Econometrica Supplementary Material

SUPPLEMENT TO "BOUNDED MEMORY AND BIASES IN INFORMATION PROCESSING"

(*Econometrica*, Vol. 82, No. 6, November 2014, 2257–2294)

BY ANDREA WILSON

APPENDIX B: PROOF THAT THE DM'S PAYOFF STRICTLY INCREASES IN M

CLAIM: *If information matters, then the maximized expected payoff* $\Pi^*(M,\eta)$ *is strictly increasing under the following condition* ((7) *in the paper*), *and so equivalent states are not optimal*:

$$\text{(B.1)}\qquad \frac{\mu_S^L}{\mu_S^H}\frac{\eta+(1-\eta)\mu_1^H}{\eta+(1-\eta)\mu_1^L} < \frac{\mu_1^H}{\mu_1^L}\frac{\mu_S^H}{\mu_S^L} < \frac{\mu_1^L}{\mu_1^H}\frac{\eta+(1-\eta)\mu_S^H}{\eta+(1-\eta)\mu_S^L}.$$

PROOF: Let $(\widetilde{g}^0,\widetilde{\sigma},\widetilde{d})$ be an optimal rule for parameters (M,η), with payoff $\Pi^*(M,\eta)$. I show here that if $\xi(1)\xi(S)\le 1$, then a strict payoff gain is possible with one extra memory state whenever the first inequality in (B.1) holds. If $\xi(1)\xi(S)\ge 1$, a symmetric argument obtains under the second inequality in (B.1).

So assume $\xi(1)\xi(S)\le 1$. Assume further that there are no equivalent memory states under (g^0,σ,d); if there are equivalent states, then, as outlined in footnote 15, use instead a payoff-equivalent rule with unused states. Now suppose the DM is given one additional state and call it state 0. To prove the claim, I will first construct a rule $(\widetilde{g}^0,\widetilde{\sigma},\widetilde{d})$ for states $\{0,1,\dots,M\}$ that earns payoff $\Pi^*(M,\eta)$, and then show that the rule is strictly suboptimal. Therefore, an optimal rule for $\{0,1,\dots,M\}$ must earn a payoff strictly above $\Pi^*(M,\eta)$, thus establishing a strict payoff gain with one additional states.

To this end, let (g^0,σ,d) be a protocol that agrees with $(\widetilde{g}^0,\widetilde{\sigma},\widetilde{d})$ on states $\{1,2,\dots,M\}$, chooses action 0 in state 0, and specifies the following transitions in state 0: (i) if state 1 is *not* absorbing, then $\sigma_{0,1}^S=1$ and $\sigma_{0,0}^s=1\ \forall s\neq S$; (ii) if state 1 *is* absorbing, then $\sigma_{0,2}^S=\gamma$, $\sigma_{0,0}^S=1-\gamma$, $\sigma_{0,1}^1=1$, and $\sigma_{0,0}^s=0\ \forall s\notin\{1,S\}$, with $\gamma\in(0,1)$ a number to be determined. Observe that under (g^0,σ,d), no states transition into state 0, while transitions among states $\{1,2,\dots,M\}$ are as specified by the original rule $\widetilde{\sigma}$. Therefore, (g^0,σ,d) yields the same terminal probabilities $(f_1^\theta,\dots,f_M^\theta)$ as the original rule (with $f_0^\theta=0$) and thus earns the original payoff, $\Pi^*(M,\eta)$.

A PROFITABLE DEVIATION IF STATE 1 IS NOT ABSORBING. Suppose first that state 1 is not absorbing. I will show that the DM strictly prefers to move $1\downarrow_1 0$, implying that (g^0,σ,d) does not correspond to a team equilibrium and, therefore, is not optimal. For any signal s, define $v_{1s}^\theta\equiv\sum_{j\in\{0,\dots,M\}}\sigma_{1,j}^s v_j^\theta$ as the DM's expected payoff following an s-signal in memory state 1. Then since $d_1=0$ if

 DOI: 10.3982/ECTA12188

information matters (otherwise the DM chooses action 1 in all memory states) and since (g^0, σ, d) must specify staying in the lowest state, 1, after the lowest signal $s = 1$, continuation payoffs in state 1 satisfy

$$v_1^\theta = \eta(0) + (1-\eta)\mu_1^\theta v_1^\theta + \sum_{s=2}^{S}(1-\eta)\mu_s^\theta v_{1s}^\theta \tag{B.2}$$

$$\Rightarrow \quad \eta v_1^\theta = \sum_{s=2}^{S}(1-\eta)\mu_s^\theta\left(v_{1s}^\theta - v_1^\theta\right).$$

In state 0, continuation payoffs are given by

$$v_0^\theta = \eta(0) + (1-\eta)\mu_S^\theta v_1^\theta + (1-\eta)\sum_{s\neq S}\mu_s^\theta v_0^\theta \tag{B.3}$$

$$\Rightarrow \quad \frac{\eta v_1^\theta}{(\eta + (1-\eta)\mu_S^\theta)} = \left(v_1^\theta - v_0^\theta\right).$$

Substituting (B.2) into (B.3)yields

$$\left(v_1^\theta - v_0^\theta\right) = \frac{\eta v_1^\theta}{(\eta + (1-\eta)\mu_S^\theta)} = \frac{\sum_{s=2}^{S}(1-\eta)\mu_s^\theta(v_{1s}^\theta - v_1^\theta)}{(\eta + (1-\eta)\mu_S^\theta)}.$$

Taking ratios and multiplying by $\ell(1)\mu_1^H/\mu_1^L$ yields

$$\xi(1)\frac{\Delta_{1,0}^H}{\Delta_{0,1}^L} = \left(\frac{\eta + (1-\eta)\mu_S^L}{\eta + (1-\eta)\mu_S^H}\right)\frac{\sum_{s=2}^{S}\mu_1^H\mu_s^H(v_{1s}^H - v_1^H)}{\sum_{s=2}^{S}\mu_1^L\mu_s^L(v_1^L - v_{1s}^L)} \tag{B.4}$$

$$< \frac{\sum_{s=2}^{S}\mu_1^H\mu_s^H(v_{1s}^H - v_1^H)}{\sum_{s=2}^{S}\mu_1^L\mu_s^L(v_1^L - v_{1s}^L)}.$$

By Lemma 1(a) and the assumption of no equivalent states, $v_{1s}^H - v_1^H \geq 0 \geq v_{1s}^L - v_1^L$ for all $s \in \mathcal{S}$, with strict inequality unless the DM stays in state 1 after

s (in which case $v_{1s}^\theta = v_1^\theta$). But then consider the final term in (B.4): for any s where $v_{1s}^\theta \neq v_1^\theta$, the sth ratio is

$$\xi(1)\xi(s)\frac{v_{1s}^H - v_1^H}{v_1^L - v_{1s}^L} < \frac{v_{1s}^H - v_1^H}{v_1^L - v_{1s}^L} \leq 1/\ell(1) \tag{B.5}$$

(the first inequality by $\xi(1)\xi(s) < \xi(1)\xi(S)$ and our assumption $\xi(1)\xi(S) \leq 1$, and the final inequality by Corollary 1 (RP), which requires that the state-1 self prefer memory state 1 to all other states). And since we are assuming here that state 1 is not absorbing, there must be some signal $s \in \{2, \dots, S\}$ for which $v_{1s}^\theta \neq v_1^\theta$, which then satisfies (B.5). But then by (B.5), the final RHS expression in (B.4) is strictly below $1/\ell(1)$. It then follows that $\ell(1)\xi(1) < \Delta_{0,1}^L/\Delta_{1,0}^H \equiv \bar{\ell}_0$; so, by Proposition 2, the DM *strictly* prefers to move from state 1 to 0 if he observes a 1-signal, as desired to establish that (g^0, σ, d) (which never uses state 0) is strictly suboptimal.

PROFITABLE DEVIATION IF STATE 1 IS ABSORBING. Suppose next that memory state 1 *is* absorbing, so that $v_1^\theta = 0$. In this case, under the protocol (g^0, σ, d), continuation payoffs in state 0 satisfy

$$\begin{aligned} v_0^\theta &= \big(\eta + (1-\eta)\mu_1^\theta\big)v_1^\theta + (1-\eta)\mu_S^\theta\gamma v_2^\theta + (1-\eta)\big(1 - \mu_1^\theta - \mu_S^\theta\gamma\big)v_0^\theta \\ \Rightarrow &\quad \big(\eta + (1-\eta)\mu_1^\theta + (1-\eta)\mu_S^\theta\gamma\big)\Delta_{0,1}^\theta = (1-\eta)\mu_S^\theta\gamma\Delta_{2,1}^\theta. \end{aligned}$$

Taking ratios yields

$$\frac{v_0^H - v_1^H}{v_1^L - v_0^L} = \frac{\mu_S^H}{\mu_S^L}\frac{\eta + (1-\eta)\mu_1^L + (1-\eta)\mu_S^L\gamma}{\eta + (1-\eta)\mu_1^H + (1-\eta)\mu_S^H\gamma}\frac{\Delta_{2,1}^H}{\Delta_{1,2}^L}. \tag{B.6}$$

By equation (2) in the paper, using the fact that state 1 is absorbing and that $g_1^0 = 0$ if information matters (the DM cannot start in absorbing state) yields

$$f_1^\theta = f_1^\theta(1-\eta) + \sum_{j=2}^M f_j^\theta \tau_{j,1}^\theta \quad \Rightarrow \quad \frac{f_1^H}{f_1^L} = \frac{\sum_{j=2}^M f_j^H \tau_{j,1}^H}{\sum_{j=2}^M f_j^L \tau_{j,1}^L} \geq \frac{f_2^H}{f_2^L}\frac{\mu_1^H}{\mu_1^L}$$

(the final inequality by the ordering of the memory states and signal realizations). Thus, $\ell(1) \geq \ell(2)\xi(1)$. Together with (B.6), this implies

$$\begin{aligned} \ell(1)\xi(S)\frac{v_0^H - v_1^H}{v_1^L - v_0^L} &\geq \ell(2)\xi(1)\big(\xi(S)\big)^2 \\ &\quad \times \frac{\eta + (1-\eta)\mu_1^L + (1-\eta)\mu_S^L\gamma}{\eta + (1-\eta)\mu_1^H + (1-\eta)\mu_S^H\gamma}\frac{\Delta_{2,1}^H}{\Delta_{1,2}^L}. \end{aligned} \tag{B.7}$$

By Corollary 2, $\ell(2)\Delta^H_{2,1}/\Delta^L_{1,2} \equiv \ell(2)/\bar{\ell}_1 \geq 1$; then, if the first inequality in (B.1) holds, the RHS of (B.7) strictly exceeds 1 for γ sufficiently small. But then $\ell(1,S)(v^H_0 - v^H_1) \geq (v^L_1 - v^L_0)$, indicating that a state-1 self who observes S strictly prefers to move from state 1 to state 0. Therefore, protocol (g^0, σ, d) (in which state 1 never moves to state 0) violates the conditions for a team equilibrium and so, by Proposition 1, cannot be optimal. *Q.E.D.*

EXAMPLE—Absorbing and Equivalent States: Absorbing states and equivalent states are not *necessarily* optimal if (B.1) fails, but may occur for some parameters. For example, consider a binary signal and $M = 3$ memory states. If the prior bias satisfies condition (B.8), which is *possible* only if (B.1) is violated, then the unique optimal protocol has the following features: state 3 is absorbing and chooses the high action, state 1 chooses the low action and jumps to state 3 with chance $\sigma^S_{1,3} = 1$ after an S-signal, and state 2 is either unused or equivalent to state 1:

$$\text{(B.8)} \qquad \frac{(\bar{\xi}(\eta))^4}{(\xi(1))^2\xi(S)} < \beta < 1, \quad \text{where} \quad \bar{\xi}(\eta) \equiv \frac{\eta + (1-\eta)\mu^H_S}{\eta + (1-\eta)\mu^L_S}.$$

This is easily established via the following chain of arguments (which rely on some omitted but straightforward computations): (i) The DM cannot start in state 3, for *all* protocols (g^0, σ, d) with initial state 3 yield $\ell(1) \geq \ell_0(\xi(1))^2/(\bar{\xi}(\eta))^2$; together with (B.8) and the optimality condition $\ell_0/\bar{\ell}_2 \geq 1$ for starting in state 3, this implies $\ell(1,S)/\bar{\ell}_2 \geq (\xi(1))^2\xi(S)/(\bar{\xi}(\eta))^2 > 1$, and so equilibrium demands $\sigma^S_{1,3} = \sigma^S_{2,3} = 1$. But all protocols with $\sigma^S_{1,3} = \sigma^S_{2,3} = 1$ yield $1/\bar{\ell}_2 \equiv \Delta^H_{3,2}/\Delta^L_{2,3} \leq |\pi^H/\pi^L|$, so $\ell_0/\bar{\ell}_2 = \beta < 1$ by (B.8), violating the optimality condition for starting in state 3. Next, (ii) *all* protocols with initial state 1 or 2 yield $\ell(3,1) \geq \ell_0\xi(S)\xi(1)/(\bar{\xi}(\eta))^2$ and $1/\bar{\ell}_2 \equiv \Delta^H_{3,2}/\Delta^L_{2,3} \geq |\pi^H/\pi^L|\xi(1)/(\bar{\xi}(\eta))^2$; together with the first inequality in (B.8), we therefore have $\ell(3,1)/\bar{\ell}_2 > 1$, and so memory state 3 must be absorbing. But (iii) $\sigma^S_{1,3}$ must then equal 1, for all optimal protocols with an absorbing state 3 and initial state 1 or 2, yield $\ell(1,S) \geq \ell_0\xi(1)\xi(S)/(\bar{\xi}(\eta))^2$, the same as our lower bound on $\ell(3,1)$; thus, $\ell(1,S)/\bar{\ell}_2 > 1$ by (B.8), implying that a state-1 self who observes an S-signal must move to state 3 with probability 1. Finally, (iv) since $\sigma^S_{1,3} = 1$ and state 3 is absorbing, state 2 is used only if it is an initial state. But all protocols with $\sigma^S_{1,3} = \sigma^1_{3,3} = 1$ and initial state 2 yield $\ell(2) \leq \ell_0$ (equality iff $\sigma^1_{2,1} = 1$), in which case $\beta < 1$ (from (B.8)) implies $\ell(2)|\pi^H/\pi^L| \leq \beta < 1$: thus, whenever state 2 is used, it must choose action 0, in which case it is equivalent to state 1 (both choose the low action and move to an absorbing state 3 after the first S-signal).

APPENDIX C: DERIVATION OF STEP 4 IN THE PROOF OF PROPOSITION 4(c)

C.1. *Preliminary Notation and Analysis From the Paper*

Recall that the paper defines coefficients $\alpha_l \equiv \prod_{j=2}^{l-1}(\sigma^1_{j+1,j}/\sigma^S_{j,j+1})$ and $\beta_l = \prod_{j=l+1}^{M-1}(\sigma^S_{j-1,j}/\sigma^1_{j,j-1})$, and finds in Step 2 that an optimal protocol must satisfy the following equations (restated here for convenience), the LHS of each expression with equality if state i is sticky up and the RHS with equality if i is sticky down:

$$\text{(C.1)}\qquad i \le i^* - 1:\quad \sum_{l=2}^{i} \frac{\alpha_l x_l}{\sigma^1_{l,l-1}} \Big/ \sum_{l=2}^{i} \frac{y_l}{\alpha_l} \le \left(\sigma^S_{1,2}\right)^2/\eta + o(\eta)/\eta \le \sum_{l=2}^{i} \frac{\alpha_l x_l}{\sigma^1_{l,l-1}} \Big/ \sum_{l=2}^{i-1} \frac{y_l}{\alpha_l},$$

$$\text{(C.2)}\qquad i^* \le i \le i_0:\quad \frac{\sum_{l=2}^{i} \frac{\alpha_l}{\sigma^1_{l,l-1}} x_l}{\sum_{l=2}^{i^*-1} \frac{y_l}{\alpha_l} + \frac{\alpha_M}{\sigma^S_{1,2}} \sum_{l=i^*}^{i} \frac{u_l}{\alpha_l}} \le \frac{(\sigma^S_{1,2})^2 + o(\eta)}{\eta} \le \frac{\sum_{l=2}^{i} \frac{\alpha_l}{\sigma^1_{l,l-1}} x_l}{\sum_{l=2}^{i^*-1} \frac{y_l}{\alpha_l} + \frac{\alpha_M}{\sigma^S_{1,2}} \sum_{l=i^*}^{i-1} \frac{u_l}{\alpha_l}},$$

$$\text{(C.3)}\qquad i \ge i_0:\quad \sum_{l=i}^{M-1} \frac{\beta_l v_l}{\sigma^S_{l,l+1}} \Big/ \sum_{l=i+1}^{M-1} \frac{u_l}{\beta_l} \ge \frac{(\sigma^1_{M,M-1})^2 + o(\eta)}{\eta} \ge \sum_{l=i}^{M-1} \frac{\beta_l v_l}{\sigma^S_{l,l+1}} \Big/ \sum_{l=i}^{M-1} \frac{u_l}{\beta_l},$$

$$\text{(C.4)}\qquad \frac{(\sigma^S_{1,2})^2}{\eta} \to \frac{\sum_{l=2}^{i_0} \frac{\alpha_l}{\sigma^1_{l,l-1}} x_l + \frac{\sigma^S_{1,2}}{\alpha_M} \sum_{l=i_0}^{M-1} \frac{\alpha_l}{\sigma^S_{l,l+1}} v_l}{\sum_{l=2}^{i^*-1} \frac{y_l}{\alpha_l} + \frac{\alpha_M}{\sigma^S_{1,2}} \sum_{l=i^*}^{M-1} \frac{u_l}{\alpha_l}},$$

$$\frac{a_M}{\sigma^S_{1,2}} \to \left(\frac{\lambda_1}{\lambda_2}\right)^{(M-1)/2} \left(\frac{\lambda^{(M-1)/2} - \sqrt{\beta}}{\sqrt{\beta}\lambda^{(M-1)/2} - 1}\right),$$

where $\lambda_1 \equiv \mu_S^H/\mu_1^H$, $\lambda_2 \equiv \mu_1^L/\mu_S^L$, and

$$\text{(C.5)} \qquad \begin{aligned} x_l &\equiv \left(\frac{\xi(1)\lambda^{l-1}-1}{\mu_1^H\lambda_1^{l-1}}\right), \\ y_l &\equiv \left(\frac{\lambda^{l-1}-1}{\lambda_2^{l-1}}\right), \\ v_l &\equiv \left(\frac{\lambda^{M-l-1}\dfrac{1}{\xi(1)}-1}{\mu_S^L\lambda_2^{M-l}}\right), \\ u_l &\equiv \left(\frac{\lambda^{M-l}-1}{\lambda_1^{M-l}}\right). \end{aligned}$$

Observe also that if the second expression in (C.2) holds, then the first expression in (C.4) reduces to the following alternative optimality condition for $i \downarrow_1 i-1$,[1] which must hold with equality if $\sigma^1_{i,i-1} \in (0,1)$:

$$\text{(C.6)} \qquad i \downarrow_1 i-1\colon \quad \frac{(\sigma^1_{M,M-1})^2 + o(\eta)}{\eta} \geq \frac{\left(\dfrac{\alpha_M}{\sigma^S_{1,2}}\right)\displaystyle\sum_{l=i+1}^{i_0}\frac{\alpha_l}{\sigma^1_{l,l-1}}x_l + \sum_{l=i_0}^{M-1}\frac{\alpha_l}{\sigma^S_{l,l+1}}v_l}{\displaystyle\sum_{l=i}^{M-1}u_l/\alpha_l}.$$

Also, by equation (20) in the paper, $\sqrt{\beta} > \max\{\lambda^{(M+1)/2-i^*}, \xi(S)\lambda^{i_0-(M+3)/2}\}$. Substituting this into the second condition in (C.4) yields

$$\text{(C.7)} \qquad \frac{a_M}{\sigma^S_{1,2}} < \min\left\{\left(\frac{\lambda_1}{\lambda_2}\right)^{(M-1)/2}\lambda^{(M+1)/2-i^*}\left(\frac{\lambda^{i^*-1}-1}{\lambda^{M-i^*}-1}\right), \left(\frac{\lambda_1}{\lambda_2}\right)^{(M-1)/2}\lambda^{i_0-(M+3)/2}\left(\frac{\lambda^{M+1-i_0}-\xi(S)}{\xi(S)\lambda^{i_0-2}-1}\right)\right\}.$$

[1]In the spirit of the notation in the paper, I use $i \downarrow_1 i-1$ to denote a downward transition i to $i-1$ after signal 1 and use $i \uparrow_S i+1$ to denote an upward transition i to $i+1$ after signal S.

C.2. *Proofs for General Signals*

For any states $k \geq 3$ and i, define the functions, with x_l, y_l, u_l, and v_l given by (C.5),

$$U^k(i) = \begin{cases} \sum_{l=k+1}^{i} x_l \Big/ \sum_{l=k}^{i} y_l & \text{for } k < i \leq i^* - 1, \\ \dfrac{\sum_{l=k+1}^{i} x_l}{\sum_{l=k}^{i^*-1} y_l + \frac{\alpha_M}{\sigma_{1,2}^S} \sum_{l=i^*}^{i} u_l} & \text{for } k \leq i^* \leq i \leq i_0,^{2} \\ \sum_{l=i}^{k} v_l \Big/ \sum_{l=i+1}^{k} u_l & \text{for } i_0 \leq i \leq k, \end{cases}$$

$$D(i) = \begin{cases} \sum_{l=2}^{i} x_l \Big/ \sum_{l=2}^{i-1} y_l & \text{if } i \leq i^* - 1, \\ \dfrac{\sum_{l=2}^{i} x_l}{\sum_{l=2}^{i^*-1} y_l + \frac{\alpha_M}{\sigma_{1,2}^S} \sum_{l=i^*}^{i-1} u_l} & \text{if } i^* \leq i \leq i_0, \end{cases}$$

$$\widetilde{D}(i) = \begin{cases} \left(\frac{\alpha_M}{\sigma_{1,2}^S} \sum_{l=i+1}^{i_0} x_l + \sum_{l=i_0}^{M-1} v_l \right) \Big/ \sum_{l=i}^{M-1} u_l & \text{for } i^* \leq i \leq i_0 - 1, \\ \sum_{l=i}^{M-1} v_l \Big/ \sum_{l=i}^{M-1} u_l & \text{for } i \geq i_0. \end{cases}$$

Also define the states

$$\text{(C.8)} \qquad i_1 \equiv \arg \min_{2 \leq i \leq i_0} D(i), \quad i_2 \equiv \arg \max_{M-1 \geq i \geq i^*} \widetilde{D}(i).$$

If $i_1 \leq i_2$ and $D(i) < \widetilde{D}(i)$ (Step C.5 shows that a sufficient condition for this is that $i_2 \geq i_1 + 2$), then define $\hat{I} \equiv \{i_1, \dots, i_2\}$; if either of these conditions fail, then let $\hat{I}$ be the empty set. Proposition C simply restates Step 4 of Section A.9.

PROPOSITION C: *If* $\lambda_1 \leq \lambda_2$, *then* (i) *all interior states are fluid up*; (ii) *all states* $i \leq i_1$ *are fluid down*, *as are all states* $i \geq i_2$; (iii) *if* $\hat{I}$ *is nonempty*, *then all states*

[2] If $i^* \leq k < i$, set the first sum in the denominator to zero, and replace the second by $\sum_{l=k}^{i} u_l$.

in $\hat{I}$ are sticky down. Also, (iv) *If $\lambda_1 \geq \lambda_2$, then all interior states $i \notin I^*$ are fluid down.*

PROOF: The idea of this proof is as follows: The function $U^k(i)$ describes a lower bound (for $i \leq i_0$, lower bound for $i \geq i_0$) on $(\sigma^S_{1,2})/\eta$ for the state-i self to find it incentive compatible to move up after S. And the function $D(i)$ describes the upper bound on $(\sigma^S_{1,2})^2/\eta$ for optimality of $i \downarrow_1 i-1$ with $i \leq i_0$, while $\widetilde{D}(i)$ describes an lower bound on $(\sigma^1_{M,M-1})^2/\eta$ for optimality of $i \downarrow_1 i-1$ when $i \geq i^*$ (for states between i^* and i_0, the two bounds coincide by (C.4)). The proof below will show that as long as $U^k(i)$ is increasing, no state i can be sticky up (and symmetrically, as long as $D(i)$ is decreasing, no state i can be sticky down): for if i is indifferent, then $(\sigma^S_{1,2})^2/\eta$ is too low for state $i+1$ to move up at all. On the other hand, $D(i-1) < D(i) < D(i+1)$ means that if i is sticky down, then so too must be $i-1$ and $i+1$. (And symmetrically, though this is not shown, if $U(i-1) > U(i) > U(i+1)$, then upward stickiness in state i implies that $i-1, i+1$ must also be upward sticky). Step C.1 below shows that if $\lambda_1 \leq \lambda_2$, then $U^k(i)$ is strictly increasing for all i; if $\lambda_2 \leq \lambda_1$, then $D(i)$ and $\widetilde{D}(i)$ are strictly decreasing $\forall i \notin I^*$. Step C.2 uses this result to prove parts (i) and (iv) of the proposition. Step C.1 also shows that if $\lambda_1 \leq \lambda_2$, then $D(i)$ and $\widetilde{D}(i)$ are decreasing to i_1, then increasing from i_1 to i_2, and decreasing for $i > i_2$. Step C.3 uses this to prove part (ii). Step C.4 then shows that if any state in $\{i_1, \dots, i_2\}$ is sticky down, then all states in this set must be sticky down, and Step C.5 shows that some state in this set (hence all states) are sticky down iff $D(i_1) < \widetilde{D}(i_2)$. Thus, if $\hat{I}$ (defined below (C.8)) is nonempty, then all states in $\hat{I}$ are sticky down. *Q.E.D.*

STEP C.1: (i) *If $\lambda_2 \geq \lambda_1$, then $U^k(i)$ is strictly increasing in i*; (ii) *if $\lambda_2 \leq \lambda_1$, then $D(i)$ and $\widetilde{D}(i)$ are strictly decreasing at all states $i \notin I^*$*; (iii) *for any $\lambda_1 \geq 1$ and $\lambda_2 \geq 1$, the states i_1 and i_2 are unique*; *$D(i)$ is decreasing at $i < i_1$ and increasing at $i > i_1$, and $\widetilde{D}(i)$ is increasing at $i < i_2$ and decreasing at $i > i_2$.*

PROOF: For (i), it suffices to prove that if $\lambda_1 \leq \lambda_2$, then the ith ratio in $U^k(i)$ is strictly increasing: that is, that (a) x_i/y_i is increasing if $i \leq i^* - 1$, (b) $x_{i^*-1}/y_{i^*-1} \leq x_{i^*}/\frac{\alpha_M}{\sigma^S_{1,2}} u_{i^*}$, (c) x_i/u_i is increasing in i if $i^* \leq i \leq i_0$, and (d) v_i/u_{i+1} is increasing if $i \geq i_0$. By (C.5),

$$\frac{\mu_1^H x_i}{y_i} = \left(\frac{\lambda_2}{\lambda_1}\right)^{i-1} \left(\frac{\xi(1)\lambda^{i-1} - 1}{\lambda^{i-1} - 1}\right), \qquad \frac{x_i}{u_i} = \frac{\lambda_1^{M+1-2i}}{\mu_1^H} \left(\frac{\xi(1)\lambda^{i-1} - 1}{\lambda^{M-i} - 1}\right),$$

$$\frac{v_i}{u_{i+1}} = \frac{1}{\lambda_1 \mu_S^L} \left(\frac{\lambda_1}{\lambda_2}\right)^{M-i} \left(\frac{\lambda^{M-i-1}\dfrac{1}{\xi(1)} - 1}{\lambda^{M-i-1} - 1}\right).$$

The expression for x_i/y_i is increasing by $\lambda_2 \geq \lambda_1$, which implies that $(\lambda_2/\lambda_1)^{i-1}$ is increasing, and by $\xi(1) < 1 < \lambda$, which implies that the final term is increasing. The expression for v_i/u_{i+1} is increasing by $\lambda_1 \leq \lambda_2$, which implies that $(\lambda_1/\lambda_2)^{M-i}$ is decreasing in $M-i$ and hence increasing in i, and by $\xi(1) < 1$, which implies that the final expression is larger than $1/\xi(1)$ and decreasing (toward $1/\xi(1)$) in $M-i-1$, hence increasing in i. The expression for x_i/u_i is increasing since $\lambda > 1$ implies that $1/(\lambda^{M-i}-1)$ is increasing in i, while $\lambda_1^{M+1-2i}(\xi(1)\lambda^{i-1}-1)$ is increasing, via straightforward calculation, by $\lambda = \lambda_1\lambda_2 > \lambda_1^2 > 1$. And finally, to show that $x_{i^*-1}/y_{i^*-1} \leq x_{i^*}/\frac{\alpha_M}{\sigma_{1,2}^S}u_{i^*}$, by (C.5) (for the first equality) and (C.7),

$$\text{(C.9)} \qquad \frac{\alpha_M}{\sigma_{1,2}^S}u_{i^*} < \left(\frac{\lambda_1}{\lambda_2}\right)^{(M-1)/2}\lambda^{(M+1)/2-i^*}\left(\frac{\lambda^{i^*-1}-1}{\lambda^{M-i^*}-1}\right)\frac{\lambda^{M-i^*}-1}{\lambda_1^{M-i^*}}$$
$$= \left(\frac{\lambda^{i^*-1}-1}{\lambda_2^{i^*-1}}\right).$$

Thus, using the expression for x_i^* from (C.5) and simplifying gives

$$\frac{x_{i^*}}{\frac{\sigma_{M,M-1}^1}{\sigma_{1,2}^S}u_{i^*}} \geq \frac{1}{\mu_1^H}\left(\frac{\lambda_2}{\lambda_1}\right)^{i^*-1}\left(\frac{\xi(1)\lambda^{i^*-1}-1}{\lambda^{i^*-1}-1}\right).$$

The RHS of this expression is strictly increasing in i^* by $\lambda_2 \geq \lambda_1$ and $\xi(1) < 1 < \lambda$; in particular, it strictly exceeds the value that would obtain if i^* were replaced by i^*-1, which is precisely x_{i^*-1}/y_{i^*-1}.

For (iii), it suffices to prove that once the ith ratio in $D(i)$ begins to increase, it continues to increase, and that once the ith ratio in $\widetilde{D}(i)$ begins to decrease, it continues to decrease. That is, (a) if $x_i/y_{i-1} < x_{i+1}/y_i$, then $x_{i+1}/y_i < x_{i+2}/y_{i+1}$; (b) if $x_{i^*-1}/y_{i^*-2} < x_{i^*}/y_{i^*-1}$, then $x_{i^*}/y_{i^*-1} \leq \frac{\sigma_{1,2}^S}{\alpha_M}x_{i^*+1}/u_{i^*}$; (c) x_i/u_{i-1} is increasing; and (d) if $v_{i-1}/u_{i-1} > v_i/u_i$, then $v_i/u_i > v_{i+1}/u_{i+1}$.[3] For (a), by (C.5),

$$\text{(C.10)} \qquad \frac{x_i}{y_{i-1}} < \frac{x_{i+1}}{y_i} \quad \Leftrightarrow \quad \frac{\lambda_2}{\lambda_1}\frac{\xi(S)\lambda^{i-1}(\lambda^{i-2}-1)-(\lambda^{i-2}-1)}{\xi(S)\lambda^{i-2}(\lambda^{i-1}-1)-(\lambda^{i-1}-1)} > 1.$$

The derivative of the LHS with respect to (w.r.t.) i has the same sign as $(\lambda-1)(\xi(S)-1)(\xi(S)\lambda^{2i-3}-1)\ln\lambda$, which is positive for $i \geq 2$ by $\lambda > \xi(S) > 1$.

[3]This uses the fact that as long as a_l/b_l is decreasing, $X(j) \equiv \sum_{l=2}^j a_l/\sum_{l=2}^j b_l$ is decreasing. Thus, $X(j) < X(j+1) \Rightarrow a_j/b_j < a_{j+1}/b_{j+1}$. Now suppose it holds that $a_j/b_j < a_{j+1}/b_{j+1} \Rightarrow a_{j+1}/b_{j+1} < a_{j+2}/b_{j+2}$: that is, that once the jth ratio begins to increase, it continues to increase. Then, if $X(j) < X(j+1)$, it follows that $X(j) < X(j+1) < a_{j+1}/b_{j+1} < a_{j+2}/b_{j+2}$, which is precisely the condition under which $X(j+1) < X(j+2)$. Thus, once $X(j)$ begins to increase, it continues to increase; while if decreasing at j, it is also decreasing at $j-1$.

So if the condition holds (i.e., x_l/y_l is increasing) at $l=i$, then it holds also at all $l \geq i$. For (c), by (C.5),

$$\frac{x_i}{u_{i-1}} < \frac{x_{i+1}}{u_i} \quad \Leftrightarrow \quad \left(\frac{\xi(1)\lambda^{i-1}-1}{\xi(1)\lambda^{i}-1}\right)\left(\frac{\lambda^{M-i}-1}{\lambda^{M+1-i}-1}\right)\lambda_1^2 < 1.$$

Since $\lambda > 1$, each of the first and second LHS terms are below $1/\lambda$; then the LHS is below $\lambda_1^2/\lambda^2 = 1/\lambda_2^2$, which is below 1 (as desired) provided $\lambda_2 > 1 \Leftrightarrow \mu_1^L \geq \mu_S^L$. (Note that this is implied by $\lambda_2 \geq \lambda_1$, which yields $\lambda_2^2 \geq \lambda_1\lambda_2 = \lambda > 1$.) For (d), by (C.5), we have (after simplifying)

$$\begin{aligned} &\frac{v_i}{u_i} > \frac{v_{i+1}}{u_{i+1}} \\ &\quad\Leftrightarrow \quad \left(\frac{\lambda_1}{\lambda_2}\right)\left(\frac{\frac{1}{\xi(S)}\lambda^{M-i}(\lambda^{M-i-1}-1)-(\lambda^{M-i-1}-1)}{\frac{1}{\xi(S)}\lambda^{M-i-1}(\lambda^{M-i}-1)-(\lambda^{M-i}-1)}\right) > 1.\end{aligned}$$

The LHS is increasing in i (the derivative w.r.t. i is proportional to $(\lambda-1)\times(\xi(S)-1)\ln\lambda$, which is positive by $\lambda > \xi(S) > 1$), and so the condition is easier to satisfy in larger states. Therefore, as desired, if v_l/u_l is decreasing at $l=i$, then it is decreasing also at $l \geq i$. Finally, for (b), for $x_{i^*}/y_{i^*-1} \leq \frac{\sigma_{1,2}^S}{\alpha_M} x_{i^*+1}/u_{i^*}$, it suffices that the following inequality holds (the LHS is the upper bound from (C.9) on $\alpha_M u_i^*/\sigma_{1,2}^S$):

$$\text{(C.11)} \qquad \begin{aligned} &\left(\frac{\lambda^{i^*-1}-1}{\lambda_2^{i^*-1}}\right) < \frac{1}{\lambda_1}\frac{\xi(1)\lambda^{i^*}-1}{\xi(1)\lambda^{i^*-1}-1}\frac{\lambda^{i^*-2}-1}{\lambda_2^{i^*-2}} \equiv \frac{x_{i^*+1}}{x_{i^*}}y_{i^*-1} \\ &\quad\Leftrightarrow \quad \frac{\lambda_1}{\lambda_2}\left(\frac{\lambda^{i^*-1}-1}{\lambda^{i^*-2}-1}\right)\frac{\xi(1)\lambda^{i^*-1}-1}{\xi(1)\lambda^{i^*}-1} < 1.\end{aligned}$$

The LHS expression in (C.11) is decreasing in i^* (the derivative is proportional to $(1-\lambda)(\xi(1)\lambda^{2(i^*-1)}-1)(\xi(S)-1)\ln\lambda$, negative $\forall i^* \geq 2$ by $\lambda > \xi(S) > 1$). Therefore, a sufficient condition for (C.11) is that it holds when i^* is replaced by i^*-1, which, by (C.5), is precisely the condition for $x_{i^*-1}/y_{i^*-2} < x_{i^*}/y$.

Finally, for part (ii), it suffices to prove that if $i \leq i^*-2$, then $\lambda_1 \geq \lambda_2$ implies that the ith ratio in $\underset{\sim}{D}(i)$, x_i/y_{i-1}, is strictly decreasing, and that if $i \geq i_0+1$, then the ith ratio in $\widetilde{D}(i)$, v_i/u_i, is strictly decreasing. For $i \geq i_0$, (C.5) yields

$$\frac{v_i}{u_i} = \left(\frac{\lambda_1}{\lambda_2}\right)^{M-i}\frac{\lambda^{M-i}\frac{1}{\xi(S)}-1}{\lambda^{M-i}-1}.$$

Since $\lambda_1 \geq \lambda_2$, the first term $(\lambda_1/\lambda_2)^{M-i}$ is increasing in $M-i$, hence decreasing in i, and the second term is also increasing in $M-i$ by $1/\xi(S) < 1 < \lambda$, hence decreasing in i, as desired to establish that v_i/u_i is decreasing. For $i \leq i^*-1$, (C.10) yields

$$\frac{x_i}{y_{i-1}} = \frac{1}{\lambda_1 \mu_1^H}\left(\frac{\lambda_2}{\lambda_1}\right)^{i-2} \frac{\xi(S)\lambda^{i-2}-1}{\lambda^{i-2}-1}.$$

The first term is decreasing by $\lambda_2/\lambda_1 \leq 1$, and the second term is greater than $\xi(S)$ and decreasing toward $\xi(S)$ as i increases. As desired, x_i/y_{i-1} is then strictly decreasing. *Q.E.D.*

STEP C.2: *There exists η^* such that in an optimal protocol with $\eta < \eta^*$,* (i) *all interior states are fluid up if $\lambda_1 \leq \lambda_2$;* (ii) *all interior states outside I^* are fluid down if $\lambda_1 \geq \lambda_2$.*

PROOF: First suppose, by contradiction, that some state $i \leq i_0 - 1$ is sticky up. Choose the smallest such i and let k be the largest state below i that is sticky down, defining $k = 2$ if all states below i are fluid down. Then, using the Step 3 (from Section A.9 in the paper) result that upward stickiness in state i implies $\sigma^1_{i,i-1} = \sigma^1_{i+1,i} = 1$, we have $\alpha_l = \prod_{j=3}^{l} \sigma^1_{j,j-1}/\sigma^S_{j-1,j} = \alpha_k \ \forall k+1 \leq l \leq i$, $\alpha_{i+1} = \alpha_k/\sigma^S_{i,i+1}$. If $3 \leq k < i \leq i^*-1$, then by (C.1), the optimality conditions for $\sigma^1_{k,k-1} \in (0,1)$ (the first expression below) and $\sigma^S_{i,i+1} \in (0,1)$ (the second expression below) are

$$\frac{\sum_{l=2}^{k} \alpha_l x_l/\sigma^1_{l,l-1}}{\sum_{l=2}^{k-1} y_l/\alpha_l} = \frac{(\sigma^S_{1,2})^2 + o(\eta)}{\eta} = \frac{\sum_{l=2}^{k} \frac{\alpha_l x_l}{\sigma^1_{l,l-1}} + \sum_{l=k+1}^{i} \frac{\alpha_l x_l}{\sigma^1_{l,l-1}}}{\sum_{l=2}^{k-1} \frac{y_l}{\alpha_l} + \sum_{l=k}^{i} \frac{y_l}{\alpha_l}}$$

$$\Rightarrow \quad \frac{(\sigma^S_{1,2})^2 + o(\eta)}{\eta} = \frac{\sum_{l=k+1}^{i} \frac{\alpha_l x_l}{\sigma^1_{l,l-1}}}{\sum_{l=k}^{i} \frac{y_l}{\alpha_l}} = \alpha_k^2 \frac{\sum_{l=k+1}^{i} x_l}{\sum_{l=k}^{i} y_l}.$$

The final term in the above expression is precisely $(\alpha_k^2)U^k(i)$. Similarly, if $i \geq i^*$, the optimality condition for $\sigma^S_{i,i+1} \in (0,1)$ (the first expression in (C.2)) reduces, using the optimality condition for $\sigma^1_{k,k-1} \in (0,1)$ and the fact that $\alpha_l = \alpha_k \ \forall k+1 \leq l \leq i$, to $(\sigma^S_{1,2})^2/\eta + o(\eta) = (\alpha_k)^2 U^k(i)$. On the other hand, consider the optimality condition for $\sigma^S_{i+1,i+2} > 0$. If $i \leq i^*-2$, then the first expression in (C.1) evaluated at $i+1$ reduces (using the optimality condition for $\sigma^1_{k,k-1} \in (0,1)$) to

$$\frac{(\sigma^S_{1,2})^2 + o(\eta)}{\eta} \geq \alpha_k^2 \frac{\sum_{l=k+1}^{i} x_l + x_{i+1} \Big/ \sigma^S_{i,i+1}}{\sum_{l=k}^{i} y_l + \sigma^S_{i,i+1} y_{i+1}} > \alpha_k^2 \frac{\sum_{l=k+1}^{i+1} x_l}{\sum_{l=k}^{i+1} y_l} \equiv \alpha_k^2 U^k(i+1)$$

(the final inequality by the fact that the second term is trivially decreasing in $\sigma^S_{i,i+1}$). By a similar calculation if $i \geq i^* - 1$, we conclude that optimality of $\sigma^S_{i+1,i+2} > 0$ requires the first inequality below and optimality of $\sigma^S_{i,i+1} \in (0,1)$ requires the equality:

$$\alpha_k^2 U^k(i+1) < \frac{(\sigma^S_{1,2})^2 + o(\eta)}{\eta} = \alpha_k^2 U^k(i).$$

This is a contradiction, since $U^k(i+1) > U^k(i)$ by Step C.1 part (i). If $k=2$, the analysis is identical, except that the numerator and denominator sums in $U^k(i)$ both begin at 2 (rather than $k+1$ and k, respectively).

The proof that states $i \geq i_0 + 1$ are fluid up is nearly identical: if not, then let i be the largest sticky-up state above i_0, and let k be the smallest sticky-down state above i. Using (C.3), the optimality condition for $\sigma^S_{i,i+1} \in (0,1)$ reduces, using the condition for $\sigma^1_{k,k-1} \in (0,1)$, to $(\sigma^1_{M,M-1})^2/\eta \to U^k(i)$, while the condition for $\sigma^S_{i-1,i} > 0$ requires $(\sigma^1_{M,M-1})^2/\eta < U^k(i-1)$—a contradiction, since $U^k(i) > U^k(i-1)$ by Step C.1.

By a symmetric argument, states outside I^* are fluid down when $\lambda_2 \leq \lambda_1$. For example, suppose some state $i \leq i^* - 1$ is sticky down and choose the smallest such i. Then all states below i are fluid down; if also all lower states are fluid up, then, by (C.1), the optimality condition for $\sigma^1_{i,i-1} > 0$ requires $\frac{(\sigma^S_{1,2})^2 + o(\eta)}{\eta} < D(i)$, with equality if $\sigma^1_{i,i-1} \in (0,1)$; moreover, if i is downward sticky, then the upper bound (from (C.1)) on $(\sigma^S_{1,2})^2/\eta$ in the optimality condition for $i+1 \downarrow_1 i$ decreases. Thus, together, the optimality conditions for $\sigma^1_{i,i-1} \in (0,1)$ and $\sigma^1_{i+1,i} > 0$ demand $D(i) = (\sigma^S_{1,2})^2/\eta + o(\eta) < D(i+1)$—a contradiction, since $D(i)$ is strictly decreasing at $i \leq i^* - 1$ by Step C.1 part (iii). The argument is similar if there is a sticky-up state $k < i$ (choose the lowest such i and define a function $D^k(i)$ that is identical to $D(i)$, except that the numerator and denominator sums begin at $k+1$ rather than 2). And similarly for states $i \geq i_0$, where $(\sigma^1_{M,M-1})^2/\eta$ must exceed a cutoff that increases as the state decreases: if the condition holds for state i to be indifferent about moving down, then state $i-1$ will not move down at all. *Q.E.D.*

STEP C.3: *Let $\lambda_1 \leq \lambda_2$. There exists η^* such that for any optimal protocol with $\eta < \eta^*$,* (i) *all states $i < i_1$ are fluid down*; (ii) *all states $i > i_2$ are fluid down*.

PROOF: For (i), suppose that some state $i < i_1$ is sticky down and choose the smallest such i. By Step C.2, no interior states are sticky up and so $\alpha_l = 1$ $\forall l \leq i-1$, while $\alpha_i = \sigma^1_{i,i-1}$ and $\alpha_{i+1} = \sigma^1_{i+1,i}\sigma^1_{i,i-1}$. Then by (C.1), optimality of $\sigma^1_{i,i-1} \in (0,1)$ requires the equality below, and optimality of $\sigma^1_{i+1,i} > 0$ requires the final inequality:

$$D(i) \equiv \sum_{l=2}^{i} x_l \Big/ \sum_{l=2}^{i-1} y_l = \frac{(\sigma^S_{1,2})^2 + o(\eta)}{\eta} \leq \left(\frac{\sum_{l=2}^{i} x_l + \sigma^1_{i,i-1} x_{i+1}}{\sum_{l=2}^{i-1} y_l + y_i \Big/ \sigma^1_{i,i-1}} \right). \tag{C.12}$$

But the RHS expression in (C.12) is increasing in $\sigma^1_{i,i-1}$ (since $x_{i+1} > 0$ and $y_i > 0$) and therefore reaches a maximum value at $\sigma^1_{i,i-1} = 1$, of $D(i+1)$. But since $i < i_1$ implies that $D(i)$ is decreasing at i by Step C.1(ii), we have $D(i) > D(i+1)$, contradicting (C.12). So no state below i_1 is sticky down in an optimal protocol.

The argument for states above i_2 is nearly identical: if some $i > i_2$ is downward sticky (choose the largest such i), then optimality for $\sigma^1_{i,i-1} \in (0,1)$ requires $(\sigma^2_{M,M-1})/\eta \to \widetilde{D}(i)$, while optimality of $\sigma^1_{i-1,i-2} > 0$ requires $(\sigma^1_{M,M-1})^2/\eta > \widetilde{D}(i-1)$—a contradiction, since $\widetilde{D}(i)$ is decreasing at $i \geq i_2$ by Step C.1(ii). *Q.E.D.*

STEP C.4: *Let $\lambda_1 \leq \lambda_2$. There exists η^* such that in an optimal protocol with $\eta < \eta^*$,* (i) *if $D(i) < D(i+1)$ and $i+1$ is sticky down, then i is sticky down*; (ii) *if $\widetilde{D}(i) < \widetilde{D}(i+1)$ and i is sticky down, then $i+1$ is sticky down.*

PROOF: For (i), suppose (by contradiction) that there is an optimal protocol such that for some state i, $D(i) < D(i+1)$, $\sigma^1_{i+1,i} \in (0,1)$, and $\sigma^1_{i,i-1} = 1$, and choose the smallest $i \geq i_1$. By Step C.2, all states below i are fluid up. If they are also fluid down, then optimality of $\sigma^1_{i,i-1} > 0$ demands the inequality below, while optimality of $\sigma^1_{i+1,i} \in (0,1)$ demands the equality below (using (C.1) if $i \leq i^* - 1$ and (C.2) if $i^* \leq i \leq i_0$, and noting that in a protocol with no mixing below $i+1$, the RHS expressions in these equations equal $D(i)$):

$$D(i+1) = \frac{(\sigma^S_{1,2})^2 + o(\eta)}{\eta} \leq D(i).$$

This is a contradiction, since $D(i) < D(i+1)$. If there is a state $k < i$ that is downward sticky, the argument is similar, choosing the largest such k (we discuss the analogous situation more carefully in (ii); see below). For (ii), suppose (by contradiction) that there is an optimal protocol such that for some state $i < i_2$, $\widetilde{D}(i) < \widetilde{D}(i+1)$, $\sigma^1_{i,i-1} \in (0,1)$, and $\sigma^1_{i+1,i} = 1$, and choose the largest

$i < i_2$. By (C.6) and (C.3), optimality of $\sigma^1_{i,i-1}$ demands the equality below, and optimality of $\sigma^1_{i+1,i} = 1$ demands the inequality below, using the Step C.2 result that all states are fluid up:[4]

$$\text{(C.13)}\qquad \frac{\left(\dfrac{\alpha_M}{\sigma^S_{1,2}}\right)\displaystyle\sum_{l=i+2}^{i_0}\alpha_{l-1}x_l + \sum_{l=i_0}^{M-1}\alpha_l v_l}{\displaystyle\sum_{l=i+1}^{M-1}\frac{u_l}{\alpha_l}} \le \frac{(\sigma^1_{M,M-1})^2 + o(\eta)}{\eta}$$

$$= \frac{\left(\dfrac{\alpha_M}{\sigma^S_{1,2}}\right)\displaystyle\sum_{l=i+1}^{i_0}\alpha_{l-1}x_l + \sum_{l=i_0}^{M-1}\alpha_l v_l}{\displaystyle\sum_{l=i}^{M-1}\frac{u_l}{\alpha_l}}.$$

By Step C.2, all states above i are fluid up. If also all states above i are fluid down, then $\alpha_l = \alpha_i\ \forall l \ge i$ and so the LHS of the above expression is precisely $\alpha_i^2\widetilde{D}(i+1)$, while the RHS is $\alpha_i^2\widetilde{D}(i)$; but since $\widetilde{D}(i) < \widetilde{D}(i+1)$, the inequality cannot hold—a contradiction. If some state $k \ge i_0 \ge i+2$ is sticky down, the argument is nearly identical, choosing the smallest such state k and substituting the optimality condition for $\sigma^1_{k,k-1} \in (0,1)$ into (C.13). For example, if $k \ge i_0$, the optimality of $\sigma^1_{k,k-1} \in (0,1)$ requires that the second expression in (C.13) holds at $i = k$. Then (C.13) reduces to an expression that is identical, except that all sums ending at $M-1$ are now truncated at $k-1$, and since (by construction) there is no mixing between i and k, so $\alpha_l = \alpha_i\ \forall i \le l \le k-1$, (C.13) reduces further to

$$\alpha_i^2\frac{\left(\dfrac{\alpha_M}{\sigma^S_{1,2}}\right)\displaystyle\sum_{l=i+2}^{i_0}x_l + \sum_{l=i_0}^{k-1}v_l}{\displaystyle\sum_{l=i+1}^{k-1}u_l} \le \frac{(\sigma^1_{M,M-1})^2 + o(\eta)}{\eta}$$

$$= \alpha_i^2\frac{\left(\dfrac{\alpha_M}{\sigma^S_{1,2}}\right)\displaystyle\sum_{l=i+1}^{i_0}x_l + \sum_{l=i_0}^{k-1}v_l}{\displaystyle\sum_{l=i}^{k-1}u_l}$$

[4]This is the expression for $i \le i_0 - 1$. If $i \ge i_0$, delete the first numerator sum in both the LHS and RHS bounds, and begin the second numerator sums (both LHS and RHS) at $l = i$.

$$\Rightarrow \quad \alpha_i^2 \frac{\left(\dfrac{\alpha_M}{\sigma_{1,2}^S}\right)\displaystyle\sum_{l=i+2}^{i_0} x_l + \sum_{l=i_0}^{k-1} v_l + \sum_{l=k}^{M-1} v_l}{\displaystyle\sum_{l=i+1}^{k-1} u_l + \sum_{l=k}^{M-1} u_l}$$

$$\leq \frac{(\sigma_{M,M-1}^1)^2 + o(\eta)}{\eta} = \alpha_i^2 \frac{\left(\dfrac{\alpha_M}{\sigma_{1,2}^S}\right)\displaystyle\sum_{l=i+1}^{i_0} x_l + \sum_{l=i_0}^{k-1} v_l + \sum_{l=k}^{M-1} v_l}{\displaystyle\sum_{l=i}^{k-1} u_l + \sum_{l=k}^{M-1} u_l}.$$

But since the LHS is $\alpha_i^2 \widetilde{D}(i+1)$ and the RHS is $\alpha_i^2 \widetilde{D}(i)$, and we have assumed $\widetilde{D}(i) < \widetilde{D}(i+1)$, this expression cannot hold—a contradiction. *Q.E.D.*

STEP C.5: *If* $\lambda_1 \leq \lambda_2$ *and* $D(i_1) \leq \widetilde{D}(i_2)$, *then all states in* $\{i_1, \dots, i_2\}$ *are sticky down*; *a sufficient condition for* $D(i_1) < \widetilde{D}(i_2)$ *is* $i_2 \geq i_1 + 2$.

PROOF: By Step C.2, $\lambda_1 \leq \lambda_2$ implies that all interior states are fluid up, and by Step C.3, all states $i \notin \{i_1, \dots, i_2\}$ are sticky down. Therefore, optimality of $i_1 \downarrow_1 i_1 - 1$ demands $(\sigma_{1,2}^S)^2/\eta + o(\eta)/\eta \leq D(i_1)$, while optimality of $i_2 \downarrow_1 i_2 - 1$ demands $(\sigma_{M,M-1}^1)^2/\eta + o(\eta) \geq \widetilde{D}(i_2)$ (using the corresponding equations in (C.1), (C.2), (C.6), or (C.3), using the fact that all states outside $\{i_1, \dots, i_2\}$ are fluid both up and down). Together, for η sufficiently near zero, these inequalities imply

$$\text{(C.14)} \qquad \frac{(\sigma_{M,M-1}^1)^2}{(\sigma_{1,2}^S)^2} \geq \frac{\widetilde{D}(i_2)}{D(i_1)}.$$

But by Step C.4, if *any* state in $\{i_1, \dots, i_2\}$, is sticky down, then all states in this set are sticky down. Therefore, it suffices to rule out a protocol with no mixing, in which $\alpha_M/\sigma_{1,2}^S = \sigma_{M,M-1}^1/\sigma_{1,2}^S$ (and $\alpha_M/\sigma_{1,2}^S \geq \sigma_{M,M-1}^1/\sigma_{1,2}^S$ if $\lambda_1 \geq \lambda_2$, so that some states are potentially upward sticky). But then the second condition in (C.4) demands

$$\left(\frac{\sigma_{M,M-1}^1}{\sigma_{1,2}^S}\right)^2 \leq \left(\frac{\alpha_M}{\sigma_{1,2}^S}\right)^2 \to \left(\frac{\lambda_1}{\lambda_2}\right)^{M-1} \left(\frac{\lambda^{(M-1)/2} - \sqrt{\beta}}{\sqrt{\beta}\lambda^{(M-1)/2} - 1}\right)^2.$$

For η sufficiently near zero, this contradicts (C.14) whenever the following inequality holds:[5]

$$\text{(C.15)}\qquad \frac{\widetilde{D}(i_2)}{D(i_1)} > \left(\frac{\lambda_1}{\lambda_2}\right)^{M-1}\left(\frac{\lambda^{(M-1)/2}-\sqrt{\beta}}{\sqrt{\beta}\lambda^{(M-1)/2}-1}\right)^2.$$

Thus, if (C.15) holds, then a protocol with no (downward) mixing cannot be optimal and so all states in $\{i_1,\dots,i_2\}$ are sticky down. For $\lambda_1 \le \lambda_2$ and $\beta > 1$, the RHS expression in (C.15) is below 1 and so the expression holds whenever $D(i_1) \le \widetilde{D}(i_2)$. This proves the first assertion. And for the second, assume $i_2 - 1 \ge i^*$ and define $a_i = x_i/\frac{\alpha_M}{\sigma^S_{1,2}}u_{i-1}\ \forall i \in \{i^*,\dots,i_0\}$. By definition of i_2 and Step C.1(iii), we know that $\widetilde{D}(i_2) > \widetilde{D}(i_2-1)$, which implies that $\widetilde{D}(i_2)$ exceeds the (i_2-1)st ratio in $\widetilde{D}(i_2-1)$:

$$\frac{(\sigma^1_{M,M-1})^2 + o(\eta)}{\eta} \ge \widetilde{D}(i_2) > \frac{\alpha_M}{\sigma^S_{1,2}}x_{i_2}\big/u_{i_2-1}.$$

By definition of i_1 and Step C.1(iii), we also have $D(i_1) < D(i_1+1)$, which implies that $D(i_1)$ is below the (i_1+1)st ratio in $D(i_1+1)$:

$$\frac{(\sigma^S_{1,2})^2 + o(\eta)}{\eta} \le \widetilde{D}(i_1) < x_{i_1+1}\Big/\frac{\alpha_M}{\sigma^S_{1,2}}u_{i_1}.$$

Together, these inequalities demand the following condition for η near zero:

$$\left(\frac{\alpha_M}{\sigma^S_{1,2}}\right)^2 \ge \frac{\frac{\alpha_M}{\sigma^S_{1,2}}x_{i_2}\big/u_{i_2-1}}{x_{i_1+1}\Big/\frac{\alpha_M}{\sigma^S_{1,2}}u_{i_1}} \quad\Rightarrow\quad x_{i_1+1}/u_{i_1} \ge x_{i_2}/u_{i_2-1}.$$

This is a contradiction if $i_1 + 1 < i_2$, since x_l/u_{l-1} is increasing by Step C.1(iii). *Q.E.D.*

C.3. *Specializing to Symmetric Signals*

STEP C.6: *For a symmetric signal*, $i_1 \le i^* + 1$.

[5]This is, in fact, a necessary and sufficient condition for there to be a downward-sticky block of interior states for any $\lambda_1 \ge 1$ and $\lambda_2 \ge 1$.

PROOF: To show that $i_1 \leq i^* + 1$, it suffices to prove that $D(i^* + 1) \leq D(i^* + 2)$, which holds iff $D(i^* + 1)$ (the LHS below, by (C.5)) is smaller than the final ratio in $D(i^* + 2)$ (the RHS below):

$$\text{(C.16)} \quad \frac{\sum_{l=2}^{i^*+1} x_l}{\sum_{l=2}^{i^*-1} y_l + \frac{\alpha_M}{\sigma_{1,2}^S} u_{i^*}} < \frac{x_{i^*+2}}{\frac{\alpha_M}{\sigma_{1,2}^S} u_{i^*+1}} \quad \Leftrightarrow \quad \frac{\sum_{l=2}^{i^*+1} x_l}{\frac{\sigma_{1,2}^S}{\alpha_M} \sum_{l=2}^{i^*-1} y_l + u_{i^*}} < \frac{x_{i^*+2}}{u_{i^*+1}}.$$

By (C.5) and $\lambda/\lambda_1 = \lambda_2$,

$$\text{(C.17)} \quad \sum_{l=2}^{i^*+1} x_l = \frac{1}{\mu_1^H} \left(\xi(1) \sum_{l=2}^{i^*+1} \lambda_2^{l-1} - \sum_{l=2}^{i^*+1} \lambda_1^{1-l} \right)$$

$$= \frac{1}{\mu_1^H} \left(\xi(1) \lambda_2 \frac{\lambda_2^{i^*} - 1}{\lambda_2 - 1} - \frac{\lambda_1^{i^*} - 1}{\lambda_1^{i^*}(\lambda_1 - 1)} \right)$$

$$\text{(C.18)} \quad = \frac{1}{\mu_1^H} \frac{1}{\lambda_1^{i^*}} \left(\lambda_1^{i^*} - 1 \right) \left(\frac{\lambda_1^{i^*} - 1}{\lambda_1 - 1} \right) \quad \text{if} \quad \lambda_1 = \lambda_2.$$

Also by (C.5),

$$\text{(C.19)} \quad \frac{x_{i^*+2}}{u_{i^*+1}} = \frac{1}{\mu_1^H} \frac{\xi(S)\lambda^{i^*} - 1}{\lambda_1^{i^*+1}} \frac{\lambda_1^{M-i^*-1}}{\lambda^{M-i^*-1} - 1},$$

$$\text{(C.20)} \quad u_{i^*} = \frac{\lambda^{M-i^*} - 1}{\lambda_1^{M-i^*}},$$

$$\text{(C.21)} \quad \sum_{l=2}^{i^*-1} y_l = \sum_{l=2}^{i^*-1} \left(\lambda_1^{l-1} - \lambda_2^{1-l} \right) = \frac{\lambda_1^{i^*-1} - \lambda_1}{\lambda_1 - 1} - \frac{\lambda_2^{i^*-2} - 1}{\lambda_2^{i^*-2}(\lambda_2 - 1)}.$$

And by (C.7),

$$\text{(C.22)} \quad \frac{\sigma_{1,2}^S}{\alpha_M} \geq \left(\frac{\lambda_2}{\lambda_1} \right)^{(M-1)/2} \frac{1}{\lambda^{(M+1)/2-i^*}} \frac{\lambda^{M-i^*} - 1}{\lambda^{i^{*-1}} - 1}.$$

Then, in the second inequality in (C.16), the LHS of (C.16) has the following upper bound, by (C.22), (C.18), (C.21), and (C.20):

$$\frac{\sum_{l=2}^{i^*+1} x_l}{\frac{\sigma_{1,2}^S}{\alpha_M}\sum_{l=2}^{i^*-1} y_l + u_{i^*}} \leq \frac{1}{\mu_1^H}\frac{1}{\lambda_2^{i^*}}\left(\frac{\mu_1^H}{\mu_S^L}\lambda_2^{i^*}-1\right)\left(\frac{\lambda_2^{i^*}-1}{\lambda_2-1}\right)$$

$$\Big/\left(\left(\frac{\lambda_2}{\lambda_1}\right)^{(M-1)/2}\frac{1}{\lambda^{(M+1)/2-i^*}}\frac{\lambda^{M-i^*}-1}{\lambda^{i^*-1}-1} \times \frac{(\lambda_1^{i^*-2}-1)(\lambda_1^{i^*-1}-1)}{\lambda_1^{i^*-2}(\lambda_1-1)} + \frac{\lambda^{M-i^*}-1}{\lambda_1^{M-i^*}}\right).$$

For a symmetric signal, with $\lambda_2 = \lambda_1$ and $\mu_1^H = \mu_S^L$, this simplifies to

$$\frac{1}{\mu_1^H}\frac{1}{\lambda_2^{i^*}}\big(\lambda_2^{i^*}-1\big)\big(\lambda_2^{i^*-1}+1\big)\frac{\lambda_1^{M-i^*}}{(\lambda^{M-i^*}-1)} < \frac{1}{\mu_1^H}\frac{1}{\lambda_2^{i^*}}\big(\lambda_2^{2i^*}-1\big)\frac{\lambda_2^{M-i^*}}{(\lambda^{M-i^*}-1)}.$$

This is less than the expression in (C.19), evaluated at $\xi(S) = \lambda_1 = \lambda_2$, whenever

$$\frac{1}{\mu_1^H}\frac{1}{\lambda_2^{i^*}}\big(\lambda_2^{2i^*}-1\big)\frac{\lambda_2^{M-i^*}}{(\lambda^{M-i^*}-1)} < \frac{1}{\mu_1^H}\frac{\lambda_2^{2i^*+1}-1}{\lambda_2^{i^*+1}}\frac{\lambda_2^{M-i^*-1}}{\lambda^{M-i^*-1}-1}$$

$$\Leftrightarrow \quad \frac{(\lambda_2^{2i^*}-1)}{(\lambda_2^{2i^*+1}-1)}\lambda_2^2\left(\frac{\lambda^{M-i^*-1}-1}{\lambda^{M-i^*}-1}\right) < 1.$$

The first LHS term is below $1/\lambda_2$ (by $\lambda_2 > 1$) and the final term (by $\lambda > 1$) is below $1/\lambda$. Thus, the LHS in the above expression is below $\lambda_2/\lambda = 1/\lambda_2 < 1$, as desired. *Q.E.D.*

STEP C.7: *For a symmetric signal*, $i_2 \geq i_0$ *for* β *near* $\xi(S)\lambda^{i_0-(M+1)/2}$ *and* $i_2 \geq i_0 - 1$ *for* β *near* $\xi(S)\lambda^{i_0-(M+3)/2}$.

PROOF: By (20) in the paper, the DM starts in a state i_0 satisfying $\xi(S) \times \lambda^{i_0-(M+1)/2} > \sqrt{\beta} > \xi(S)\lambda^{i_0-(M+3)/2}$. At the top end of this range, the DM is indifferent between starting in i_0 and $i_0 + 1$; here, symmetric signals yield $i_2 \geq i_0$ (so that i_0 is downward sticky), as I will prove below. At the bottom end of the range, the DM is close to indifferent between starting in i_0 and $i_0 - 1$, at which point i_0 is fluid down, and i_2 only exceeds (in fact equals) $i_0 - 1$. The two calculations are nearly identical, so I prove only the first assertion in the Step C.7 statement.

To this end, to prove that $\widetilde{D}(i)$ is maximized at $i_2 \geq i_0$, it suffices, by Step C.1, to prove that $\widetilde{D}(i_0 - 1) < \widetilde{D}(i_0)$, which holds iff the first ratio (for $i = i_0 - 1$) in $\widetilde{D}(i_0 - 1)$ is smaller than $\widetilde{D}(i_0)$:

$$\frac{\alpha_M}{\sigma^S_{1,2}} x_{i_0} \Big/ u_{i_0-1} < \sum_{l=i_0}^{M-1} v_l \Big/ \sum_{l=i_0}^{M-1} u_l.$$

By (C.5), for a symmetric signal,

$$x_l \equiv \frac{1}{\mu_1^H}\left(\frac{\lambda_1^{2l-3} - 1}{\lambda_1^{l-1}}\right), \quad y_l \equiv \left(\frac{\lambda_1^{2l-2} - 1}{\lambda_1^{l-1}}\right),$$
$$v_l \equiv \frac{1}{\mu_1^H}\left(\frac{\lambda_1^{2M-2l-1} - 1}{\lambda_1^{M-l}}\right), \quad u_l \equiv \left(\frac{\lambda_1^{2M-2l} - 1}{\lambda_1^{M-l}}\right).$$

Thus,

$$\sum_{l=i_0}^{M-1} v_l = \frac{1}{\mu_1^H} \sum_{l=i_0}^{M-1} \left(\lambda_1^{M-l-1} - \lambda_1^{l-M}\right) = \frac{1}{\mu_1^H} \frac{(\lambda_1^{M-i_0} - 1)^2}{\lambda_1^{M-i_0}(\lambda_1 - 1)},$$
$$\sum_{l=i_0}^{M-1} u_l = \sum_{l=i_0}^{M-1} \frac{\lambda_1^{2M-2l} - 1}{\lambda_1^{M-l}} = \frac{(\lambda_1^{M+1-i_0} - 1)(\lambda_1^{M-i_0} - 1)}{\lambda_1^{M-i_0}(\lambda_1 - 1)},$$
$$x_{i_0} = \frac{1}{\mu_1^H}\left(\frac{\lambda_1^{2i_0-3} - 1}{\lambda_1^{i_0-1}}\right), \quad u_{i_0-1} = \frac{\lambda_1^{2M+2-2i_0} - 1}{\lambda_1^{M-i_0+1}}.$$

At $\sqrt{\beta} = \xi(S)\lambda^{i_0-(M+1)/2}$, which becomes $\sqrt{\beta} = \lambda_1^{2i_0-M}$ for symmetric signals, the second expression in (C.4) yields (using $\lambda_1 = \lambda_2$)

$$\frac{\alpha_M}{\sigma^S_{1,2}} \to \left(\frac{\lambda^{(M-1)/2} - \sqrt{\beta}}{\sqrt{\beta}\lambda^{(M-1)/2} - 1}\right) = \lambda_1^{2i_0-M}\left(\frac{\lambda_1^{2M-1-2i_0} - 1}{\lambda_1^{2i_0-1} - 1}\right).$$

Substituting into the desired inequality, we obtain the following sufficient inequality, after factoring:

$$\left(\frac{\lambda_1^{2M-1-2i_0} - 1}{(\lambda_1^{M+1-i_0} + 1)(\lambda_1^{M-i_0} - 1)}\right) \frac{\lambda_1^{2i_0-1} - \lambda_1^2}{\lambda_1^{2i_0-1} - 1} < 1.$$

The first LHS term is less than 1: it is sufficient to show that the numerator minus the denominator is negative, which yields the inequality

$$-\lambda_1^{2M-1-2i_0}\big(\lambda_1^2-1\big)+\lambda_1^{M-i_0}(\lambda_1-1)$$
$$<-\lambda_1^{M-i_0}(\lambda_1-1)\big(\lambda_1^{M-1-i_0}-1\big)<0.$$

The second term is less than 1 by $\lambda_1>1$. As desired, this establishes that when the DM is close to indifferent between starting in i_0 and i_0+1, we obtain $i_2\geq i_0$. *Q.E.D.*

STEP C.8: *For a symmetric signal, $\sqrt{\beta}>\lambda$ implies that the set $I\equiv\{i_1,\dots,i_2\}$ is nonempty, and that all states in this set are sticky down.*

PROOF: I prove the result for M odd and $i^*\geq 3$; a similar calculation yields the desired result if $i^*=2$ or if M is even. For a symmetric signal, Step C.2 implies that all interior states are fluid up, Step C.3 implies that all interior states outside I are fluid down, and Step C.4 implies that if any state in I is sticky down, then all states in I are sticky down. Therefore, to prove the result, it suffices to show that there is no optimal protocol in which all interior states are fluid both up and down.

Suppose, by contradiction, that there is an optimal protocol for a symmetric signal and $\sqrt{\beta}>\lambda$ in which all interior states are fluid, both up and down. For $\sqrt{\beta}$ near λ, equation (20) in the paper yields $i^*=\frac{M-1}{2}$ and $i_0=\frac{M+3}{2}$. Using (C.2) and our assumption that all interior states are fluid, optimality of $i^*\downarrow_1 i^*-1$ demands the first inequality below, and the second inequality is for optimality of $i_0\uparrow_S i_0+1$:

$$\frac{(\sigma_{1,2}^S)^2}{\eta}\leq\sum_{l=2}^{(M-1)/2}x_l\Big/\sum_{l=2}^{(M-3)/2}y_l\quad\text{and}$$

$$\frac{(\sigma_{M,M-1}^1)^2}{\eta}\geq\sum_{l=(M+3)/2}^{M-1}v_l\Big/\sum_{l=(M+3)/2}^{M-1}u_l=\sum_{l=2}^{(M-1)/2}x_l\Big/\sum_{l=2}^{(M-1)/2}y_l,$$

where the final equality above follows from the fact that for a symmetric signal, $x_l=v_{M+1-l}$ and $y_l=u_{M+1-l}$. Combining inequalities, and using the fact that $\alpha_M=\sigma_{M,M-1}^1$ when there is no interior mixing, we deduce (using (C.5) for the final inequality)

$$\left(\frac{\alpha_M}{\sigma_{1,2}^S}\right)^2\geq\frac{\sum_{l=2}^{(M-1)/2}x_l\Big/\sum_{l=2}^{(M-1)/2}y_l}{\sum_{l=2}^{(M-1)/2}x_l\Big/\sum_{l=2}^{(M-3)/2}y_l}=\sum_{l=2}^{(M-3)/2}y_l\Big/\sum_{l=2}^{(M-1)/2}y_l=\frac{\lambda_1^{(M-3)/2}-\lambda_1}{\lambda_1^{(M-1)/2}-1}.\tag{C.23}$$

On the other hand, the first expression in (C.4) yields the first inequality below for $\sqrt{\beta} > \lambda$ and $\lambda_1 = \lambda_2$, and the inequality is by $\lambda = \lambda_1^2$ (for a symmetric signal):

$$\text{(C.24)}\quad \left(\frac{\alpha_M}{\sigma^S_{1,2}}\right)^2 \leq \left(\frac{\lambda^{(M-3)/2}-\lambda}{\lambda^{(M+1)/2}-1}\right)^2 = \left(\frac{\lambda_1^{(M-3)/2}-\lambda_1}{\lambda^{(M+1)/2}-1}\right)^2\left(\frac{\lambda_1^{(M-3)/2}+\lambda_1}{\lambda^{(M+1)/2}+1}\right)^2$$
$$< \left(\frac{\lambda_1^{(M-3)/2}-\lambda_1}{\lambda^{(M-1)/2}-1}\right)\left(\frac{\lambda_1^{(M-3)/2}-\lambda_1}{\lambda^{(M+1)/2}-1}\right)$$
$$\times \left(\frac{\lambda_1^{(M-3)/2}+\lambda_1}{\lambda^{(M+1)/2}+1}\right)^2 \quad (\text{by } \lambda > 1).$$

But since $\lambda_1 > 1$, the product of the second two terms in (C.24) is below 1, while the first term is smaller than the RHS expression in (C.23). Therefore, the RHS of (C.24) is smaller than the RHS of (C.23), and so the two expressions are incompatible—a contradiction. *Q.E.D.*

APPENDIX D: INFORMATION MATTERS

CLAIM: *Given the normalization* $\beta \geq 1$, *information matters if and only if the following condition holds*:

$$\text{(D.1)}\quad \beta < \min_{s^*\in\{2,\dots,S\}} \left(\frac{\mu_1^L}{\mu_1^H}\frac{\eta+(1-\eta)\sum_{s=s^*}^{S}\mu_s^H}{\eta+(1-\eta)\sum_{s=s^*}^{S}\mu_s^L}\right)^{M-1} \equiv \left(\bar{\xi}(\eta)/\xi(1)\right)^{M-1}.$$

PROOF: Let $s^* \in \arg\min_{s\in\{2,\dots,S\}}(\eta + (1-\eta)\sum_{s=s^*}^S \mu_s^L)/(\eta + (1-\eta) \times \sum_{s=s^*}^S \mu_s^H)$.

FIRST DIRECTION: $\beta < (\bar{\xi}(\eta)/\xi(1))^{M-1} \Rightarrow$ *information matters.*

PROOF: Suppose (by way of contradiction) that (D.1) holds, but information does not matter. Then any memory protocol that chooses action 1 in all memory states is optimal. In particular, for all $\varepsilon \in (0,1)$, the following protocol is optimal: start in state M, and in each memory state $i \in \mathcal{M}$, choose action 1, transition to memory state M after any signal realization $s \geq s^*$, transition to memory state $i-1$ with probability ε after signal realization $s = 1$, and otherwise stay in state i. An immediate application of footnote 10 (noting that $\omega^\theta_{i,j} = 0$ whenever $j \notin \{i-1, i, M\}$) then yields the following expression for the terminal probabilities f_j^θ:

$$f_j^\theta = \frac{y_j^\theta}{\sum_{i=1}^{M} y_i^\theta}, \quad \text{where}$$

$$y_j^\theta = \Big(\prod_{i\le j-1} \omega_{i,M}^\theta\Big)\Big(\prod_{i\ge j+1} \omega_{i,i-1}^\theta\Big)$$
$$= \left(\eta + (1-\eta)\sum_{s=s^*}^{S} \mu_s^\theta\right)^{j-1} \big((1-\eta)\varepsilon\mu_1^\theta\big)^{M-j}.$$

Taking ratios in memory state $j = 1$ and then considering the limit as $\varepsilon \to 0$ yields

$$\text{(D.2)} \qquad \lim_{\varepsilon\to 0}\frac{f_1^H}{f_1^L} = \left(\frac{\mu_1^H}{\mu_1^L}\right)^{M-1} \lim_{\varepsilon\to 0}\frac{\sum_{i=1}^{M} y_i^L}{\sum_{i=1}^{M} y_i^H} = \left(\frac{\mu_1^H}{\mu_1^L}\frac{\eta + (1-\eta)\sum_{s=s^*}^{S}\mu_s^L}{\eta + (1-\eta)\sum_{s=s^*}^{S}\mu_s^H}\right)^{M-1}.$$

Then by continuity and (D.1), we can choose ε small enough that under the suggested protocol, the following expression holds:

$$\ell(1)\left|\frac{\pi^H}{\pi^L}\right| = \ell_0\left|\frac{\pi^H}{\pi^L}\right|\frac{f_1^H}{f_1^L} < 1 \quad \Rightarrow \quad p(1)\pi^H + \big(1 - p(1)\big)\pi^L < 0.$$

But then action 0 is strictly better than the specified action choice 1 in memory state 1, contradicting optimality of the specified protocol. This establishes the first direction: under (D.1), information matters. *Q.E.D.*

SECOND DIRECTION: *If* (D.1) *is violated*, *then information does not matter*.

PROOF: For this, I will prove that any optimal protocol yields the bound

$$\text{(D.3)} \qquad \frac{f_1^H}{f_1^L} \ge \left(\frac{\mu_1^H}{\mu_1^L}\frac{\eta + (1-\eta)\sum_{s=s^*}^{S}\mu_s^L}{\eta + (1-\eta)\sum_{s=s^*}^{S}\mu_s^H}\right)^{M-1}.$$

Then if the inequality in (D.1) is violated, it follows that $\ell_0|\pi^H/\pi^L|(f_1^H/f_1^L) \ge 1$, so that action 1 is preferred in memory state 1. Together with the ordering of the memory states and Proposition 1, it then follows that any optimal protocol must always choose action 1, hence information does not matter.

To prove (D.3), let (g^0, σ, d) be an optimal protocol with initial state i_0. By Appendix A.2, the terminal distribution f^θ is the steady-state distribution of a perturbed Markov process with transition chance $\omega^\theta_{i,j}$ from state i to j, where $\omega^\theta_{i,j} = \tau^\theta_{i,j}$ if $j \neq i_0$, and $\omega^\theta_{i,i_0} = \eta + \tau^\theta_{i,i_0}$. Then, for any subset of states $A \subseteq \mathcal{M}$, the probability of entering A equals the probability of leaving A. For memory state $i \leq i_0$ and the set $A = \{1, 2, \dots, i-1\}$, this yields

$$\sum_{j \leq i-1} f^\theta_j \left(\eta + \sum_{k \geq i} \tau^\theta_{j,k} \right) = \sum_{j \geq i} f^\theta_j \sum_{k \leq i-1} \tau^\theta_{j,k}.$$

Taking ratios and recalling that memory states are ordered so that f^H_j / f^L_j is weakly increasing, it follows that

$$\text{(D.4)} \qquad \frac{f^H_{i-1}}{f^L_{i-1}} \frac{\left(\eta + \sum_{k \geq i} \tau^H_{j,k} \right)}{\left(\eta + \sum_{k \geq i} \tau^L_{j,k} \right)} \geq \frac{\sum_{j \leq i-1} f^H_j \left(\eta + \sum_{k \geq i} \tau^H_{j,k} \right)}{\sum_{j \leq i-1} f^L_j \left(\eta + \sum_{k \geq i} \tau^L_{j,k} \right)}$$

$$= \frac{\sum_{j \geq i} f^H_j \sum_{k \leq i-1} \tau^H_{j,k}}{\sum_{j \geq i} f^L_j \sum_{k \leq i-1} \tau^L_{j,k}} \geq \frac{f^H_i}{f^L_i} \frac{\mu^H_1}{\mu^L_1}$$

$$\Rightarrow \quad \frac{f^H_{i-1}}{f^L_{i-1}} \geq \frac{f^H_i}{f^L_i} \left(\frac{\mu^H_1}{\mu^L_1} \frac{\eta + (1-\eta) \sum_{s=s^*}^{S} \mu^L_s}{\eta + (1-\eta) \sum_{s=s^*}^{S} \mu^H_s} \right).$$

If $i_0 = M$ (or if $i_0 \leq M - 1$ but all states above i_0 are unused), then (D.3) follows immediately by iterating (D.4), noting that $f^H_M / f^L_M \geq 1$ by the ordering of the memory states.

If $i_0 \leq M - 1$ and if some state above i_0 is used, then (i) optimality of starting in state i_0 demands $\ell_0 \leq \bar{\ell}_{i_0}$,[6] while (ii) any state $j > i_0$ with $f^\theta_j > 0$ has a belief

[6]This simply rearranges the condition stating that the DM must earn a weakly higher continuation payoff in state i_0 than $i_0 + 1$ at the prior p_0. Recall, from the discussion leading up to Proposition 2 in the paper that $\bar{\ell}_i \equiv \Delta^L_{i,i+1} / \Delta^H_{i+1,i}$ is the belief likelihood threshold for indifference between states i and $i+1$, with belief likelihoods below $\bar{\ell}_i$ favoring state i.

likelihood ratio $\ell(j) \equiv \ell_0 f_j^H/f_j^L \geq \bar{\ell}_{i_0}$ (by Corollary 2 in the paper). Together, these inequalities imply

$$\text{(D.5)} \qquad f_j^H/f_j^L \geq 1 \quad \text{for all } j \geq i_0+1 \text{ with } f_j^\theta > 0.$$

Also, f^θ satisfies the following steady-state relationship, stating that the probability of entering $\{1, 2, \ldots, i_0\}$ equals the probability of leaving:

$$\sum_{j\leq i_0} f_j^\theta \sum_{k\geq i_0+1} \tau_{j,k}^\theta = \sum_{j\geq i_0+1} f_j^\theta \left(\eta + \sum_{k\leq i_0} \tau_{j,i}^\theta\right).$$

Adding $\eta \sum_{j\leq i_0} f_j^\theta$ to both sides and taking ratios gives

$$\text{(D.6)} \qquad \frac{\sum_{j\leq i_0} f_j^H\left(\eta + \sum_{k\geq i_0+1} \tau_{j,k}^H\right)}{\sum_{j\leq i_0} f_j^L\left(\eta + \sum_{k\geq i_0+1} \tau_{j,k}^L\right)} = \frac{\eta + \sum_{j\geq i_0+1} f_j^H\left(\sum_{k\leq i_0} \tau_{j,i}^H\right)}{\eta + \sum_{j\geq i_0+1} f_j^L\left(\sum_{k\leq i_0} \tau_{j,i}^L\right)}$$

$$\text{(D.7)} \qquad \Rightarrow \quad \frac{f_{i_0}^H}{f_{i_0}^L} \frac{\eta + (1-\eta)\sum_{s=s^*}^{S} \mu_s^H}{\eta + (1-\eta)\sum_{s=s^*}^{S} \mu_s^L} \geq \frac{\mu_1^H}{\mu_1^L}.$$

(The LHS of (D.7) is an upper bound on the LHS of (D.6) by the ordering of the memory states; the RHS of (D.7) is a lower bound on the LHS of (D.6), using $\tau_{j,i}^H/\tau_{j,i}^L \geq \mu_1^H/\mu_1^L$, (D.5), and $1 > \mu_1^H/\mu_1^L$.) Together with (D.4) (iterated), it then follows that

$$\frac{f_1^H}{f_1^L} \geq \frac{f_{i_0}^H}{f_{i_0}^L} \left(\frac{\mu_1^H}{\mu_1^L} \frac{\eta + (1-\eta)\sum_{s=s^*}^{S} \mu_s^L}{\eta + (1-\eta)\sum_{s=s^*}^{S} \mu_s^H} \right)^{i_0-1}$$

$$\geq \left(\frac{\mu_1^H}{\mu_1^L} \frac{\eta + (1-\eta)\sum_{s=s^*}^{S} \mu_s^L}{\eta + (1-\eta)\sum_{s=s^*}^{S} \mu_s^H} \right)^{i_0},$$

which establishes (D.3) by $i_0 \leq M-1$. *Q.E.D.*

Q.E.D.

APPENDIX E: AN ANALYTICAL EXAMPLE

I now illustrate the theory with a symmetric three-state example. Assume the high and low states are equally likely, so $p_0 = \frac{1}{2}$, and that payoffs are $\pi^H = -\pi^L = 1$. The DM observes an "extreme" signal l or h with probability $(1-\phi)$, with likelihood ratios $\xi(h) = 1/\xi(l) = q/(1-q)$, and an "intermediate" signal $\bar{m}$ or $\underline{m}$ with probability ϕ, with $\xi(\bar{m}) = 1/\xi(\underline{m}) = p/(1-p)$; assume $q > p > \frac{1}{2}$. I restrict attention to protocols with beliefs $\ell(1) < \ell(2) < \ell(3)$ and with no absorbing or equivalent states. Assume that the DM starts in memory state 2 and uses the decision rule $d = (0, d_2, 1)$, with d_2 the *probability* of choosing the high action in memory state 2. The optimal transition rule then maximizes the objective function

$$\frac{1}{2}\big(d_2 f_2^H + f_3^H\big) - \frac{1}{2}\big(d_2 f_2^L + f_3^L\big). \tag{E.1}$$

In Step E.1, I will illustrate why intermediate signals are optimally ignored for η near zero. In Steps E.2–E.4, specializing to a binary signal ($\phi = 0$), I will demonstrate some features of optimal protocols that easily hold for general termination chances $\eta \in (0,1)$, not necessarily small; the remainder of the section finishes computing the optimal transition rules for the action rules $d = (0, \frac{1}{2}, 1)$ and $d = (0,1,1)$, which are referenced in Step E.4.

STEP E.1: *Only react to extreme signals for* η *near* 0.

To see this most easily, consider a symmetric decision rule $d = (0, \frac{1}{2}, 1)$, which intuitively implies a symmetric optimal transition rule. Then, with $f_2^H = f_2^L$ and $f_3^L = f_1^H$ (by symmetry) and the identity $\sum_{i=1}^3 f_i^H = 1$, we can rewrite (E.1) as $\frac{1}{2}(f_3^H - f_1^H) = \frac{1}{2}f_2^H + f_3^H - \frac{1}{2}$. Thus, defining $x(\sigma) \equiv (f_1^H + \frac{1}{2}f_2^H)/(f_3^H + \frac{1}{2}f_2^H)$, the goal is to maximize $\frac{1}{2}f_2^H + f_3^H = (1 + x(\sigma))^{-1}$, which clearly is achieved by making $x(\sigma)$ as small as possible.

By symmetry and the ordering of the states,

$$x(\sigma) \equiv \frac{f_1^H + \dfrac{1}{2}f_2^H}{f_1^L + \dfrac{1}{2}f_2^L} \geq \frac{f_1^H}{f_1^L} \quad \text{with equality iff } f_2^H = f_2^L = 0. \tag{E.2}$$

And by equation (2) in the paper, f^θ is the steady-state distribution of a Markov process with transition chances $\omega_{i,j}^\theta$, and so the probability of entering state 1 equals the probability of leaving:

$$f_1^\theta\big(\omega_{1,2}^\theta + \omega_{1,3}^\theta\big) = f_2^\theta \omega_{2,1}^\theta + f_3^\theta \omega_{3,1}^\theta. \tag{E.3}$$

Taking ratios, and recalling $\omega^H_{i,j}/\omega^L_{i,j} \geq \xi(l)$ (with equality if $\sigma^s_{i,j} = 0\ \forall s \neq l$) and $f^H_3/f^L_3 > f^H_2/f^L_2 = 1$ yields

$$\text{(E.4)}\qquad \frac{f^H_1}{f^L_1} \geq \xi(l)\frac{\omega^L_{1,2} + \omega^L_{1,3}}{\omega^H_{1,2} + \omega^H_{1,4}} \quad \text{with equality iff } \sigma^m_{2,1} = 0 \text{ and } \omega^\theta_{3,1} = 0$$

$$\text{(E.5)}\qquad \equiv \xi(l)\frac{\eta + (1-\eta)\mu^L_{\bar{m}}(\sigma^{\bar{m}}_{1,2} + \sigma^{\bar{m}}_{1,3}) + (1-\eta)\mu^L_h(\sigma^h_{1,2} + \sigma^h_{1,3})}{\eta + (1-\eta)\mu^H_{\bar{m}}(\sigma^{\bar{m}}_{1,2} + \sigma^{\bar{m}}_{1,3}) + (1-\eta)\mu^H_h(\sigma^h_{1,2} + \sigma^h_{1,3})}$$

$$\text{(E.6)}\qquad \geq \xi(l)/\xi(h) \quad \text{with equality iff } \eta = \sigma^{\bar{m}}_{1,2} + \sigma^{\bar{m}}_{1,3} = 0.$$

Thus, the ratio $x(\sigma)$ that the DM wishes to minimize has a lower bound $\xi(l)/\xi(h)$, achieved as $\eta \to 0$ only if intermediate signals are ignored (by (E.4), (E.6), and symmetry), there are no jumps (by (E.4) and symmetry), and $f^H_2 = f^L_2 = 0$ (achieved as $\eta \to 0$ by leaving the extremal states with a vanishing chance).[7]

Now specialize to a binary signal ($\phi = 0$), and recall from Proposition 2 in the paper that an equilibrium is characterized by thresholds $\bar{\ell}_1 \leq \bar{\ell}_2$ such that a self transitions to 1 if his posterior is below $\bar{\ell}_1$, to state 3 if his posterior is above $\bar{\ell}_2$, and to state 2 for posteriors in between.

STEP E.2—Indifference Thresholds: *$d_2 = 1 \Rightarrow \bar{\ell}_1 < \bar{\ell}_2$ and $\xi(l)/\bar{\ell}_1 < \xi(h)/\bar{\ell}_2$.*

PROOF: By equation (5) in the paper, the payoff differential $\Delta^\theta_{3,2}$ satisfies

$$\text{(E.7)}\qquad \big(\eta + (1-\eta)\mu^\theta_l\big(\sigma^l_{3,1} + \sigma^l_{3,2}\big) + (1-\eta)\mu^\theta_h\sigma^h_{2,3}\big)\Delta^\theta_{3,2} = (1-\eta)\mu^\theta_l\big(\sigma^l_{2,1} - \sigma^l_{3,1}\big)\Delta^\theta_{2,1}.$$

With no equivalent states, Lemma 1(a) demands $\Delta^H_{3,2} > 0$ and $\Delta^H_{2,1} > 0$, requiring $\sigma^l_{2,1} > \sigma^l_{3,1}$ by (E.7). Then taking ratios in (E.7) and recalling $\bar{\ell}_2 \equiv \Delta^L_{2,3}/\Delta^H_{3,2}$ and $\bar{\ell}_1 \equiv \Delta^L_{1,2}/\Delta^H_{2,1}$ yields

$$\text{(E.8)}\qquad \frac{\eta + (1-\eta)(1-q)(\sigma^l_{3,1} + \sigma^l_{3,2}) + (1-\eta)q\sigma^h_{2,3}}{\eta + (1-\eta)q(\sigma^l_{3,1} + \sigma^l_{3,2}) + (1-\eta)(1-q)\sigma^h_{2,3}}\frac{1}{\bar{\ell}_2} = \frac{1-q}{q}\frac{1}{\bar{\ell}_1}.$$

The RHS of (E.7) is $\xi(l)/\bar{\ell}_1$, and the LHS is strictly above $\xi(l)/\bar{\ell}_2$ and strictly below $\xi(h)/\bar{\ell}_2$; this immediately yields the desired result. *Q.E.D.*

[7]Of course, ignoring intermediate signals is not optimal for larger η. This is easily seen from (E.5), which is minimized by setting $\sigma^{\bar{m}}_{1,2} + \sigma^{\bar{m}}_{1,3}$ equal to 1 (instead of zero) whenever $\frac{(\eta+(1-\eta)\mu^H_h)}{(\eta+(1-\eta)\mu^L_h)} < \xi(\bar{m})$. This relates to the condition in footnote 11 in the paper for information to matter, and holds for smaller values of η the closer are $\xi(\bar{m})$ and $\xi(h)$.

Step E.2 has the following immediate implications: (i) State 2 cannot be sticky in both directions, for if a state-2 self is indifferent about moving down, $\ell(2)\xi(l)/\bar{\ell}_1 = 1$, then he strictly prefers moving up after an h-signal, as $\ell(2)\xi(h)/\bar{\ell}_2 > 1$; thus, $\sigma^l_{2,1} \in (0, 1) \Rightarrow \sigma^h_{2,3} = 1$. And (ii) if a state-1 self ever jumps to state 3 after an h-signal, so $\ell(1, h)/\bar{\ell}_2 \geq 1$, then he never stays in state 1, as $\ell(1, h)/\bar{\ell}_1 > 1$ (indicating a strict preference for state 2 over state 1); thus, $\sigma^h_{1,3} > 0 \Rightarrow \sigma^h_{1,2} + \sigma^h_{1,3} = 1$.

STEP E.3: *No jumps if* $d_2 \in \{0, 1\}$.

PROOF: In an equilibrium with $\sigma^h_{1,3} > 0$, a state-1 self who observes an h-signal must weakly prefer memory state 3 to 2, while a state-3 self who observes an l-signal must weakly prefer state 2 to 3 (since we have ruled out absorbing states). Thus, the following inequality must hold:

$$\frac{f_1^H}{f_1^L}\frac{q}{1-q} \equiv \ell(1, h) \geq \bar{\ell}_2 \geq \ell(3, l) \equiv \frac{f_3^H}{f_3^L}\frac{1-q}{q}. \tag{E.9}$$

I will now derive a contradiction. By Step E.2 implication (ii), $\sigma^h_{1,3} > 0 \Rightarrow \sigma^h_{1,2} + \sigma^h_{1,3} = 1$. Then by (E.3), adding $f_1^\theta(1-\eta)\mu_l^\theta$ to both sides and taking ratios,

$$\frac{f_1^H}{f_1^L} = \frac{1-q}{q}\frac{f_1^H + f_2^H\sigma^l_{2,1} + f_3^H\sigma^l_{3,1}}{f_1^L + f_2^L\sigma^l_{2,1} + f_3^L\sigma^l_{3,1}}. \tag{E.10}$$

Now note first that $\sigma^l_{3,1} \leq \sigma^l_{2,1}$ at equilibrium, for if $\sigma^l_{3,1} > 0$ is optimal, so $\ell(3, l)/\bar{\ell}_1 \leq 1$; then $\ell(2, l)/\bar{\ell}_1$ is strictly below 1 and so $\sigma^l_{2,1}$ must equal 1. But then by the ordering of the states and the identity $\sum_{i=1}^3 f_i^\theta = 1$, the RHS expression in (E.10) is strictly below 1 whenever $\sigma^l_{3,1} < 1$, which was established just below (E.7) for any equilibrium with $d_2 = 1$. Thus, $d_2 = 1 \Rightarrow \ell(1, h) < 1$. By a symmetric argument, $\ell(3, l) \geq 1$, with strict inequality in any equilibrium with $d_2 = 0$. Thus $d_2 \in \{0, 1\} \Rightarrow \ell(1, h) < \ell(3, l)$, contradicting (E.9).[8] *Q.E.D.*

STEP E.4: *The optimal memory protocol is asymmetric*, *with* $d_2 \in \{0, 1\}$.

[8]If $d_2 = \frac{1}{2}$, then there *is* a symmetric team equilibrium with $\sigma^h_{1,3} = \sigma^l_{3,1} = 1$, but it is Pareto-dominated by the symmetric equilibrium with no jumps described in Step E.4. This is easily seen by noting that if the strategies were perturbed, say to $\sigma^h_{1,2} = \varepsilon$ and $\sigma^h_{1,3} = 1 - \varepsilon$, we would have $\ell(1, h) < \ell(3, l)$, indicating that the jumps between states 1 and 3 are *strictly* suboptimal ((E.9) is violated). But then for any $\varepsilon > 0$, it is a profitable deviation to increase ε further, and decrease the chance $(1 - \varepsilon)$ of jumps. This equilibrium with $\sigma^h_{1,3} = \sigma^l_{3,1} = 1$ also fails Marple and Shoham's (2012) "distributed perfect equilibrium" concept, which essentially requires that strategies be robust to trembles.

PROOF—Sketch: In the next two subsections, I will finish characterizing the optimal protocol when $d_2 = 1$, establishing that $\sigma^l_{2,1} = \sigma^h_{2,3} = 1$ and $\sigma^h_{1,2} < \sigma^l_{3,2}$. I then find the best *symmetric* equilibrium, with action rule $\widehat{d} = (0, \frac{1}{2}, 1)$, obtaining the closed-form expressions for the best symmetric transition rule $\widehat{\sigma}$,[9]

$$\widehat{\sigma}^l_{2,1} = \widehat{\sigma}^h_{2,3} = 1, \quad \widehat{\sigma}^h_{1,3} = \widehat{\sigma}^l_{3,1} = 0, \quad \text{and}$$
$$\widehat{\sigma}^h_{1,2} = \widehat{\sigma}^l_{3,2} = \frac{-\eta + \sqrt{\eta(2-\eta)}}{1-\eta}.$$

To see that $(\widehat{\sigma}, \widehat{d})$ is Pareto-dominated for $\eta \in (0, 1)$, note that it necessarily leaves the DM indifferent between actions in the middle memory state, with $f^H_2 = f^L_2$. Thus, he would earn exactly the same payoff if he again followed $\widehat{\sigma}$, but switched to the action rule $d = (0, 1, 1)$. But we have just established that the *optimal* transition rule σ when $d_2 = 1$ is necessarily *asymmetric*, with $\sigma^h_{1,2} < \sigma^l_{3,2}$, and so (σ, d) must earn a strictly higher payoff than $(\widehat{\sigma}, \widehat{d})$. The intuition is that it is not optimal to waste memory resources on a completely uninformed state. If the DM *knows* that he will choose the high action in the middle state, then it is payoff-improving to reduce the chance of ending up there in the low state of the world, while increasing the relative chance of ending up there in the high state of the world. Since he is most often in state 1 (or 3) in the low (or high) state of the world, this is accomplished by reducing $\sigma^h_{1,2}$ and increasing $\sigma^l_{3,2}$.[10] Of course, the rule $(\widehat{\sigma}, \widehat{d})$ is asymptotically efficient: part (d) of Proposition 4 guarantees that the chance of making a decision in an interior memory state i vanishes as $\eta \to 0$, and so the limit payoff does not depend on d_i.

E.1. *Completing the Example if* $d_2 = 1$

It remains to prove that the optimal transition rule σ satisfies $\sigma^l_{2,1} = \sigma^h_{2,3} = 1$ and $\sigma^h_{1,2} < \sigma^l_{3,2}$.

STEP E.5: $\sigma^l_{2,1} \in (0, 1) \Rightarrow \sigma^h_{1,2} = 1$ *and* $\sigma^h_{2,3} \in (0, 1) \Rightarrow \sigma^l_{3,2} = 1$.

PROOF: I will prove the first assertion; the second follows by a symmetric argument. Consider an equilibrium with $\sigma^l_{2,1} \in (0, 1)$, if such an equilibrium

[9]Note that these transitions out of the extremal states are strictly increasing in η, but below $1\ \forall \eta < 1$; thus, stickiness optimally persists.

[10]The rule $(\widehat{\sigma}, d)$ also does not correspond to a team equilibrium: each state-1 self knows that the action will change for sure if he "passes the baton" to state 2 and so he is more reluctant to do so ($\sigma^h_{1,2}$ must fall), while each state-3 self more readily passes the baton down, knowing that the action will not change.

exists. Optimality of $\sigma^l_{2,1} \in (0, 1)$ demands that a state-2 self be indifferent between states 1 and 2 after observing an l-signal, so $\ell(2, l)/\bar{\ell}_1 = 1$. But then by Step E.2,

$$\ell(2, h)/\bar{\ell}_2 \equiv \ell(2)\xi(h)/\bar{\ell}_2 > \ell(2)\xi(l)/\bar{\ell}_1 \equiv \ell(2, l)/\bar{\ell}_1 = 1, \tag{E.11}$$

while taking ratios in (E.5), using the Step E.3 result that $\sigma^h_{1,3} = \sigma^l_{3,1} = 0$, yields

$$\begin{aligned}\ell(1, h) &\equiv \frac{f^H_1}{f^L_1}\frac{q}{1-q} > \frac{f^H_1}{f^L_1}\frac{\eta + (1-\eta)q\sigma^h_{1,2}}{\eta + (1-\eta)(1-q)\sigma^l_{1,2}} \\ &= \frac{f^H_2}{f^L_2}\frac{\mu^H_l}{\mu^L_l} \equiv \ell(2, l) = \bar{\ell}_1.\end{aligned} \tag{E.12}$$

But $\ell(2, h)/\bar{\ell}_2 > 1$ (from (E.11)) implies that a state-2 self strictly prefers state 3 to 2 after observing an h-signal, and so at equilibrium, he moves with probability $\sigma^h_{2,3} = 1$. Similarly, $\ell(1, h)/\bar{\ell}_1 > 1$ (from (E.12)) implies that a state-1 self strictly prefers state 2 to 1 after observing an h-signal, and so at equilibrium, he moves with probability $\sigma^h_{1,2} = 1$. *Q.E.D.*

STEP E.6: $\sigma^h_{2,3} = 1$.

PROOF: By (E.8) and (E.11), using the Step E.2 result that $\sigma^l_{3,1} = 0$,

$$\begin{aligned}\frac{\ell(1, h)}{\bar{\ell}_1}\Big/\frac{\ell(2, h)}{\bar{\ell}_2} &= \frac{\eta + (1-\eta)(1-q)\sigma^l_{3,2} + (1-\eta)q\sigma^h_{2,3}}{\eta + (1-\eta)q\sigma^l_{3,2} + (1-\eta)(1-q)\sigma^h_{2,3}} \\ &\quad\times \frac{\eta + (1-\eta)(1-q)\sigma^h_{1,2}}{\eta + (1-\eta)q\sigma^h_{1,2}}.\end{aligned} \tag{E.13}$$

If there is an optimal protocol with $\sigma^h_{2,3} \in (0, 1)$, then $\sigma^l_{3,2} = 1$ by Step E.5, implying that the RHS expression in (E.13) is strictly below 1. Also, in an equilibrium with $\sigma^h_{2,3} \in (0, 1)$, a state-2 self must be indifferent about moving up after an h-signal, so $\ell(2, h)/\bar{\ell}_2 = 1$. Thus $\ell(1, h)/\bar{\ell}_1 < \ell(2, h)/\bar{\ell}_2 = 1$, indicating that the state-1 self strictly prefers to remain in state 1 after an h-signal, which is a contradiction, since absorbing states are not optimal. *Q.E.D.*

STEP E.7: *Calculating* $\ell(3, l)/\bar{\ell}_2$ *and* $\ell(1, h)/\bar{\ell}_1$.

By footnote 10 in the paper, using the Step E.3 result that there are no transitions between states 1 and 3, the terminal distribution f^θ satisfies $f^\theta_i =$

$y_i^\theta / \sum_{j=1}^3 y_j^\theta$, where $y_1^\theta = \omega_{3,2}^\theta \omega_{2,1}^\theta$, $y_2^\theta = \omega_{1,2}^\theta \omega_{3,2}^\theta$, and $y_3^\theta = \omega_{1,2}^\theta \omega_{2,3}^\theta$.[11] Next, comparing (5) from the paper in states 1 and 2 yields the following expression for payoff gap $\Delta_{2,1}^\theta$, using $\sigma_{2,3}^h = 1$ (just shown) and $\sigma_{3,1}^l = \sigma_{1,3}^h = 0$:

$$\big(\eta + (1-\eta)\sigma_{2,1}^l \mu_l^\theta + (1-\eta)\sigma_{1,2}^h \mu_h^\theta\big)\Delta_{2,1}^\theta = \eta\pi^\theta + (1-\eta)\mu_h^\theta \Delta_{3,2}^\theta.$$

Substituting (E.7) into this expression and solving for $\Delta_{2,1}^\theta$ and $\Delta_{3,2}^\theta$ yields

$$\Delta_{2,1}^\theta = \eta\pi^\theta \frac{(\eta + (1-\eta)\sigma_{3,2}^l \mu_l^\theta + (1-\eta)\mu_h^\theta)}{\sum_{j=1}^3 y_j^\theta} \quad \text{and}$$

$$\Delta_{3,2}^\theta = \eta\pi^\theta \frac{(1-\eta)\sigma_{2,1}^l \mu_l^\theta}{\sum_{j=1}^3 y_j^\theta}.$$

Setting $\Gamma \equiv \sum_{j=1}^3 y_j^L / \sum_{j=1}^3 y_j^H$, we then obtain

$$\text{(E.14)} \quad \frac{\ell(1,h)}{\bar{\ell}_1} \equiv \frac{y_1^H}{y_1^L} \frac{q}{1-q} \frac{\Delta_{2,1}^H}{\Delta_{1,2}^L}\Gamma = \frac{\eta + (1-\eta)\sigma_{3,2}^l(1-q)}{\eta + (1-\eta)\sigma_{3,2}^l q} \times \frac{\eta + (1-\eta)\sigma_{3,2}^l(1-q) + (1-\eta)q}{\eta + (1-\eta)\sigma_{3,2}^l q + (1-\eta)(1-q)}\Gamma^2,$$

$$\text{(E.15)} \quad \frac{\ell(3,l)}{\bar{\ell}_2} \equiv \frac{y_3^H}{y_3^L} \frac{1-q}{q} \frac{\Delta_{3,2}^H}{\Delta_{2,3}^L}\Gamma = \frac{\eta + (1-\eta)\sigma_{1,2}^h q}{\eta + (1-\eta)\sigma_{1,2}^h(1-q)} \frac{1-q}{q}\Gamma^2,$$

where

$$\text{(E.16)} \quad \Gamma = \Big(\big(\eta + (1-\eta)\sigma_{1,2}^h(1-q) + (1-\eta)\sigma_{2,1}^l q\big) \times \big(\eta + (1-\eta)(1-q) + (1-\eta)\sigma_{3,2}^l q\big) - (1-\eta)^2 q(1-q)\sigma_{2,1}^l\Big) \Big/ \Big(\big(\eta + (1-\eta)\sigma_{1,2}^h q + (1-\eta)(1-q)\sigma_{2,1}^l\big) \times \big(\eta + (1-\eta)q + (1-\eta)\sigma_{3,2}^l(1-q)\big) - (1-\eta)^2 q(1-q)\sigma_{2,1}^l\Big).$$

[11]For example, note that there are in principle three 1-trees: $3 \to 2 \to 1$, $2 \to 3 \to 1$, and $2, 3 \to 1$. This yields $y_1^\theta = \omega_{3,2}^\theta \omega_{2,1}^\theta + \omega_{2,3}^\theta \omega_{3,1}^\theta + \omega_{2,1}^\theta \omega_{3,1}^\theta$, but the last two terms equal zero given $\omega_{3,1}^\theta = 0$.

STEP E.8: $\sigma^h_{1,2} < 1 = \sigma^l_{2,1}$.

PROOF: First, I show that $\sigma^l_{3,2} \in (0,1) \Rightarrow \sigma^h_{1,2} \in (0,1)$: The equilibrium condition for $\sigma^l_{3,2} \in (0,1)$ is $\ell(3,l)/\bar{\ell}_2 = 1$, which, by (E.15), can only hold if $\Gamma > 1$; but by (E.16), this requires $\sigma^h_{1,2} < 1$, as $\sigma^h_{1,2} = 1 \Rightarrow \Gamma \leq 1$ (equality iff $\sigma^l_{2,1} = \sigma^l_{3,2} = 1$).

Now, to prove that $\sigma^h_{1,2} < 1 = \sigma^l_{2,1}$, we have just shown that if $\sigma^h_{1,2} = 1$, then $\sigma^l_{3,2}$ must also equal 1. But in such an equilibrium, the RHS expression in (E.14) is strictly below Γ^2, while (E.16) yields $\Gamma \leq 1$ at $\sigma^h_{1,2} = \sigma^l_{3,2} = 1$ (with equality iff $\sigma^l_{2,1} = 1$). Thus, if $\sigma^h_{1,2} = 1$, then (E.14) yields $\ell(1,h)/\bar{\ell}_1 < 1$, a violation of the equilibrium condition in state 1: a state-1 self strictly prefers to stay in state 1 after an h-signal. We therefore conclude that there is no equilibrium with $\sigma^h_{1,2} = 1$; thus, $\sigma^h_{1,2} \in (0,1)$, which implies $\sigma^l_{2,1} = 1$ by Step E.5. *Q.E.D.*

STEP E.9: *Optimal transition chances* $\sigma^h_{1,2} < \sigma^l_{3,2}$.

It is now immediate that $\sigma^l_{3,2} > \sigma^h_{1,2}$ at equilibrium: For if $\sigma^l_{3,2} \leq \sigma^h_{1,2} < 1$ is optimal, then the RHS expression in (E.14) must equal 1; this is only possible if $\Gamma > 1$, which, by (E.16) evaluated at $\sigma^l_{2,1} = 1$, demands $\sigma^l_{3,2} > \sigma^h_{1,2}$—a contradiction.

It remains to choose $\sigma^h_{1,2} \in (0,1)$ and $\sigma^l_{3,2}$ such that the RHS expression in (E.14) equals 1, and the RHS expression in (E.15) is at most 1, with equality if $\sigma^l_{3,2} \in (0,1)$. For large η (specifically, η satisfying $\gamma(1-\gamma)/(1-\eta) < \eta\sqrt{\gamma+\eta}$), the optimal protocol has $\sigma^l_{3,2} = 1$ and $\sigma^h_{1,2} = (-\eta + (1-\gamma)\sqrt{\eta+\gamma})/(1-\eta-\gamma)$. For small η (whenever this inequality does not hold), the transition from state 3 to 2 is also mixed: a closed-form solution is not possible, but $\sigma^l_{3,2}$ may be found numerically by setting $\sigma^h_{1,2} = \frac{\eta/\gamma + \sigma^l_{3,2}(1-\sigma^l_{3,2})}{(\sigma^l_{3,2})^2/\eta - (1-\sigma^l_{3,2})^2}$ (this equates the RHS expressions in (E.14) and (E.15)) and then choosing $\sigma^l_{3,2} \in (0,1)$ to set the RHS of (E.15) equal to 1.

E.2. *Completing the Example if* $d_2 = \frac{1}{2}$

We have already shown that jumps are not optimal, and so it remains to calculate the optimal transition chances $\sigma^l_{3,2} = \sigma^h_{1,2}$ and $\sigma^l_{2,1} = \sigma^h_{2,3}$ to maximize

$$\text{(E.17)} \quad \frac{1}{2}\left(\frac{1}{2}f^H_2 + f^H_3\right) - \frac{1}{2}\left(\frac{1}{2}f^L_2 + f^L_1\right) =_{\text{symmetry}} \frac{1}{2}f^H_2 + f^H_3 - \frac{1}{2}.$$

With $d_2 = \frac{1}{2}$, the expression for $\ell(1,h)/\bar{\ell}_1$ from (E.14) changes to the expression (using symmetry to set $\sigma^l_{3,2} = \sigma^h_{1,2}$ and $\Gamma = 1$)

$$\text{(E.18)} \quad \frac{\ell(1,h)}{\bar{\ell}_1} = \left(\frac{\eta + (1-\eta)(1-q)\sigma^h_{1,2}}{\eta + (1-\eta)q\sigma^h_{1,2}}\right) \times \left(\frac{\eta + (1-\eta)(1-q)\sigma^h_{1,2} + 2(1-\eta)q}{\eta + (1-\eta)q\sigma^h_{1,2} + 2(1-\eta)(1-q)}\right).$$

The first ratio is $\ell(1,h)$ and the second is $1/\bar{\ell}_1$: Noting that $1/\bar{\ell}_1$ is strictly below $\xi(h)$, we immediately have $\ell(2,l)/\bar{\ell}_1 = \ell_0\xi(l)/\bar{\ell}_1 < 1$ (recalling that $\ell(2) = f^H_2/f^L_2 = 1$ in a symmetric equilibrium), and so $\sigma^l_{2,1}$ must equal 1; symmetrically, $\sigma^h_{2,3} = 1$.

To calculate the optimal transitions *into* the middle state, we need to choose $\sigma^h_{1,2}$ such that either $\ell(1,h)/\bar{\ell}_1 = 1$ and $\sigma^h_{1,2} \in (0,1)$, or $\ell(1,h)/\bar{\ell}_1 > 1$ and $\sigma^h_{1,2} = 1$. In (E.18), the numerator exceeds the denominator by $(1-\eta) \times (2q-1)(2\eta(1-\sigma^h_{1,2}) - (1-\eta)(\sigma^h_{1,2})^2)$. This difference is strictly negative at $\sigma^h_{1,2} = 1$, and so $\ell(1,h) < \bar{\ell}_1$, while the difference is strictly positive at $\sigma^h_{1,2} = 0$, and so $\ell(1,h) > \bar{\ell}_1$. Both contradict optimality. Any equilibrium therefore requires mixing. In fact, by choosing $\sigma^h_{1,2} = (-\eta + \sqrt{\eta(2-\eta)})/(1-\eta)$, we get $\ell(1,h) = \bar{\ell}_1$. The condition for the transition from 3 to 2 is symmetric.

Finally, let us consider the equilibrium payoff. Substituting the optimal transition chances into (E.17) and simplifying with $\sqrt{\rho^*} \equiv (1-q)^2/q^2$ yields the payoff

$$\frac{1}{2}\frac{(1-\eta)(1-\sqrt{\rho^*})}{1+\sqrt{\rho^*}+2\sqrt{\eta(2-\eta)}(1-q)/q}.$$

The limit as $\eta \to 0$ is $\frac{1}{2}(1-\sqrt{\rho^*})/(1+\sqrt{\rho^*})$, that is, the bound at $\beta = 1$ in Proposition 3 of the paper.

Dept. of Economics, University of Wisconsin–Madison, 1180 Observatory Dr., Madison, WI 53706, U.S.A.; awilson@ssc.wisc.edu.

Manuscript received May, 2004; final revision received June, 2014.