Appendix A. Proof of Proposition

Proof. We solve the game backwards and prove the claims by mathematical induction, initially focusing on $r^M > r^F$. First, we derive conditions under which $AM$ will be played in $M$’s last move on the equilibrium path, $n^M = N^M$ (the inductive basis). Specifically, part A) of the proof will examine the case $R = 0$, and part B) the case $R > 0$. Second, supposing that this holds for some $n^M = N^M$, we show in part C) the conditions under which the same is true for $n^M - 1$ as well.

A) $n^M = N^M$ under $R = 0$. Here we have $T(r^M, r^F) = r^M$, and therefore $N^M = 1$ and $N^F = \frac{r^M}{r^F}$. Solving backwards, we know $F$ would like to play the best response to $M$’s initial action, $F_n^* \in B(M_1), \forall n^F$. From her second move till the end of the dynamic stage game $F$ can observe $M_1$, and will hence rationally respond with $PF$ to $M_1$, and $AF$ to $M_1^P$.

Moving backwards, $M$ uses this information and hence knows that if he opens with $AM$ he will from period $r^F$ onwards be rewarded by payoff $a$. But $M$ also knows that such inducement play may be costly, payoff $b$, if $F$ plays $F_1^A$. Therefore, to achieve the Ricardian World $M$’s victory reward must more than offset his conflict cost, in which case $M$’s optimal play in period 1 will be $AM$ even if he knows with certainty that $F_1^A$ will be played. Formally, the incentive compatibility condition (15) in the main text needs to hold. Using (7)-(8) and rearranging yields equation (16). The $r^M(0)$ threshold is therefore the necessary and sufficient degree of $M$ commitment that delivers the Ricardian World under $R = 0$.

B) $n^M = N^M$ under $R > 0$: We know that the number of $M$’s moves is $N^M = \frac{T(r^M, r^F)}{r^M} > 1$. A condition analogous to (15) is $br^F + a(r^M - r^F R) > dr^M$ which implies, using (7)-(8) and rearranging,

$$r^M > \frac{\frac{1}{2} - \rho_M}{\phi_M - \rho_M} R r^F.$$  

C) $n^M + 1 \rightarrow n^M$ (if applicable, ie if $1 \leq n^M < N^M$): The proof proceeds by induction. We first assume that $M$’s unique best play in the $(n^M+1)$-th step is $AM$ regardless

\[ \text{of parameter values satisfying (12)} \]
of $F$’s preceding play (i.e. that $M_{n+1}$ is history-independent), and we attempt to prove that this implies the same assertion for the $n^M$-th step. Intuitively, this means that if $M$ inflates he finds it optimal to immediately disinflated. Two scenarios are possible in terms of the underlying $F$ behaviour that determines the costs of the disinflation. If $F$ runs a deficit, $AF$, the conflict costs $b$ and $w$ will occur for at least one period. In contrast, if $F$ switches to $PF$ pre-emptively (in anticipation of the disinflation, it will only be accompanied by the payoffs $a$ and $v$ and hence costless. This implies that one of the following two conditions will apply at any move $n^M$.

\begin{align}
\text{(21)} & \quad bk_n + a(r^M - k_n) + a[r^F - (r^F - k_{n+1})] > dr^M + b[r^F - (r^F - k_{n+1})] \quad , \\
& \quad (AM, AF) : \text{costly disinflation} \\
\text{(22)} & \quad bk_n + a(r^M - k_n) > d[r^M - (r^F - k_{n+1})] + a(r^F - k_{n+1}) \quad , \\
& \quad (AM, PF) : \text{costless disinflation}
\end{align}

Analogously to (15), the left-hand and the right-hand sides report $M$’s payoff from playing $AM$ and $PM$ respectively, using the information on what the subsequent action will be from the induction argument. Which of these two conditions is relevant for a certain $n^M$ depends on $F$’s payoffs $\{v, w, y, z\}$, and importantly on $k_{n+1}$. In particular, if

\begin{equation}
\begin{aligned}
(AM, AF) & \quad z(r^F - k_{n+1}) + wk_{n+1} \geq y(r^F - k_{n+1}) + vk_{n+1} \\
(AM, PF) & \quad (PM, AF) \quad (PM, PF)
\end{aligned}
\end{equation}

then (21) obtains, otherwise (22) is the relevant condition. Now, we will show that if the conditions (21) and (22) are satisfied at $n^M = 1$, then they hold in all other $n^M$ as well. This interesting feature notably simplifies the solution of the game.

**Lemma 1.** Consider the Game of Chicken in which (7)-(8) hold and $\delta_F = \delta_M = 1$. For any given $R$, out of the necessary and sufficient conditions for the Ricardian World, $\{M^A_n\} = B(F^A)$, the one regarding the initial move $n^M = 1$ yields as least as high $r^M(R)$ as any other $n^M$. Therefore, $\{M^A_1\} = B(F^A_1)$ is the sufficient condition.

**Proof.** Equations (21) and (22) can be, respectively, rearranged into

\begin{equation}
\begin{aligned}
r^M & > (k_n - k_{n+1}) (a - b) \quad \text{and} \quad r^M > r^F + \frac{k_n (a - b)}{a - d} - k_{n+1}. \\
\end{aligned}
\end{equation}

Using the specific payoffs in (7)-(8) yields

\begin{equation}
\begin{aligned}
r^M & > \frac{(k_n - k_{n+1}) (1 - \phi_M \rho_M)}{\phi_M - \rho_M} \quad \text{and} \quad r^M > r^F + \frac{k_n (1 - \phi_M \rho_M)}{\phi_M - \rho_M} - k_{n+1}. \\
\end{aligned}
\end{equation}

The strength of both conditions is increasing in $k_n$ and decreasing in $k_{n+1}$. Thus the strongest condition is guaranteed by the maximum of $(k_n - k_{n+1})$. From (10) it follows that $k_n - k_{n+1} \leq R r^F$. The fact that $k_1 - k_2 = R r^F$ then proves the claim for $R > 0$. Realizing that for $R = 0$ we have $N^M = 1$ finishes the proof.

Continuing the proof of Proposition 1, Lemma 1 means that regardless of the exact dynamics/asynchrony $R$, it suffices to focus on the initial simultaneous move (similarly to a one-shot game) assuming that all further relevant conditions hold. If the strongest
condition for \( n^M = 1 \) is satisfied we then know that a unique (type of) equilibrium outcome obtains throughout. Lemma 1 therefore implies, in combination with the recursive scheme, that throughout the proof we can use the following:

\[
k_n = k_1 = r^F \quad \text{and} \quad k_{n+1} = k_2 = (1 - R)r^F.
\]

Substituting (26) into (25) we obtain, together with (16)

\[
r^M > \tilde{r}^M(R) = \begin{cases} 
\frac{1}{3} \rho_M^F r^F & \text{if } R = 0, \\
\left(\frac{2}{3} \rho_M^F - \rho_M^F + \tilde{R}\right) r^F & \text{if } R \leq \tilde{R} = \frac{1}{2} - 2\rho_F, \\
\frac{1}{3} \rho_M^F R r^F & \text{if } R > \tilde{R} = \frac{1}{2} - 2\rho_F,
\end{cases}
\]

where the threshold \( \tilde{R} \in (0, 1) \) is derived from (23). The \( \tilde{r}^M(R) \) variable is the necessary and sufficient threshold for the Ricardian World (note that all three are at least as strong as the condition for \( N^M \) in (20)). By inspection, \( \tilde{r}^M(R) \) is, for all \( R \), monotonically increasing in \( r^F \) and \( \rho_M^F \), and decreasing in \( M \). It is also increasing in \( F \) which follows from the fact that the condition for \( R \geq \tilde{R} \) is stronger than the one for \( R > \tilde{R} \), and hence a higher \( \rho_F \) increases \( \tilde{R} \) and leads to strengthening of (27). This completes the proof of claim (i).

In terms of claim (ii), by symmetry the necessary and sufficient condition for the non-Ricardian World is

\[
r^F > \tilde{r}^F(R) = \begin{cases} 
\frac{1}{3} \phi_F^M r^M & \text{if } R = 0, \\
\left(\frac{2}{3} \phi_F^M - \phi_F^M + \tilde{R}\right) r^M & \text{if } R \leq \tilde{R} = \frac{1}{2} - 2\phi_F, \\
\frac{1}{3} \phi_F^M R r^M & \text{if } R > \tilde{R} = \frac{1}{2} - 2\phi_F.
\end{cases}
\]

Notice that the threshold \( \tilde{r}^F(R) \) is just a ‘mirror-image’ of the threshold \( \tilde{r}^M(R) \). Furthermore, the former threshold can be expressed in terms of \( r^M \) rather than \( r^F \) to obtain the threshold \( \tilde{r}^M(R) \) in the main text. Specifically, switching sides of \( r^M \) and \( r^F \) (28) can be re-written as

\[
r^M < \tilde{r}^M(R) = \begin{cases} 
\rho_F^M r^F & \text{if } R = 0, \\
\left(\frac{2}{3} \rho_F^M - \rho_F^M + \tilde{R}\right) r^F & \text{if } R \leq \tilde{R} = \frac{1}{2} - 2\phi_M, \\
\frac{1}{3} \rho_F^M R r^F & \text{if } R > \tilde{R} = \frac{1}{2} - 2\phi_M.
\end{cases}
\]

By inspection, \( \tilde{r}^M(R) \) is, for all \( R \), increasing in \( r^F \) and \( \rho_F \), and decreasing in \( \phi_F \). In addition, given that the condition for \( R \leq \tilde{R} \) is now weaker than the one for \( R > \tilde{R} \), the threshold \( \tilde{r}^M(R) \) is also decreasing in \( \phi_M \) (in a step manner). This completes the proof of Proposition 1.

\[\text{Equation (27) implies that the conditions for the } R > 0 \text{ cases only differ quantitatively from the } R = 0 \text{ case, not qualitatively.}\]
Appendix B. Proof of Proposition 2

Proof. To prove this existence claim it suffices to provide a specific example. Let us consider the simplest case of $R > 0$, namely $r^M = 3, r^F = 2$ (implying $R = \frac{1}{2}$) and the payoffs in (7)-(8) with $\rho_M = 0$. To prove that there exists no Ricardian SPNE it suffices to show that $F$ will play $AF$ in one of her moves regardless of the preceding move of $M$. To prove that there exists at least one non-Ricardian SPNE it suffices to show that in neither of his moves $M$ will play $AM$ regardless of $F$’s preceding move.

Focus on the condition for $M$’s last move to be uniquely $AM$ in equation (20), $r^M < \frac{R}{4\phi_M}$. Notice that since $R = \frac{1}{2}$, under $\phi_M < \frac{\phi_M}{2}$ the condition is not satisfied. Therefore, $M_2$ is not history-independent and it will be the best response to $F$’s preceding move, $F_2$. Moving backwards, player $F$ takes this into account in comparing the continuation payoffs from $F^P_2$ and $F^A_2$. Under $M^A_1$ the continuation payoff from playing $F^P_2$ is $-4\rho_F$, whereas from playing $F^A_2$ it is $-\frac{1}{4} - 3\phi_F$. Therefore, if $\rho_F > \frac{1}{4} + \frac{3\phi_F}{4}$ then $F_2$ is history-independent - regardless of $M$’s preceding move, $M_1$, $F$ will uniquely play $F^A_2$ in order to ensure the non-Ricardian regime levels for the rest of the stage game. This proves that in this case there exists no Ricardian SPNE as there will never be $F^P_2$ on the equilibrium path.

In order to prove that there exists a non-Ricardian SPNE it suffices to note that, similarly to $M_2$, the $M^A_1$ level is not a unique play regardless of the level played in $F_1$. Put differently, we have $M^P_1 \in B(F^A)$ since $M$ knows that $F^A_2$ is always played and there would be no victory reward from $M^A_1$. This implies that the non-Ricardian SPNE with $(F^A_1, M^P_1, F^A_2, M^P_2, F^A_3)$ on the equilibrium path belongs to the set of SPNE. □

Appendix C. Proof of Proposition 3 (can be removed)

Proof: The derivation of the generalized necessary and sufficient threshold is analogous in all its aspects to that of Proposition 1. In part A) the condition corresponding to (15) under $M$’s impatience, $\delta_M < 1$, is

\[
\sum_{t=1}^{r^F} \delta_M^{t-1} + a \sum_{t=r^F+1}^{r^M} \delta_M^{t-1} > d \sum_{t=1}^{r^M} \delta_M^{t-1}.
\]

This can, using the formula for a sum of a finite series and rearranging, be written as

\[
\delta_M^* < \frac{(a - b) \delta_M^F + b - d}{a - d}.
\]

Taking the logarithms yields

\[
r^M > r^M (0) = \log \frac{(a - b) \delta_M^F + b - d}{a - d}.
\]

The condition of part B) is again weaker than that. To prove part C) let us extend the result of Lemma 1 under the general payoffs and players’ impatience.

Lemma 2. Lemma 1 holds \( \forall \delta_M \leq 1, \forall \delta_F \leq 1, \) and any general payoffs satisfying (11).

Proof. Lemma 1 shows this claim to hold under $\delta_M = \delta_F = 1$. The proof of Proposition 1 showed that the payoffs of the less committed player, $F$ in our case, only affect the
necessary and sufficient condition through the threshold $\tilde{R}$. The same will thus be true for the value of $\delta_F$. Let us therefore consider the effect of $M$’s impatience. Under $\delta_M < 1$, the inequality in (21) that applies to the case of $R > \tilde{R}$ becomes

$$\sum_{t=1}^{k_n} b \delta_{M,t-1} + a \sum_{t=k_n+1}^{r_M} \delta_{M,t-1} + a \sum_{t=r_M+1}^{r_{M+k_n+1}} \delta_{M,t-1} > d \sum_{t=1}^{r_M} \delta_{M,t-1} + b \sum_{t=r_M+1}^{r_{M+k_n+1}} \delta_{M,t-1}. \tag{32}$$

This can be, after some manipulation, rearranged into

$$\sum_{t=1}^{r_M} (a - b) \delta_{M,t-1} - (a - d) \sum_{t=1}^{r_M} \delta_{M,t-1} < (a - b) \delta_{M} k_n \frac{1 - \delta_{M}^{r_{M+k_n+1} - k_n}}{1 - \delta_{M}}. \tag{33}$$

Since $\delta_M < 1$ we see that, analogously to Lemma 1, the strength of the condition is increasing in $k_n$ and decreasing in $k_{n+1}$. Hence the same argument applies. We can readily check, using (22) under $M < 1$, that the same is true for $R$. \hfill \square

We will now complete the proof of Proposition 3 using this result. Lemma 2 implies that we need to substitute (26) into (33) for the costly disinflation case. Using formulas for finite sums, rearranging, and taking the logarithms yields

$$r_M > \log_{\delta_M} \frac{(a - d) - (a - b) \left(1 - \delta_F^{r_M}\right)}{(a - d) - (a - b) \left(1 - \delta_F^{r_M(1-R)}\right)}. \tag{34}$$

For the costless disinflation case, the analog of (22) under $\delta_M < 1$ is, using Lemma 2

$$\sum_{t=1}^{r_F} b \delta_{F,t-1} + a \sum_{t=r_F+1}^{r_M - r_F R} \delta_{M,t-1} + a \sum_{t=r_M - r_F R + 1}^{r_M} \delta_{M,t-1}, \tag{35}$$

and after rearranging

$$r_M > \log_{\delta_M} \frac{b \left(1 - \delta_F^{r_F}\right) + a \delta_F^{r_F} - d}{a \left(1 + \delta_F^{r_F} - \delta_F^{r_F(1-R)}\right) - d \delta_F^{r_F R}}. \tag{36}$$

The threshold $\tilde{R}$ determining whether the costly disinflation case of (34) or the costless disinflation case of (36) applies is derived from the generalization of (24) under $F$’s impatience. Specifically, under $\delta_F < 1$ if

$$\sum_{t=1}^{r_F R} b \delta_{F,t-1} + w \sum_{t=r_F R + 1}^{r_F} \delta_{M,t-1} > y \sum_{t=1}^{r_F R} \delta_{M,t-1} + v \sum_{t=r_F R + 1}^{r_F} \delta_{M,t-1}, \tag{37}$$

then (34) obtains, otherwise (36) is the relevant condition. After rearranging this implies the following threshold

$$\tilde{R} = \frac{1}{r_F} \log_{\delta_F} \frac{z - y + (v - w) \delta_F^{r_F}}{z - y + v - w}. \tag{38}$$
Combining (31), (34), (36), and \( \hat{R} \) from (38) yields the following generalized necessary and sufficient condition for the Ricardian World

\[
M > \hat{M}(R) = \begin{cases} 
\log \frac{(a-b)\delta_M + b-d}{a-d} & \text{if } R = 0, \\
\log \frac{(a-d)-(a-b)(1-\delta_F)}{a-b(1-\delta_F)} & \text{if } R \leq \hat{R}, \\
\log \frac{b(1-\delta_M) + \delta_F}{a(1+\delta_M - \delta_F(1-R)) - \delta_M^{-1}} & \text{if } R > \hat{R}.
\end{cases}
\]

We can now use this condition to prove the claims of Proposition 3. Examining (38) and (39) reveals that \( \hat{M}(R) \) is a function of \( r^F \), both players’ discount factors \( \delta_M \) and \( \delta_F \), and all the payoffs except \( c \).

In terms of the patience threshold, consider the logarithm’s numerator of (31), (34), and (36). For the threshold \( \hat{M}(R) \) to exist for all \( R \) it must hold that \( (a-b)\delta_M + b-d > 0 \). Rearranging this inequality yields the necessary patience threshold \( \delta_M \) in (17).

Finally, note that if \( \delta_F \) is below a certain threshold \( \hat{d}_F \) then there are cases in which \( \hat{M}_R = r^F \) as claimed in (18) where \( \hat{M}_R \geq r^F \). In such case any \( \hat{M} > r^F \) uniquely ensures discipline of both policies. The easiest way to see this is to consider \( \delta_F = 0 \). Such an impatient \( F \) will never reduce spending before the start of disinflation as she fully ignores the future. Therefore, disinflation will always be costly for both players, ie (36) no longer applies and (34) becomes the relevant condition \( \forall \hat{M}_R, R \in (0,1), \) and for all \( \{a,b,d,v,w,y,z\} \) satisfying (11). This completes the proof of Proposition 3.

\[ \square \]

APPENDIX D. PROOF OF PROPOSITION 4

Proof. After joining the union, there is a decrease in \( s_j \) and a possible increase in \( m_j \). If in country \( j \) the degree of free-riding is above a certain threshold

\[
m > m_j^*(s_j, v_j^R) \quad \text{where} \quad \frac{\partial m_j^*(s_j, v_j^R)}{\partial s_j} > 0 \quad \text{and} \quad \frac{\partial m_j^*(s_j, v_j^R)}{\partial v_j^R} > 0,
\]

then the value of \( v_j^A \) will fall below \( w_j = 0 \). This follows, using a continuity argument, from the monotonicity of \( v_j(m_j) \) and the assumed \( v_j(m_j = 1) < w_j \). In such case the underlying game after accession for country \( j \) is no longer the Game of Chicken but the Neglect scenario since \( AF \) becomes a strictly dominant strategy. Therefore, we will observe \( AF \) for any level of \( \delta_{M,j} \) and \( r_j^F \), even if the common central bank has \( \delta_M = 1 \) and \( r_M \rightarrow \infty \).

In terms of claim (ii), denote the number of countries in which \( m_j > m_j^*(s_j, v_j^R) \) by \( \gamma \in \mathbb{N} \), and order the member countries such that those \( \{1, \ldots, \gamma\} \) feature \( m_j > m_j^* \), and those \( \{\gamma + 1, \ldots, J\} \) feature \( m_j \leq m_j^* \). Let us report the conditions only for the special case \( R = 0 \) and \( \delta_M = 1 \) as it was shown to be representative of the other cases as well.
The condition analogous to (15) becomes
\[
\begin{align*}
    \sum_{j=1}^{J} b s_j r_j^F + b \sum_{j=1}^{\gamma} s_j (r^M - r_j^F) + a \sum_{j=\gamma+1}^{J} s_j (r^M - r_j^F) > \left. \frac{d r^M}{(PM, AF); \forall j} \right. \\
    \text{(AM, AF): } \forall j \leq \gamma \\
    \text{(AM, AF): } \forall j > \gamma \\
\end{align*}
\]

Note that the condition only differs from the no free-riding case in the second element on the left hand side, which is now the payoff \( b \) rather than \( a \) since the \( \gamma \) countries with \( m_j > \overline{m}_j \) will not switch to \( PF \). Intuitively, the cost of conflict is higher and the victory reward is lower. Rearranging yields
\[
(41) \\
    r^M > \overline{r^M} = \frac{(a - b) \sum_{j=\gamma+1}^{J} s_j r_j^F}{b \sum_{j=1}^{\gamma} s_j + a \sum_{j=\gamma+1}^{J} s_j - d}. \\
\]

By inspection, \( \overline{r^M} \) is increasing in the total size of the \( \gamma \) free-riders, \( \sum_{j=1}^{\gamma} s_j \). Since the numerator is positive, if the denominator is negative then the threshold \( \overline{r^M} \) does not exist. This means that even \( \bar{d}_M = 1 \) and \( \overline{r^M} \to \infty \) do not guarantee the \( AM \) outcome. By inspection this happens if \( \sum_{j=1}^{\gamma} s_j \) is above a certain threshold that is an increasing function of \( b \), and a decreasing function of \( a \) and \( d \). \( \square \)

\[32\] Alternatively, this condition can be expressed as \( d > \bar{d} \left( a, b, \gamma, s_j \right) \), where \( \bar{d} \) is decreasing in \( s_j \) for all \( j \leq \gamma \), and increasing in \( s_j \) for all \( j > \gamma \).