have $N_L(\boldsymbol{\xi}') = N$ and $N_L(\boldsymbol{\xi}') \geq N_H(\boldsymbol{\xi}')$. Since $\hat{\boldsymbol{\xi}}$ satisfies (A.5) and (A.6), we have

$$v^* \leq \left\{\sum_{t=0}^{\hat{t}-1}(1-\delta)\delta^t u(\hat{\xi}_t)\right\} + (1-\delta)\{\delta^{\hat{t}} + 2\delta^{\hat{t}+1}\} + \left\{\sum_{t=\hat{t}+2}^{s}(1-\delta)\delta^t u(\hat{\xi}_t)\right\} + \delta^{s+1}\underline{v}$$

and for every $\hat{t}+2 \leq k \leq s$,

$$v^* > \left\{\sum_{t=0}^{\hat{t}-1}(1-\delta)\delta^t u(\hat{\xi}_t)\right\} + (1-\delta)\{\delta^{\hat{t}} + 2\delta^{\hat{t}+1}\} + \left\{\sum_{t=\hat{t}+2}^{k}(1-\delta)\delta^t u(\hat{\xi}_t)\right\} + \delta^{s+1}\frac{\underline{v}-2(1-\delta)}{\delta}.$$

Additionally, since $\hat{\xi}_{\hat{t}} = L$ and $N = N_L(\hat{\xi}_1, \ldots, \hat{\xi}_{\hat{t}-1}) \geq N_H(\hat{\xi}_1, \ldots, \hat{\xi}_{\hat{t}-1})$, and (A.6) must hold because $t-1 < s$,

$$v^* > \left\{\sum_{t=0}^{\hat{t}-1}(1-\delta)\delta^t u(\hat{\xi}_t)\right\} + \delta^{\hat{t}+1}\underline{v},$$

since otherwise, $\hat{\xi}_1, \ldots, \hat{\xi}_{\hat{t}-1}$ would contradict the inductive hypothesis. Since

$$\frac{\underline{v}-2(1-\delta)}{\delta} = \frac{1+2\delta}{1+\delta} < \frac{2+\delta}{1+\delta} = \underline{v},$$

we have:

$$\begin{aligned}
v^* &\leq \sum_{t=0}^{\hat{t}+1}(1-\delta)\delta^t u(\hat{\xi}_t) + \sum_{t=\hat{t}}^{s-2}(1-\delta)\delta^t u(\hat{\xi}_{t+2}) + \delta^{s-1}\underline{v} + (1-\delta^2)\left(\frac{1+2\delta}{1+\delta}\delta^{\hat{t}} - \delta^{s-1}\underline{v}\right) \\
&\quad - (1-\delta^2)\sum_{t=\hat{t}+1}^{s-2}(1-\delta)\delta^t u(\hat{\xi}_{t+2}) \\
&\leq \sum_{t=0}^{\hat{t}}(1-\delta)\delta^t u(\hat{\xi}_t) + \sum_{t=\hat{t}}^{s-2}(1-\delta)\delta^t u(\hat{\xi}_{t+2}) + \delta^{s-1}\underline{v} + (1-\delta^2)(\delta^{\hat{t}} - \delta^{s-1})\underline{v} \\
&\quad - (1-\delta^2)(\delta^{\hat{t}} - \delta^{s-1})\underline{v} \\
&\leq \left\{\sum_{t=0}^{\hat{t}-1}(1-\delta)\delta^t u(\hat{\xi}_t)\right\} + \left\{\sum_{t=\hat{t}}^{s-2}(1-\delta)\delta^t u(\hat{\xi}_{t+2})\right\} + \delta^{s-1}\underline{v}.
\end{aligned} \tag{A.9}$$

Since $v(h^{\hat{t}}) > \underline{v}$, by the construction of ξ', either $\xi'_t = H$ or $\xi'_t = L$ and the continuation value is greater than $\underline{v}$; therefore, the lower bound on continuation values in (A.2) continues to hold, and

$$v^* > \sum_{t=0}^{\hat{t}+1}(1-\delta)\delta^t u(\hat{\xi}_t) + \sum_{t=\hat{t}}^{k-2}(1-\delta)\delta^t u(\hat{\xi}_{t+2}) + \delta^{k-1}.\frac{\underline{v}-2(1-\delta)}{\delta} \tag{A.10}$$

That is, $\boldsymbol{\xi}'$ satisfies both (A.5) and (A.6). This contradicts our induction hypothesis since $N_L(\boldsymbol{\xi}') = N$ and $N_L(\boldsymbol{\xi}') \geq N_H(\boldsymbol{\xi}')$.

The two parts together imply that for every $\boldsymbol{\xi}$ such that $N_L(\boldsymbol{\xi}) \geq N_H(\boldsymbol{\xi})$, $\boldsymbol{\xi}$ cannot satisfy both (A.5) and (A.6). So in order for $\boldsymbol{\xi}$ to satisfy both (A.5) and (A.6), we need $N_L(\boldsymbol{\xi}) < N_H(\boldsymbol{\xi})$. □

The change in the log-likelihood ratio in each active learning period in which (H, H) is played is 0 if $v(h^t) \leq \underline{v}$ and $-\log(2)$ if $v(h^t) > \underline{v}$. In the active-learning periods in which (L, L) is played, if $v(h^t) > \underline{v}$ it increments by $\log \frac{1-\frac{\varepsilon}{2}}{1-\varepsilon}$, and if $v(h^t) \leq \underline{v}$, then if $L(h^t) \leq 1$, the log-likelihood ratio increments by at most $\log \frac{1-\frac{\varepsilon}{2}}{1-\frac{3\varepsilon}{2}}$. Let

$$\Delta(\xi_t) \equiv \begin{cases} -\log 2 & \text{if } \xi_t = H \\ \log \frac{1-\frac{\varepsilon}{2}}{1-\frac{3\varepsilon}{2}} & \text{if } \xi_t = L. \end{cases} \tag{A.11}$$

By definition, this is an upper bound on the change in the log likelihood ratio if $l(h^t) \leq 0$. Therefore, if $l_t \leq 0$ for every period $t \leq s$ in which $v(h^t) \leq \underline{v}$, then player 2's posterior log likelihood ratio in period $s+1$, denoted by l_{s+1} satisfies:

$$l_{s+1} \leq l_0 + \sum_{t=0}^{s} \Delta(\xi_t).$$

Indeed, $l_0 \leq 0$, so by Lemma 2 and by induction on periods in which $v(h^t) \leq \underline{v}$, we have

$$l_0 + \sum_{t=0}^{s} \Delta(\xi_s) \leq 0. \tag{A.12}$$

Hence $l(h^t) \leq 0$ at every history h^t such that $l(h^t)$ is finite and $v(h^t) \leq \underline{v}$. Thus, the constructed strategy profile is feasible for all $\varepsilon < 2/5$ and $\delta > \max\{\underline{\delta}, 1 - \varepsilon/7\}$.

Appendix B. Comparison with existing approaches

Ekmekci et al. (2012) (hereafter, EGW) derives a lower bound on the patient player's payoff using the entropy approach. Their approach does not apply directly to our setting, because the short-run players observe the long-run player's announcement before moving, and announcements and actions are imperfectly recalled.

To summarize, EGW considers a model where the long-lived player 1 is replaced in every period with probability $\rho > 0$ and their type is redrawn after each replacement. They show that when player 1 is commitment type $\widehat{\omega}$ with probability $\mu(\widehat{\omega})$, the rational type's payoff from imitating commitment type $\widehat{\omega}$ is at least:

$$w_{\widehat{\omega}}\Big(-(1-\delta)\log \mu(\widehat{\omega}) - \log(1-\rho)\Big),$$

where $w_{\widehat{\omega}}(x)$ decreases in x and converges to type $\widehat{\omega}$'s optimal commitment payoff when $x \to 0$. Their result implies that player 1 can guarantee his commitment payoff when δ is close to 1 and ρ is close to 0. Their proof uses the following bound on the divergence between the equilibrium distribution of player 2 histories and that induced by the commitment type. In particular, let $P_{\sigma}^{2,n}$ be the probability measure over player 2_n's histories under equilibrium σ, and let $\widehat{P}_{\sigma'}^{2,n}$ be the probability measure over player 2_n's histories when player 1 deviates and imitates commitment type $\widehat{\omega}$. Let $d(\cdot||\cdot)$ be the KL-divergence between two distributions. They show that for every $n \in \mathbb{N}$,

$$\sum_{t=0}^{n-1} \mathbb{E}[d(\Pr_\sigma(h_2^{t+1}|h_2^t)||\Pr_{\widehat{\omega}}(h_2^{t+1}|h_2^t))] = d\left(\widehat{P}_{\sigma'}^{2,n} \middle\| P_\sigma^{2,n}\right) \leq -\log\mu(\widehat{\omega}) - n\log(1-\rho), \tag{B.1}$$

i.e., that the chain rule for relative entropy applies to h^n and $d\left(\widehat{P}_{\sigma'}^{2,n} \middle\| P_\sigma^{2,n}\right)$ is bounded from above by a linear function of n, with coefficient $-\log(1-\rho)$.

Let $d_\delta^{\sigma,\widehat{\omega}}$ be the expected discounted-average relative entropy between the predictions of player 2s over their next signal when they rely on the equilibrium distribution and when they rely on the distribution under player 1's deviation. EGW shows that

$$\begin{aligned} d_\delta^{\sigma,\widehat{\omega}} &:= (1-\delta)\sum_{t=0}^{\infty} \delta^t \mathbb{E}[d(\Pr_\sigma(h_2^{t+1}|h_2^t)||\Pr_{\widehat{\omega}}(h_2^{t+1}|h_2^t))] \\ &\leq -(1-\delta)\log\mu(\widehat{\omega}) - \log(1-\rho). \end{aligned} \tag{B.2}$$

Following the argument in Gossner (2011), the ex-ante expected payoff of the long-run player from imitating commitment type $\widehat{\omega}$ is bounded by $w_{\widehat{\omega}}(d_\delta^{\sigma,\widehat{\omega}})$.

To directly apply EGW's approach to show that player 1 receives at least their commitment payoff, we need the coefficient in the RHS of (B.1) to converge to 0 in the relevant limit, and the limit of ε-confirming best response payoffs to approach the commitment payoff. However, in our setting the announcements occur prior to the short-run player's move. Recall that $h_2^t = \{y_1, \ldots, y_{t-1}, m_t\}$. We cannot apply the chain rule for relative entropy to conclude that the sum of expected relative entropy under equilibrium and $\widehat{\omega}$ is equal to the ex-ante KL divergence of histories under the two strategy profiles: because m_t is imperfectly recalled, it is part of h_2^t but no other period's history, so the left-hand side of (B.1) fails to hold.

Intuitively, the deviating strategy (e.g., the naive commitment strategy) may not coincide with the equilibrium strategy of any type. The expected discounted entropy approach uses the chain rule to bound the expected total discounted surprise of the short run player over histories by a constant term, irrespective of δ, using the fact that the long run player plays exactly like the positive-probability commitment type $\widehat{\omega}$. In our setting, there is always a nonzero probability that an honest player makes a sequence of unlikely announcements, but this event depends on independent draws of $\tilde{A}_t$, period by period. Although the decrease in the long-run player's reputation (i.e., likelihood ratio between the honest type and the opportunistic type) after every such announcement is bounded, the sum over all periods may be unbounded as $\delta \to 1$, and so the discounted average does not vanish.

Note that the average expected entropy of outcomes for short run players prior to their observation of the announcement, i.e. with respect to histories $\hat{h}_2^t = \{y_1, \ldots, y_{t-1}\}$, does vanish. However, it does not yield the desired payoff bound. The reason is that, prior to knowing the period's announcement, the distribution of y_t does not identify *which* promises the long-run player will keep. Indeed, if, under $\widehat{\omega}$ the long-run player breaks their promise with small probability (say, η), then the set of strategies for the long-run player that generate the same outcome distribution includes a profile in which they announce $\widehat{\omega}$ with probability $\frac{\eta}{2-3\eta}$; and conditional on doing so, they keep their promise with probability only η, whereas if they announce a different action, they keep their promise with probability $(1-\eta)$. This distribution of play is possible when the long-run player is very likely to be opportunistic (note that there is no explicit bound on the distribution of the long-run player's type, only on the induced distribution of outcomes). The short-run player's best response to such a strategy need not yield the long-run player playing $\widehat{\omega}$ a payoff approaching what they achieve when the short-run player best responds to $\widehat{\omega}$. As a result, one cannot directly apply EGW's approach to show that player 1 can secure his commitment payoff in the limit where η approaches 0 and δ approaches 1 by playing as if they were a commitment type.

References

Abreu, D., Gul, F., 2000. Bargaining and reputation. Econometrica 68, 85–117.
Abreu, D., Pearce, D., 2007. Bargaining, reputation, and equilibrium selection in repeated games with contracts. Econometrica 75, 653–710.
Atakan, A., Ekmekci, M., 2015. Reputation in the long-run with imperfect monitoring. J. Econ. Theory 157, 553–605.
Azuma, K., 1967. Weighted sums of certain dependent random variables. Tohoku Math. J. (2) 19, 357–367.
Bagwell, K., 2018. Commitment with private resolve. Working Paper. Stanford University.
Benabou, R., Laroque, G., 1992. Using privileged information to manipulate markets: insiders, gurus, and credibility. Q. J. Econ. 107, 921–958.
Best, J., Quigley, D., 2020. Persuasion for the long run. Available at SSRN 2908115.
Board, S., Meyer-ter-Vehn, M., 2013. Reputation for quality. Econometrica 81, 23812462.
Bohren, J.A., Hauser, D.N., 2021. Learning with heterogeneous misspecified models: characterization and robustness. Econometrica 89, 3025–3077.
Celentani, M., Fudenberg, D., Levine, D., Pesendorfer, W., 1996. Maintaining a reputation against a long-lived opponent. Econometrica 64, 691–704.
Charness, G., Dufwenberg, M., 2006. Promises and partnership. Econometrica 74, 1579–1601.
Chen, Y., 2011. Perturbed communication games with honest senders and naive receivers. J. Econ. Theory 146, 401–424.
Chen, Y., Kartik, N., Sobel, J., 2008. Selecting cheap-talk equilibria. Econometrica 76, 117–136.
Daley, B., Green, B., 2012. Waiting for news in the market for lemons. Econometrica 80, 1433–1504.
Ekmekci, M., Gossner, O., Wilson, A., 2012. Impermanent types and permanent reputations. J. Econ. Theory 147, 162–178.
Ely, J., Valimaki, J., 2003. Bad reputation. Q. J. Econ. 118, 785814.
Esponda, I., Pouzo, D., 2016. Berk-Nash equilibrium: a framework for modeling agents with misspecified models. Econometrica 84, 1093–1130.
Esponda, I., Pouzo, D., 2020. Equilibrium in misspecified Markov decision processes. Theor. Econ.
Esponda, I., Pouzo, D., Yamamoto, Y., 2021. Asymptotic behavior of Bayesian learners with misspecified models. J. Econ. Theory 195.
Fudenberg, D., Gao, Y., Pei, H., 2020. A reputation for honesty. arXiv preprint. arXiv:2011.07159.
Fudenberg, D., Lanzani, G., Strack, P., 2021. Limit points of endogenous misspecified learning. Econometrica 89, 1065–1098.
Fudenberg, D., Levine, D., 1989. Reputation and equilibrium selection in games with a patient player. Econometrica 57, 759–778.
Fudenberg, D., Levine, D., 1992. Maintaining a reputation when strategies are imperfectly observed. Rev. Econ. Stud. 59, 561–579.
Fudenberg, D., Maskin, E., 1991. On the dispensability of public randomization in discounted repeated games. J. Econ. Theory 53, 428–438.
Gneezy, U., 2005. Deception: the role of consequences. Am. Econ. Rev. 95, 384–394.
Gneezy, U., Kajackaite, A., Sobel, J., 2018. Lying aversion and the size of the lie. Am. Econ. Rev. 108, 419–453.
Gossner, O., 2011. Simple bounds on the value of a reputation. Econometrica 79 (5), 1627–1641.
He, K., 2021. Mislearning from censored data: the gambler's fallacy in optimal-stopping problems. arXiv preprint. arXiv:1803.08170.
Jullien, B., Park, I.-U., 2020. Communication, feedbacks and repeated moral hazard with short-lived buyers. SSRN 3655043.
Kambe, S., 1999. Bargaining with imperfect commitment. Games Econ. Behav. 28, 217–237.
Kartik, N., 2009. Strategic communication with lying costs. Rev. Econ. Stud. 76, 1359–1395.
Kartik, N., Ottaviani, M., Squintani, F., 2007. Credulity, lies, and costly talk. J. Econ. Theory 134, 93–116.
Kim, K., 2009. The Coase conjecture with incomplete information on the monopolist's commitment. Theor. Econ. 4, 17–44.
Mathevet, L., Pearce, D., Stacchetti, E., 2019. Reputation and information design. NYU Working Paper.
Olszewski, W., 2004. Informal communication. J. Econ. Theory 117, 180200.

Ottaviani, M., Sørensen, P.N., 2006. Reputational cheap talk. Rand J. Econ. 37, 155–175.
Pei, H., 2020. Repeated communication with private lying cost. SSRN 3528641.
Pei, H., 2021. Trust and betrayals: reputational payoffs and behaviors without commitment. Theor. Econ. 16, 449–475.
Sanktjohanser, A., 2020. Optimally stubborn. Cowles Foundation Discussion Paper No. 2255.
Schmidt, K., 1993. Commitment through incomplete information in a simple repeated bargaining game. J. Econ. Theory 60, 114–139.
Sobel, J., 1985. A theory of credibility. Rev. Econ. Stud. 52, 557–573.