Is Your Machine Better Than You? You May Never Know.

Francis de Véricourt, Huseyin Gurkan
ESMT Berlin, {francis.devericourt, huseyin.gurkan}@esmt.org

Appendix A: No-Interaction and No-Overriding Benchmarks

We prove Theorems 1-2 by first deriving a recursive expression (in terms of b_{t-1}) for belief b_t using Bayes' rule (see, e.g., Winkler 1972). Then, we focus on the log-likelihood ratio process L_t defined by $L_t = \log(b_t/(1-b_t))$. Observe that when $L_t \to \infty$ (and/or $L_t \to -\infty$) almost surely, then it immediately follows that $b_t \to 1$ (and resp., $b_t \to 0$) due to the continuous mapping theorem (Resnick 2014, p. 261) L_t is a continuous monotone transformation of b_t.

Proof of Theorem 1. In the no-interaction benchmark, the DM acts only if $S_t^{\mathsf{H}} = +$, so it follows that

$$b_t = \left[1 + \frac{\bar{b}_{t-1}}{b_{t-1}}\zeta\right]^{-1} \tag{1}$$

where

$$\zeta \triangleq \left\{ \left[\left(\frac{\bar{\alpha}^{\mathsf{W}}}{\bar{\alpha}^{\mathsf{B}}}\right)^{1_{\{S_t^{\mathsf{M}}=-\}}} \left(\frac{\alpha^{\mathsf{W}}}{\alpha^{\mathsf{B}}}\right)^{1_{\{S_t^{\mathsf{M}}=+\}}} \right]^{1_{\{\Theta_t=\mathsf{A}\}}} \left[\left(\frac{\beta^{\mathsf{W}}}{\beta^{\mathsf{B}}}\right)^{1_{\{S_t^{\mathsf{M}}=-\}}} \left(\frac{\bar{\beta}^{\mathsf{W}}}{\bar{\beta}^{\mathsf{B}}}\right)^{1_{\{S_t^{\mathsf{M}}=+\}}} \right]^{1_{\{\Theta_t=\mathsf{NA}\}}} \right\}^{1_{\{S_t^{\mathsf{H}}=+\}}} \tag{2}$$

Here, $1_{\{\cdot\}}$ is the indicator variable. Using the definition of L_t, we obtain $L_t = L_{t-1} + R_t$, where

$$\begin{aligned} R_t \triangleq & 1_{\{S_t^{\mathsf{H}}=+, S_t^{\mathsf{M}}=+, \Theta_t=\mathsf{A}\}} \log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right) + 1_{\{S_t^{\mathsf{H}}=+, S_t^{\mathsf{M}}=-, \Theta_t=\mathsf{A}\}} \log\left(\frac{\bar{\alpha}^{\mathsf{B}}}{\bar{\alpha}^{\mathsf{W}}}\right) \\ & + 1_{\{S_t^{\mathsf{H}}=+, S_t^{\mathsf{M}}=+, \Theta_t=\mathsf{NA}\}} \log\left(\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right) + 1_{\{S_t^{\mathsf{H}}=+, S_t^{\mathsf{M}}=-, \Theta_t=\mathsf{NA}\}} \log\left(\frac{\beta^{\mathsf{B}}}{\beta^{\mathsf{W}}}\right). \end{aligned} \tag{3}$$

Therefore, L_t is a random walk with i.i.d. random jumps R_t. When the machine's type is $\Gamma \in \{\mathsf{B}, \mathsf{W}\}$, the mean of the random jump is

$$\mathbb{E}^{\Gamma}[R_t] = p\alpha^{\mathsf{H}} \left[\alpha^{\Gamma} \log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right) + \bar{\alpha}^{\Gamma} \log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right)\right] + \bar{p}\bar{\beta}^{\mathsf{H}} \left[\bar{\beta}^{\Gamma} \log\left(\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right) + \beta^{\Gamma} \log\left(\frac{\beta^{\mathsf{B}}}{\beta^{\mathsf{W}}}\right)\right]. \tag{4}$$

The mean $\mathbb{E}^{\Gamma}[R_t]$ is positive (negative) when $\Gamma = \mathsf{B}$ (and resp., $\Gamma = \mathsf{W}$). This is because, the terms inside the square brackets are the Kullback-Leibler (KL) divergence (Csiszar 1975) when $\Gamma = \mathsf{B}$, i.e., $D_{KL}(\mathbb{P}^{\mathsf{B}}(\cdot \,|\, \Theta = \mathsf{A}) \,||\, \mathbb{P}^{\mathsf{W}}(\cdot \,|\, \Theta = \mathsf{A}))$ inside the first square brackets and $D_{KL}(\mathbb{P}^{\mathsf{B}}(\cdot \,|\, \Theta =$

$\mathsf{NA}) \,||\, \mathbb{P}^{\mathsf{W}}(\cdot \,|\, \Theta = \mathsf{NA}))$ inside the first square brackets. Similarly, $-D_{KL}(\mathbb{P}^{\mathsf{W}}(\cdot \,|\, \Theta = \mathsf{A}) \,||\, \mathbb{P}^{\mathsf{B}}(\cdot \,|\, \Theta = \mathsf{A}))$ is inside the first square brackets and $-D_{KL}(\mathbb{P}^{\mathsf{W}}(\cdot \,|\, \Theta = \mathsf{NA}) \,||\, \mathbb{P}^{\mathsf{B}}(\cdot \,|\, \Theta = \mathsf{NA}))$ is inside the second square brackets when $\Gamma = \mathsf{W}$. Gibbs' inequality (see Section 2.6 in MacKay 2003) implies that KL divergence is always positive when the distributions are not the same, which implies the result. Q.E.D.

Proof of Theorem 2. In the no-overriding benchmark, the DM acts only if $S_t^{\mathsf{M}} = +$, so it follows that

$$b_t = \frac{b_{t-1}\left(\alpha^{\mathsf{B}}\right)^{1_{\{S_t^{\mathsf{M}}=+,\Theta_t=\mathsf{A}\}}}\left(\bar{\beta}^{\mathsf{B}}\right)^{1_{\{S_t^{\mathsf{M}}=+,\Theta_t=\mathsf{NA}\}}}}{b_{t-1}\left(\alpha^{\mathsf{B}}\right)^{1_{\{S_t^{\mathsf{M}}=+,\Theta_t=\mathsf{A}\}}}\left(\bar{\beta}^{\mathsf{B}}\right)^{1_{\{S_t^{\mathsf{M}}=+,\Theta_t=\mathsf{NA}\}}} + \bar{b}_{t-1}\left(\alpha^{\mathsf{W}}\right)^{1_{\{S_t^{\mathsf{M}}=+,\Theta_t=\mathsf{A}\}}}\left(\bar{\beta}^{\mathsf{W}}\right)^{1_{\{S_t^{\mathsf{M}}=+,\Theta_t=\mathsf{NA}\}}}}. \tag{5}$$

Using the definition of L_t, we obtain $L_t = L_{t-1} + R_t$, where

$$R_t \triangleq 1_{\{S_t^{\mathsf{M}}=+,\Theta_t=\mathsf{A}\}} \log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right) + 1_{\{S_t^{\mathsf{M}}=+,\Theta_t=\mathsf{NA}\}} \log\left(\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right). \tag{6}$$

Therefore, L_t is a random walk with i.i.d. random jumps R_t.

When the machine's type is $\Gamma \in \{\mathsf{B}, \mathsf{W}\}$, then the mean of the random jump is

$$\mathbb{E}^{\Gamma}[R_t] = p\alpha^{\Gamma} \log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right) + \bar{p}\bar{\beta}^{\Gamma} \log\left(\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right). \tag{7}$$

If $p < p^{\Gamma}$, it follows that the mean $\mathbb{E}^{\Gamma}[R_t]$ and, hence, the drift of the random walk L_t is negative so $L_t \to -\infty$ (see Gut 2009, Theorem 9.1). The reverse condition (with strict inequality) implies the divergence to ∞. If the mean $\mathbb{E}^{\Gamma}[R_t]$ equals 0 ($p = p^{\Gamma}$), then L_t is a martingale; hence, b_t oscillates (see Theorem 8.3.4 in Chung 2001).

Finally, p^{B} and p^{W} are such that $p^{\mathsf{B}} < p^{\mathsf{W}}$ because $\alpha^{\mathsf{B}}/\bar{\beta}^{\mathsf{B}} > \alpha^{\mathsf{W}}/\bar{\beta}^{\mathsf{W}}$, which is implied by Substitution (4)-(5). Q.E.D.

Appendix B: Main Set-up

Proof of Lemma 1. This lemma follows from the fact that posterior probabilities are continuous and monotone in b_{t-1} and the boundary values 1 and 0 are on different sides of r due to Substitution (4)-(5). In particular, the posterior probabilities are

$$\begin{aligned}
\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = +, S_t^{\mathsf{M}} = -, b_{t-1}) &= \frac{\alpha^{\mathsf{H}}(b_{t-1}\bar{\alpha}^{\mathsf{B}} + \bar{b}_{t-1}\bar{\alpha}^{\mathsf{W}})p}{\alpha^{\mathsf{H}}(b_{t-1}\bar{\alpha}^{\mathsf{B}} + \bar{b}_{t-1}\bar{\alpha}^{\mathsf{W}})p + \bar{\beta}^{\mathsf{H}}(b_{t-1}\beta^{\mathsf{B}} + \bar{b}_{t-1}\beta^{\mathsf{W}})\bar{p}}, \\
\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = -, S_t^{\mathsf{M}} = +, b_{t-1}) &= \frac{\bar{\alpha}^{\mathsf{H}}(b_{t-1}\alpha^{\mathsf{B}} + \bar{b}_{t-1}\alpha^{\mathsf{W}})p}{\bar{\alpha}^{\mathsf{H}}(b_{t-1}\alpha^{\mathsf{B}} + \bar{b}_{t-1}\alpha^{\mathsf{W}})p + \beta^{\mathsf{H}}(b_{t-1}\bar{\beta}^{\mathsf{B}} + \bar{b}_{t-1}\bar{\beta}^{\mathsf{W}})\bar{p}}.
\end{aligned}$$

By solving the following equations for b_{t-1},

$$\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = +, S_t^{\mathsf{M}} = -, b_{t-1}) = r \text{ and } \mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = -, S_t^{\mathsf{M}} = +, b_{t-1}) = r \tag{8}$$

we obtain the thresholds

$$b^- = \frac{\left(\frac{\bar{r}p\alpha^{\mathsf{H}}}{r\bar{p}\bar{\beta}^{\mathsf{H}}}\right)\bar{\alpha}^{\mathsf{W}} - \beta^{\mathsf{W}}}{\left(\frac{\bar{r}p\alpha^{\mathsf{H}}}{r\bar{p}\bar{\beta}^{\mathsf{H}}}\right)[\alpha^{\mathsf{B}} - \alpha^{\mathsf{W}}] + \beta^{\mathsf{B}} - \beta^{\mathsf{W}}} \text{ and } b^+ = \frac{\bar{\beta}^{\mathsf{W}} - \left(\frac{\bar{r}p\bar{\alpha}^{\mathsf{H}}}{r\bar{p}\beta^{\mathsf{H}}}\right)\alpha^{\mathsf{W}}}{\left(\frac{\bar{r}p\bar{\alpha}^{\mathsf{H}}}{r\bar{p}\beta^{\mathsf{H}}}\right)[\alpha^{\mathsf{B}} - \alpha^{\mathsf{W}}] + \beta^{\mathsf{B}} - \beta^{\mathsf{W}}}. \tag{9}$$

Thus, the result follows. Q.E.D.

Before proving Theorems 3 and 4, we first provide two constructive lemmas and their proofs.

LEMMA 3. *In the main set-up, the log-likelihood ratio process L_t is as follows.*

Case 1: If $b^+ > b^-$, then we have $L_t = L_{t-1} + R_t^{HM}(L_{t-1})$, where

$$R_t^{HM}(L_{t-1}) \triangleq \begin{cases} 1_{\{S_t^M=+\}}\left[1_{\{\Theta_t=A\}}\log\left(\frac{\alpha^B}{\alpha^W}\right) + 1_{\{\Theta_t=NA\}}\log\left(\frac{\bar{\beta}^B}{\bar{\beta}^W}\right)\right] & \text{if } L_{t-1} \geq L_h \\ 1_{\{S_t^H=+,S_t^M=+\}}\left[1_{\{\Theta_t=A\}}\log\left(\frac{\alpha^B}{\alpha^W}\right) + 1_{\{\Theta_t=NA\}}\log\left(\frac{\bar{\beta}^B}{\bar{\beta}^W}\right)\right] & \text{if } L_h > L_{t-1} > L_\ell \\ 1_{\{S_t^H=+\}}[\upsilon_1 + \upsilon_2] & \text{if } L_\ell \geq L_{t-1} \end{cases}$$

with $L_h \triangleq \log\left(\frac{b^+}{\bar{b}^+}\right)$, $L_\ell \triangleq \log\left(\frac{b^-}{\bar{b}^-}\right)$ *and*

$$\upsilon_1 \triangleq 1_{\{\Theta_t=A\}}\left(1_{\{S_t^M=-\}}\log\left(\frac{\bar{\alpha}^B}{\bar{\alpha}^W}\right) + 1_{\{S_t^M=+\}}\log\left(\frac{\alpha^B}{\alpha^W}\right)\right),$$

$$\upsilon_2 \triangleq 1_{\{\Theta_t=NA\}}\left(1_{\{S_t^M=-\}}\log\left(\frac{\beta^B}{\beta^W}\right) + 1_{\{S_t^M=+\}}\log\left(\frac{\bar{\beta}^B}{\bar{\beta}^W}\right)\right).$$

Case 2: If $b^+ \leq b^-$, then we have $L_t = L_{t-1} + R_t^{HM}(L_{t-1})$ where

$$R_t^{HM}(L_{t-1}) \triangleq \begin{cases} 1_{\{S_t^M=+\}}\left[1_{\{\Theta_t=A\}}\log\left(\frac{\alpha^B}{\alpha^W}\right) + 1_{\{\Theta_t=NA\}}\log\left(\frac{\bar{\beta}^B}{\bar{\beta}^W}\right)\right] & \text{if } L_{t-1} > L_h \\ 1_{\{\Theta_t=A\}}\nu_1 + 1_{\{\Theta_t=NA\}}\nu_2 & \text{if } L_h \geq L_{t-1} \geq L_\ell \\ 1_{\{S_t^H=+\}}[\upsilon_1 + \upsilon_2] & \text{if } L_\ell > L_{t-1} \end{cases} \tag{10}$$

with $L_h \triangleq \log\left(\frac{b^-}{\bar{b}^-}\right)$, $L_\ell \triangleq \log\left(\frac{b^+}{\bar{b}^+}\right)$ *and*

$$\nu_1 \triangleq 1_{\{S_t^M=+\}}\log\left(\frac{\alpha^B}{\alpha^W}\right) + 1_{\{S_t^M=-\}}1_{\{S_t^H=+\}}\log\left(\frac{\bar{\alpha}^B}{\bar{\alpha}^W}\right),$$

$$\nu_2 \triangleq 1_{\{S_t^M=+\}}\log\left(\frac{\bar{\beta}^B}{\bar{\beta}^W}\right) + 1_{\{S_t^M=-\}}1_{\{S_t^H=+\}}\log\left(\frac{\beta^B}{\beta^W}\right).$$

Proof of Lemma 3. As in the proof of Theorem 2, we derive a recursive expression for b_t in terms of b_{t-1} using Bayes' rule but this time using the decision-making procedure characterized in Lemma 1.

Case 1. When $b^+ > b^-$, we have the following three regimes

- $b_{t-1} \geq b^+$:

$$\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = -, S_t^{\mathsf{M}} = +, b_{t-1}) \geq r \tag{11}$$

$$\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = +, S_t^{\mathsf{M}} = -, b_{t-1}) < r \tag{12}$$

In this case, $S_t^{\mathsf{M}} = +$ is sufficient and necessary to act, which implies the machine overrides the human's signal $S_t^{\mathsf{H}} = -$. Further, $S_t^{\mathsf{H}} = +$ is also overruled by $S^{\mathsf{M}} = -$.

- $b^+ > b_{t-1} > b^-$:

$$\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = -, S_t^{\mathsf{M}} = +, b_{t-1}) < r \tag{13}$$

$$\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = +, S_t^{\mathsf{M}} = -, b_{t-1}) < r \tag{14}$$

In this case, $S_t^{\mathsf{M}} = S_t^{\mathsf{H}} = +$ is the only condition for acting.

- $b^- \geq b_{t-1}$:

$$\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = -, S_t^{\mathsf{M}} = +, b_{t-1}) < r \tag{15}$$

$$\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = +, S_t^{\mathsf{M}} = -, b_{t-1}) \geq r \tag{16}$$

In this case, $S_t^{\mathsf{H}} = +$ is sufficient and necessary to act. The signal of the human overrides the machine's signal in both conflicting cases.

Finally, the belief update when $b^+ > b^-$ is as follows.

$$b_t = \begin{cases} \left[1 + \frac{\bar{b}_{t-1}}{b_{t-1}} \left(\frac{\alpha^{\mathsf{W}}}{\alpha^{\mathsf{B}}}\right)^{1_{\{\Theta_t = \mathsf{A}, S_t^{\mathsf{M}} = +\}}} \left(\frac{\bar{\beta}^{\mathsf{W}}}{\bar{\beta}^{\mathsf{B}}}\right)^{1_{\{\Theta_t = \mathsf{NA}, S_t^{\mathsf{M}} = +\}}}\right]^{-1} & \text{if } b_{t-1} \geq b^+ \\ \left[1 + \frac{\bar{b}_{t-1}}{b_{t-1}} \left(\frac{\alpha^{\mathsf{W}}}{\alpha^{\mathsf{B}}}\right)^{1_{\{\Theta_t = \mathsf{A}, S_t^{\mathsf{M}} = +, S_t^{\mathsf{H}} = +\}}} \left(\frac{\bar{\beta}^{\mathsf{W}}}{\bar{\beta}^{\mathsf{B}}}\right)^{1_{\{\Theta_t = \mathsf{NA}, S_t^{\mathsf{M}} = +, S_t^{\mathsf{H}} = +\}}}\right]^{-1} & \text{if } b^+ > b_{t-1} > b^- \\ \left[1 + \frac{\bar{b}_{t-1}}{b_{t-1}} \zeta\right]^{-1} & \text{if } b^- \geq b_{t-1} \end{cases} \tag{17}$$

where ζ is defined in (2).

Case 2. When $b^+ \leq b^-$ we have the following three regimes

- $b_{t-1} > b^-$:

$$\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = -, S_t^{\mathsf{M}} = +, b_{t-1}) \geq r \tag{18}$$

$$\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = +, S_t^{\mathsf{M}} = -, b_{t-1}) < r \tag{19}$$

In this case, $S_t^{\mathsf{M}} = +$ is sufficient and necessary to act, which implies the machine overrides the human's signal $S_t^{\mathsf{H}} = -$. Further, $S_t^{\mathsf{H}} = +$ is also overruled by $S_t^{\mathsf{M}} = -$.

- $b^- \geq b_{t-1} \geq b^+$:

$$\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = -, S_t^{\mathsf{M}} = +, b_{t-1}) \geq r \tag{20}$$

$$\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = +, S_t^{\mathsf{M}} = -, b_{t-1}) \geq r \tag{21}$$

In this case, $S_t^{\mathsf{M}} = +$ or $S_t^{\mathsf{H}} = +$ is sufficient for acting.

- $b^+ > b_{t-1}$:

$$\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = -, S_t^{\mathsf{M}} = +, b_{t-1}) < r \tag{22}$$

$$\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = +, S_t^{\mathsf{M}} = -, b_{t-1}) \geq r \tag{23}$$

In this case, $S_t^{\mathsf{H}} = +$ is sufficient and necessary to act. The signal of the human overrides the machine's signal in both conflicting cases.

Finally the belief update if $b^- \leq b^+$ is as follows.

$$b_t = \begin{cases} \left[1 + \frac{\bar{b}_{t-1}}{b_{t-1}} \left(\frac{\alpha^{\mathsf{W}}}{\alpha^{\mathsf{B}}}\right)^{1_{\{\Theta_t = \mathsf{A}, S_t^{\mathsf{M}} = +\}}} \left(\frac{\bar{\beta}^{\mathsf{W}}}{\bar{\beta}^{\mathsf{B}}}\right)^{1_{\{\Theta_t = \mathsf{NA}, S_t^{\mathsf{M}} = +\}}}\right]^{-1} & \text{if } b_{t-1} > b^- \\ \left[1 + \frac{\bar{b}_{t-1}}{b_{t-1}} \iota\right]^{-1} & \text{if } b^- \geq b_{t-1} \geq b^+ \\ \left[1 + \frac{\bar{b}_{t-1}}{b_{t-1}} \zeta\right]^{-1} & \text{if } b^+ > b_{t-1} \end{cases} \tag{24}$$

where ζ is defined in (2) and

$$\iota \triangleq \left[\left(\frac{\alpha^{\mathsf{W}}}{\alpha^{\mathsf{B}}}\right)^{1_{\{S_t^{\mathsf{M}} = +\}}} \left(\frac{\bar{\alpha}^{\mathsf{W}}}{\bar{\alpha}^{\mathsf{B}}}\right)^{1_{\{S_t^{\mathsf{M}} = -, S_t^{\mathsf{H}} = +\}}}\right]^{1_{\{\Theta_t = \mathsf{A}\}}} \left[\left(\frac{\bar{\beta}^{\mathsf{W}}}{\bar{\beta}^{\mathsf{B}}}\right)^{1_{\{S_t^{\mathsf{M}} = +\}}} \left(\frac{\beta^{\mathsf{W}}}{\beta^{\mathsf{B}}}\right)^{1_{\{S_t^{\mathsf{M}} = -, S_t^{\mathsf{H}} = +\}}}\right]^{1_{\{\Theta_t = \mathsf{NA}\}}}$$

The log-likelihood ratio process L_t is then obtained by $L_t = \log\left(b_t/\bar{b}_t\right)$ in both cases. Q.E.D.

LEMMA 4. *Consider the following sequences of random variables*

- *i.i.d.* $Y_{1,t}$ *with* $\mathbb{E}[Y_{1,t}] > 0$ *and* $|Y_{1,t}| \leq Y_{1,h}$ *where* $\infty > Y_{1,h} > 0$ *for* $t = 1, \ldots,$
- *i.i.d.* $Y_{2,t}$ *with* $\mathbb{P}(Y_{2,t} > 0) > 0$ *and* $|Y_{2,t}| \leq Y_{2,h}$ *where* $\infty > Y_{2,h} > 0$ *for* $t = 1, \ldots,$
- *and, i.i.d.* $Y_{3,t}$ *with* $\mathbb{E}[Y_{3,t}] \geq 0$ *and* $|Y_{3,t}| \leq Y_{3,h}$ *where* $\infty > Y_{3,h} > 0$ *for* $t = 1, \ldots.$

Let Z_t *be a discrete stochastic process governed by* $Y_{i,t}$ *and let two thresholds* Z_ℓ *and* Z_h *be such that* $Z_h > Z_\ell$ *as follows.*

$$Z_{t+1} = Z_t + Y_{1,t} 1_{\{Z_t \geq Z_h\}} + Y_{2,t} 1_{\{Z_t \in (Z_\ell, Z_h)\}} + Y_{3,t} 1_{\{Z_t \leq Z_\ell\}}\,. \tag{25}$$

Then, $Z_t \xrightarrow{a.s.} \infty$.

Proof of Lemma 4. Note that the divergence immediately follows when the mean of $Z_{t+1} - Z_t$ is positive for any Z_t because in that case in all three regions process Z_t is driven by random walks drifting to ∞. Thus, we focus on the regime where the mean of $Z_{t+1} - Z_t$ is nonpositive in (Z_ℓ, Z_h) and 0 for $Z_t \leq Z_\ell$. To prove this result, we show that there exists a finite (but random) period τ such that process Z_t remains above $Z_t \geq Z_h$ for all $t > \tau$. This is sufficient for divergence to ∞ because process Z_t is governed by a random walk drifting to ∞ above Z_h. We prove this in two steps, but first, we define the following stopping times recursively by assuming, without loss of generality, that $Z_0 > Z_h$.

$$T_1 = \inf\{t : Z_t < Z_h\}, \tag{26}$$

$$\tilde{T}_1 = \inf\{t > T_1 : Z_t \geq L_h\}, \tag{27}$$

$$T_i = \inf\{t > \tilde{T}_{i-1} : Z_t < Z_h\} \ \forall i > 1, \tag{28}$$

$$\tilde{T}_i = \inf\{t > T_i : Z_t \geq Z_h\} \ \forall i > 1\,. \tag{29}$$

Here, if one of the sets above is empty, i.e., no such t exists, we assign $T_i = \infty$ (and respectively $\tilde{T}_i = \infty$) for the corresponding set.

Step 1. In the first step, we prove that $\mathbb{P}(\tilde{T}_i < \infty \,|\, T_i < \infty) = 1$ for all i. In particular, if process Z_t goes below Z_h once, then with probability one, it will cross up Z_h in finite steps. This result also implies that the sequence of infinite stopping times (if it exists) is started by $T_j = \infty$ (but not $\tilde{T}_j = \infty$) for some j. To prove this result, first fix i and assume that T_i is finite; then we define a new sequence of stopping times for process Z_t.

$$V_1 = \inf\{t > T_i : Z_t \le Z_\ell\} \tag{30}$$

$$\tilde{V}_1 = \inf\{t > V_1 : Z_t > Z_\ell\} \tag{31}$$

$$V_i = \inf\{t > \tilde{V}_{i-1} : Z_t \le Z_\ell\} \;\; \forall i > 1, \tag{32}$$

$$\tilde{V}_i = \inf\{t > V_i : Z_t > Z_\ell\} \;\; \forall i > 1. \tag{33}$$

First, note that V_i and $\tilde{V}_i$ for $i \ge 1$ are proper random variables, i.e., they take finite values with probability one. This is because

$$\mathbb{P}(V_i < \infty \,|\, \tilde{V}_{i-1} < \infty) = 1 \tag{34}$$

$$\mathbb{P}(\tilde{V}_i < \infty \,|\, V_i < \infty) = 1\,. \tag{35}$$

The first equality holds because as discussed at the beginning of this proof, we focus on the case where the mean of $Z_{t+1} - Z_t$ in (Z_ℓ, Z_h) is either negative or 0. If negative, Z_t is governed by a random walk drifting to $-\infty$ in (Z_ℓ, Z_h); thus it crosses Z_ℓ with probability one in finite steps (see, Theorem 9.1 in Gut 2009, p. 70). If 0, Z_t in (Z_ℓ, Z_h) is a martingale and oscillates (see, Theorem 8.3 in Gut 2009, p. 68). The second equality holds because Z_t is governed again by an oscillating random walk when $Z_t < Z_\ell$ thus it crosses Z_ℓ in finite steps. Furthermore, $T_i < \infty$; thus, V_1 is also finite. Therefore, it follows that $V_i < \infty$ and $\tilde{V}_i < \infty$ for $i \ge 1$.

To prove that $\tilde{T}_i$ is finite, we will show that there exists a finite j such that $\tilde{T}_i < V_j$. To do so, define events $A_t = \{\tilde{T}_i \ge V_t\}$. Then, the Borel-Cantelli lemma (Resnick 2014, p. 102) implies the following,

$$\sum_{t=T_i}^{\infty} \mathbb{P}(A_t) < \infty \Rightarrow \mathbb{P}(\limsup_{i\to\infty} A_t) = 0 \tag{36}$$

First, consider $\mathbb{P}(A_1) = \mathbb{P}(\tilde{T}_i \ge V_1)$, this probability is bounded, i.e., $\mathbb{P}(A_1) < \delta < 1$ because $Z_h - Z_\ell$ is bounded and Z_t has positive size jumps with positive probability in (Z_ℓ, Z_h). Proceeding similarly, we obtain the following

$$\mathbb{P}(\tilde{T}_i \ge V_t) = \mathbb{P}(\tilde{T}_i \ge V_{t-1})\mathbb{P}(\tilde{T}_i \ge V_t \,|\, \tilde{T}_i \ge V_{t-1}) < \delta^t \tag{37}$$

Therefore, it follows that $\mathbb{P}(\limsup_{i\to\infty} A_t) = 0$, which implies that there exists a finite j such that $\tilde{T}_i < V_j$. As discussed, V_j is finite; thus, it follows that $\mathbb{P}(\tilde{T}_i < \infty \,|\, T_i < \infty) = 1$.

Step 2. In this step, we show that there exists a finite j such that $T_j = \infty$ and $\tilde{T}_k < \infty$ for all $k < j$. Define the following events $E_t = \{T_t < \infty\}$ for $t > 1$. Then, the Borel-Cantelli lemma implies the following

$$\sum_{t=1}^{\infty} \mathbb{P}(E_t) < \infty \Rightarrow \mathbb{P}(\limsup_{t\to\infty} E_t) = 0\,. \tag{38}$$

In particular, if the summation condition is satisfied, then events E_t cannot occur infinitely many times, i.e., there exists a finite j such that $T_j = \infty$. To show that the summation condition is indeed satisfied, we construct a finite upper bound for it. Above Z_h, process Z_t is driven by a random walk drifting to ∞. Thus, stopping time T_1 is defective, i.e., $\mathbb{P}(T_1 < \infty) < \varphi < 1$ for some φ (see Theorem 9.1 in Gut 2009, p.70). Moreover, we bound $\mathbb{P}(T_2 < \infty)$ as follows:

$$\mathbb{P}(T_2 < \infty) = \mathbb{P}(T_2 < \infty \,|\, \tilde{T}_1 < \infty)\mathbb{P}(\tilde{T}_1 < \infty \,|\, T_1 < \infty)\mathbb{P}(T_1 < \infty) < \varphi^2\,. \tag{39}$$

In Step 1 of this proof, we show that $\mathbb{P}(\tilde{T}_1 < \infty \,|\, T_1 < \infty) = 1$. Moreover, we have $\mathbb{P}(T_2 < \infty \,|\, \tilde{T}_1 < \infty) < \varphi < 1$ because $\tilde{T}_1 < \infty$ implies that process Z_t crosses Z_h up in finite steps, and after crossing Z_h, process Z_t is again driven by the same random walk drifting to ∞. Proceeding similarly, it follows that $\mathbb{P}(E_t) < \varphi^t$. Hence, there exists a finite (but random) j such that $T_j = \infty$, and Step 1 implies that $\tilde{T}_k < \infty$ for $k < j$ because $T_k < \infty$. Therefore, after sufficiently large t, process Z_t always remains above Z_h and diverges to ∞. Q.E.D.

Proof of Theorem 3. In this proof, we focus on the log-likelihood ratio process L_t because as also discussed in the proof of Theorem 2, the continuous mapping theorem implies that the limit of L_t characterizes the limit of b_t. Our proof is based on analyzing the mean of $L_{t+1} - L_t$ when L_t takes different values in comparison to L_h and L_ℓ for Case 1 and Case 2 characterized in Lemma 3. In both cases, process L_t is governed by different random walks with random jumps whose means and size change depending on the previous state[1] L_{t-1}. Trivial cases arise when the mean of $L_{t+1} - L_t$ always remains positive regardless of L_t. In particular, L_t diverges to ∞ when the mean of $L_{t+1} - L_t$ is always positive despite changing values of L_t due to the strong law of large numbers (Resnick 2014) because three random walks (for $L_t \geq L_h$, $L_t \in (L_h, L_\ell)$ and $L_t \leq L_\ell$ under Case 1) that establish the trajectory of process L_t all diverge to ∞ (see Theorem 8.3 in Gut 2009, p. 68). As a result, b_t converges to 0 as L_t diverges to ∞. We consider the different values of p in the statement of the theorem separately.

---

[1] The log-likelihood ratio process L_t is similar to the *oscillating random walk* defined in Kemperman (1974) such that the partition of $\mathbb{R}$ consists of $(-\infty, L_\ell]$, (L_ℓ, L_h) and $[L_h, \infty)$.

Step 1. Let $p \le p^{\mathsf{B}}$. To prove that L_t and hence b_t oscillate, we need to show that there exists a number that process L_t crosses infinitely often (see, for instance, Vatutin and Wachtel 2009 for a mathematical definition of oscillation). This property (oscillation) holds for a random walk with a noise term whose mean is 0 (see Theorem 8.2 in Gut 2009, p. 68). However, process L_t does not satisfy this property because the mean of $L_{t+1} - L_t$ is always positive when $L_t \le L_\ell$ in Case 1 ($L_t < L_\ell$ in Case 2) as discussed. Nevertheless, we can show that process L_t oscillates because the mean of $L_{t+1} - L_t$ is either negative or zero when $p \le p^{\mathsf{B}}$ for $L_t \ge L_h$ in Case 1 ($L_t > L_h$ in Case 2). Therefore, process L_t returns to interval (L_ℓ, L_h) in finite steps after lying outside it. Oscillation is regardless of the sign of the mean of $L_{t+1} - L_t$ in (L_ℓ, L_h) because process L_t goes out of (L_ℓ, L_h) in finite steps. Specifically, in finite steps, process L_t i) crosses L_h up if the mean of $L_{t+1} - L_t$ in (L_ℓ, L_h) is positive, ii) crosses L_ℓ down if it is negative, and iii) goes out of (L_ℓ, L_h) if it is zero.

To illustrate this process, we focus on Case 1 and the setting where the mean of $L_{t+1} - L_t$ in (L_ℓ, L_h) is negative. Define the following stopping times

$$J = \inf\{t \ge 1 : L_t \le L_\ell\} \tag{40}$$

$$\tilde{J} = \inf\{t \ge 1 : L_t > L_\ell\} \tag{41}$$

If J and $\tilde{J}$ are proper random variables (i.e., $\mathbb{P}(J < \infty) = \mathbb{P}(\tilde{J} < \infty) = 1$), then process L_t oscillates. Assume $L_0 > L_\ell$ without loss of generality; then J is a proper (almost surely finite) random variable because process L_t is driven by a random walk drifting to $-\infty$ for $L_t > L_\ell$. Thus, in finite steps it will cross L_ℓ down. After J steps, process L_t is driven by a random walk drifting to[2] ∞, thus it crosses L_ℓ up in finite steps, so $\tilde{J} - J$ is a proper random variable. Since J is also proper, $\tilde{J} = \tilde{J} - J + J$ is also proper. Since L_t crosses L_ℓ infinitely often, b_t crosses $1/(1 + \exp(L_\ell))$ infinitely often and, hence, oscillates. The same approach can be repeated for the remaining setting where the mean of $L_{t+1} - L_t$ is positive or zero in (L_ℓ, L_h).

The oscillation property also implies that the process L_t and hence b_t is recurrent because the mean of the random jumps is bounded. With probability 1, process L_t will revisit interval (L_ℓ, L_h) for Case 1 in finite steps interval $[L_\ell, L_h]$ for Case 2 in finite steps. Furthermore, intervals that can be reached from (L_ℓ, L_h) with positive probability are also recurrent, which implies Corollary 1.

Step 2. Let $p > p^{\mathsf{B}}$. Then, the mean of $L_{t+1} - L_t$ when $L_t \le L_\ell$ for Case 1 and $L_t < L_\ell$ for Case 2 is positive (see Footnote 2). Further, we know from the proof of Theorem 2 that the mean of $L_{t+1} - L_t$ when $L_t \ge L_h$ for Case 1 and $L_t > L_h$ for $p > p^{\mathsf{B}}$ is also positive. In Case 1, the mean of $L_{t+1} - L_t$ for $L_t \in (L_\ell, L_h)$ is

$$\alpha^{\mathsf{H}}\alpha^{\mathsf{B}} p \log\left[\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right] + \bar{\beta}^{\mathsf{H}}\bar{\beta}^{\mathsf{B}}\bar{p}\log\left[\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right] = \bar{\beta}^{\mathsf{H}}\left[\alpha^{\mathsf{B}} p \log\left[\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right] + \bar{\beta}^{\mathsf{B}}\bar{p}\log\left[\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right]\right] + \bar{\beta}^{\mathsf{H}}\alpha^{\mathsf{B}}\left[\frac{\alpha^{\mathsf{H}}}{\bar{\beta}^{\mathsf{H}}} - 1\right]\log\left[\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right].$$

---

[2] This claim can be proved by invoking Lemma A.2 in Harrison et al. (2012) to show that $\mathbb{E}^{\mathsf{B}}[R_t^{\mathsf{HM}}(L_{t-1})] > 0$ for $L_{t-1} \le L_\ell$ in Case 1, and $L_{t-1} < L_\ell$ in Case 2.

Note that the term above is positive for $p > p^{\mathsf{B}}$. Hence, as discussed L_t diverges to ∞ and b_t converges to 1 for Case 1 when p is larger than p^{B}. In Case 2, the mean of $L_{t+1} - L_t$ for $L_t \in [L_\ell, L_h]$ is not necessarily positive. When it is positive, we immediately obtain the same result. Nevertheless, we obtain the same divergence despite having negative or zero mean of $L_{t+1} - L_t$ for $L_t \in [L_\ell, L_h]$ as long as the means of $L_{t+1} - L_t$ for $L_t < L_\ell$ and $L_t > L_h$ are positive, which is true as proven in Lemma 4. Thus, using Lemma 4, we capture all possible values of L_t with respect to L_h and L_ℓ in Cases 1 and 2 for $p > p^{\mathsf{B}}$, and show that L_t diverges to ∞, which implies b_t converges to 1. Hence, we conclude the proof. Q.E.D.

Proof of Theorem 4. The first part of the result in this theorem when $p \le p^{\mathsf{W}}$ follows from Lemma 4 by considering the reflection of the stochastic process. In particular, updating the conditions in the statement of Lemma 4 as $\mathbb{E}[Y_{1,t}] \le 0$, $\mathbb{P}(Y_{2,t} < 0) > 0$ and $\mathbb{E}[Y_{3,t}] < 0$ would imply $Z_t \xrightarrow{\text{a.s.}} -\infty$ in that lemma. This is because, the mean of $L_{t+1} - L_t$ is nonpositive when $L_t \in [L_h, \infty)$ in Case 1 (and $L_t \in (L_h, \infty)$ in Case 2); and is negative when $L_t \in (-\infty, L_\ell]$ in Case 1 (and $L_t \in (-\infty, L_\ell)$ in Case 2).[3] Thus, the case for $p \le p^{\mathsf{W}}$ follows from Lemma 4 with a slight modification.

To prove the second part of this result $p > p^{\mathsf{W}}$, we focus on Case 1 (Case 2 can be addressed by adjusting weak and strict inequalities in the same way) by assuming the mean of $L_{t+1} - L_t$ is negative when $L_t \in (L_\ell, L_h)$ and define the following stopping times by assuming $L_0 > L_h$ without loss of generality.

$$T_1 = \inf\{t : L_t < L_h\} \tag{42}$$

$$\tilde{T}_1 = \inf\{t > T_1 : L_t \ge L_h\} \tag{43}$$

$$T_i = \inf\{t > \tilde{T}_{i-1} : L_t < L_h\} \qquad \text{for } i \ge 2 \tag{44}$$

$$\tilde{T}_i = \inf\{t > T_i : L_t \ge L_h\} \qquad \text{for } i \ge 2 \tag{45}$$

Here, if a set is empty, then the stopping times takes the value of ∞. Therefore, if one of the stopping times T_i or $\tilde{T}_i$ is not finite for some i, then all the following stopping times for $j > i$ also are ∞. We first show that $\sum_{i=1}^{\infty} \mathbb{P}(T_i < \infty) < \infty$ and then use the Borel-Cantelli lemma to deduce that there exists a finite j such that $\mathbb{P}(T_j = \infty \text{ or } \tilde{T}_j = \infty) = 1$. If $T_j = \infty$, then $L_t \to \infty$; otherwise, ($\tilde{T}_j = \infty$) then $L_t \to -\infty$. Thus, $L_t \to L$, where $L \in \{-\infty, \infty\}$; hence b_t converges to a Bernoulli random variable.

First, consider T_1; it follows that $\mathbb{P}(T_1 < \infty) < \xi < 1$ (i.e., T_1 is a defective random variable) because process L_t is driven by a random walk drifting to ∞ when $L_t > L_h$ and $p > p^{\mathsf{W}}$. Next, considering $\tilde{T}_1$ and T_2; we obtain that

$$\mathbb{P}(\tilde{T}_1 < \infty) = \mathbb{P}(\tilde{T}_1 < \infty \,|\, T_1 < \infty)\mathbb{P}(T_1 < \infty) < \xi^2 \tag{46}$$

---

[3] The second part of this claim can be proved by invoking Lemma A.2 in Harrison et al. (2012) to show that $\mathbb{E}^{\mathsf{B}}[R_t^{\mathsf{HM}}(L_{t-1})] < 0$ for $L_{t-1} \le L_\ell$ in Case 1, and $L_{t-1} < L_\ell$ in Case 2.

$$\mathbb{P}(T_2 < \infty) = \mathbb{P}(T_2 < \infty \,|\, \tilde{T}_1 < \infty)\mathbb{P}(\tilde{T}_1 < \infty) < \xi^3 \tag{47}$$

Here, the first term is strictly less than 1, i.e., $\mathbb{P}(\tilde{T}_1 < \infty \,|\, T_1 < \infty) < \xi < 1$ because process L_t is driven by random walks drifting to $-\infty$ after T_1. The inequality in the second line follows because L_t is driven by a random walk drifting to ∞. Proceeding similarly, it follows that $\mathbb{P}(T_i < \infty) < \xi^{(2i-1)}$ and hence $\sum_{i=1}^{\infty} \mathbb{P}(T_i < \infty)$ is finite.

Note that we assume at the beginning that the mean of $L_{t+1} - L_t$ is negative when $L_t \in (L_\ell, L_h)$. If it is positive, the same steps can be followed by redefining stopping times using L_ℓ as the threshold instead of L_h. If it is zero, then stopping times T_i is defined by considering the time when process L_t enters (L_ℓ, L_h) and $\tilde{T}_i$ is the time when L_t exists (L_ℓ, L_h) in any direction. The approach follows because outside (L_ℓ, L_h), process L_t diverges (either to ∞ or $-\infty$ depending on being above L_h or below L_ℓ) and it oscillates in (L_ℓ, L_h), which guarantees it remains in (L_ℓ, L_h) for finite steps.

More specifically, the main driver of this result is that process L_t may diverge to ∞ when above L_h and to $-\infty$ when below L_ℓ. The sign of the mean between L_ℓ and L_h does not affect this characteristic in the limit. Thus, we conclude the proof. Q.E.D.

Appendix C: Mistrust Bias against the Machine

Proof of Lemma 2. Note that Informativeness (1)-(3) imply that the posterior after + (after $-$) is larger than or equal to r (resp. lower than r). Thus, there exist thresholds $\lambda_\Gamma^-, \lambda_\Gamma^+ \in (0,1)$ for $\Gamma \in \{\mathsf{B}, \mathsf{W}\}$ given by

$$\lambda_\Gamma^- \triangleq \frac{r - \mathbb{P}^\Gamma(\Theta_t = \mathsf{A} \,|\, S^\mathsf{M} = -)}{\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S^\mathsf{H} = +) - \mathbb{P}^\Gamma(\Theta_t = \mathsf{A} \,|\, S^\mathsf{M} = -)}, \tag{48}$$

$$\lambda_\Gamma^+ \triangleq \frac{\mathbb{P}^\Gamma(\Theta_t = \mathsf{A} \,|\, S^\mathsf{M} = +) - r}{\mathbb{P}^\Gamma(\Theta_t = \mathsf{A} \,|\, S^\mathsf{M} = +) - \mathbb{P}(\Theta_t = \mathsf{A} \,|\, S^\mathsf{H} = -)}. \tag{49}$$

These equations imply that

$$\lambda \mathbb{P}(\Theta_t = \mathsf{A} \,|\, S^\mathsf{H} = +) + \bar{\lambda} \mathbb{P}^\mathsf{B}(\Theta_t = \mathsf{A} \,|\, S^\mathsf{M} = -) \geq r \;\; \Leftrightarrow \lambda \geq \lambda_\mathsf{B}^-, \tag{50}$$

$$\lambda \mathbb{P}(\Theta_t = \mathsf{A} \,|\, S^\mathsf{H} = -) + \bar{\lambda} \mathbb{P}^\mathsf{B}(\Theta_t = \mathsf{A} \,|\, S^\mathsf{M} = +) \leq r \;\; \Leftrightarrow \lambda \geq \lambda_\mathsf{B}^+, \tag{51}$$

$$\lambda \mathbb{P}(\Theta_t = \mathsf{A} \,|\, S^\mathsf{H} = +) + \bar{\lambda} \mathbb{P}^\mathsf{W}(\Theta_t = \mathsf{A} \,|\, S^\mathsf{M} = -) \leq r \;\; \Leftrightarrow \lambda \leq \lambda_\mathsf{W}^-, \tag{52}$$

$$\lambda \mathbb{P}(\Theta_t = \mathsf{A} \,|\, S^\mathsf{H} = -) + \bar{\lambda} \mathbb{P}^\mathsf{W}(\Theta_t = \mathsf{A} \,|\, S^\mathsf{M} = +) \geq r \;\; \Leftrightarrow \lambda \leq \lambda_\mathsf{W}^+, \tag{53}$$

because the left-hand sides of the first and third inequalities are increasing and the second and the fourth ones are decreasing in λ.

First, note that λ_Γ^+ is increasing in $\mathbb{P}^\Gamma(\Theta_t = \mathsf{A} \,|\, S^\mathsf{M} = +)$, and λ_Γ^- is decreasing in $\mathbb{P}^\Gamma(\Theta_t = \mathsf{A} \,|\, S^\mathsf{M} = -)$. We also know from Substitution (4)-(5) that $\mathbb{P}^\mathsf{B}(\Theta_t = \mathsf{A} \,|\, S^\mathsf{M} = +) > \mathbb{P}^\mathsf{W}(\Theta_t = \mathsf{A} \,|\, S^\mathsf{M} = +)$, and $\mathbb{P}^\mathsf{W}(\Theta_t = \mathsf{A} \,|\, S^\mathsf{M} = -) > \mathbb{P}^\mathsf{B}(\Theta_t = \mathsf{A} \,|\, S^\mathsf{M} = -)$. We conclude the proof by defining $\lambda_{min} \triangleq \max(\lambda_\mathsf{W}^-, \lambda_\mathsf{W}^+)$ and $\lambda_{max} = \min(\lambda_\mathsf{B}^-, \lambda_\mathsf{B}^+)$. Q.E.D.

Before proving Theorem 5, we first provide a constructive lemma and its proof.

LEMMA 5. *For $\lambda \in (\lambda_{min}, \lambda_{max})$, unique thresholds $b_\lambda^- \in (0,1)$ and $b_\lambda^+ \in (0,1)$ exist such that*

$$\lambda \mathbb{P}(\Theta_t = \mathsf{A} \,|\, S^H = +) + (1-\lambda)\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S^M = -, b_{t-1}) \geq r \;\; \Leftrightarrow b_{t-1} \leq b_\lambda^- \,, \tag{54}$$

$$\lambda \mathbb{P}(\Theta_t = \mathsf{A} \,|\, S^H = -) + (1-\lambda)\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S^M = +, b_{t-1}) \geq r \;\; \Leftrightarrow b_{t-1} \leq b_\lambda^+ \,. \tag{55}$$

Proof of Lemma 5. We obtain the thresholds by solving the following equations.

$$\lambda \mathbb{P}(\Theta_t = \mathsf{A} \,|\, S^{\mathsf{H}} = +) + (1-\lambda)\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S^{\mathsf{M}} = -, b_\lambda^-) = r \,, \tag{56}$$

$$\lambda \mathbb{P}(\Theta_t = \mathsf{A} \,|\, S^{\mathsf{H}} = -) + (1-\lambda)\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S^{\mathsf{M}} = +, b_\lambda^+) = r \,. \tag{57}$$

The closed-form expressions are

$$b_\lambda^- = \frac{\varphi_\lambda^-(1-\alpha^{\mathsf{W}}) - \beta^{\mathsf{W}}}{\beta^{\mathsf{B}} - \beta^{\mathsf{W}} + \varphi_\lambda^-(\alpha^{\mathsf{B}} - \alpha^{\mathsf{W}})} \text{ and } b_\lambda^+ = \frac{\bar{\beta}^{\mathsf{W}} - \varphi_\lambda^+ \alpha^{\mathsf{W}}}{\varphi_\lambda^+(\alpha^{\mathsf{B}} - \alpha^{\mathsf{W}}) + \beta^{\mathsf{B}} - \beta^{\mathsf{W}}}$$

where

$$\varphi_\lambda^- = \frac{p}{\bar{p}} \left[\frac{(1-\lambda) - r + \lambda \mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = +)}{r - \lambda \mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = +)} \right] \text{ and } \varphi_\lambda^+ = \frac{p}{\bar{p}} \left[\frac{(1-\lambda) - r + \lambda \mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = -)}{r - \lambda \mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = -)} \right].$$

Finally, we can similarly define $b_\lambda^{\mathsf{H}} = \min(b_\lambda^-, b_\lambda^+)$ and $b_\lambda^{\mathsf{M}} = \max(b_\lambda^-, b_\lambda^+)$ as in Corollary 1. Q.E.D.

Proof of Theorem 5. Note that Lemma 5 shows that there exist thresholds $b_\lambda^-, b_\lambda^+ \in (0,1)$ for $\lambda \in (\lambda_{min}, \lambda_{max})$ as in the case of Lemma 1. Thus, the exact same steps at which thresholds $b^-, b^+ \in (0,1)$ are replaced by $b_\lambda^-, b_\lambda^+ \in (0,1)$ in the proofs of Lemma 3, Theorems 3-4 will imply this result because the results in those theorems do not depend on the exact values of thresholds b^- and b^+, as long as they are interior to $(0,1)$. Q.E.D.

Proof of Theorem 6. Note that Lemma 1 also characterizes the DM's decision rule for the biased case. Nevertheless, the belief updating (6) is replaced with the one in Section 7.2 with a bias term μ when the machine's prediction is not correct. Thus, we need to modify Lemma 3 slightly to incorporate this change. In particular, the updated log-likelihood ratio process $\tilde{L}_t$ is as follows. Case 1: If $b^+ > b^-$, we have $\tilde{L}_t = \tilde{L}_{t-1} + \tilde{R}_t^{\mathsf{HM}}(\tilde{L}_{t-1})$ where

$$\tilde{R}_t^{\mathsf{HM}}(\tilde{L}_{t-1}) \triangleq \begin{cases} 1_{\{S_t^{\mathsf{M}}=+\}} \left[1_{\{\Theta_t=\mathsf{A}\}} \log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right) + 1_{\{\Theta_t=\mathsf{NA}\}} \mu \log\left(\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right) \right] & \text{if } \tilde{L}_{t-1} \geq L_h \\ 1_{\{S_t^{\mathsf{H}}=+, S_t^{\mathsf{M}}=+\}} \left[1_{\{\Theta_t=\mathsf{A}\}} \log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right) + 1_{\{\Theta_t=\mathsf{NA}\}} \mu \log\left(\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right) \right] & \text{if } L_h > \tilde{L}_{t-1} > L_\ell \\ 1_{\{S_t^{\mathsf{H}}=+\}} \left[\tilde{\upsilon}_1 + \tilde{\upsilon}_2 \right] & \text{if } L_\ell \geq \tilde{L}_{t-1} \end{cases}$$

with $L_h \triangleq \log\left(\frac{b^+}{\bar{b}^+}\right)$, $L_\ell \triangleq \log\left(\frac{b^-}{\bar{b}^-}\right)$ and

$$\tilde{\upsilon}_1 \triangleq 1_{\{\Theta_t=\mathsf{A}\}} \left(1_{\{S_t^{\mathsf{M}}=-\}} \mu \log\left(\frac{\bar{\alpha}^{\mathsf{B}}}{\bar{\alpha}^{\mathsf{W}}}\right) + 1_{\{S_t^{\mathsf{M}}=+\}} \log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right) \right),$$

$$\tilde{\upsilon}_2 \triangleq 1_{\{\Theta_t=\mathsf{NA}\}}\left(1_{\{S_t^{\mathsf{M}}=-\}}\log\left(\frac{\beta^{\mathsf{B}}}{\beta^{\mathsf{W}}}\right)+1_{\{S_t^{\mathsf{M}}=+\}}\mu\log\left(\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right)\right).$$

Case 2: If $b^+ \leq b^-$, we have $\tilde{L}_t = \tilde{L}_{t-1} + \tilde{R}_t^{\mathsf{HM}}(\tilde{L}_{t-1})$ where

$$\tilde{R}_t^{\mathsf{HM}}(\tilde{L}_{t-1}) \triangleq \begin{cases} 1_{\{S_t^{\mathsf{M}}=+\}}\left[1_{\{\Theta_t=\mathsf{A}\}}\log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right)+1_{\{\Theta_t=\mathsf{NA}\}}\mu\log\left(\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right)\right] & \text{if } \tilde{L}_{t-1} > L_h \\ 1_{\{\Theta_t=\mathsf{A}\}}\tilde{\nu}_1 + 1_{\{\Theta_t=\mathsf{NA}\}}\tilde{\nu}_2 & \text{if } L_h \geq \tilde{L}_{t-1} \geq L_\ell \\ 1_{\{S_t^{\mathsf{H}}=+\}}\left[\tilde{\upsilon}_1+\tilde{\upsilon}_2\right] & \text{if } L_\ell > \tilde{L}_{t-1} \end{cases} \tag{58}$$

with $L_h \triangleq \log\left(\frac{b^-}{\bar{b}^-}\right)$, $L_\ell \triangleq \log\left(\frac{b^+}{\bar{b}^+}\right)$ and

$$\tilde{\nu}_1 \triangleq 1_{\{S_t^{\mathsf{M}}=+\}}\log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right)+1_{\{S_t^{\mathsf{M}}=-\}}1_{\{S_t^{\mathsf{H}}=+\}}\mu\log\left(\frac{\bar{\alpha}^{\mathsf{B}}}{\bar{\alpha}^{\mathsf{W}}}\right),$$

$$\tilde{\nu}_2 \triangleq 1_{\{S_t^{\mathsf{M}}=+\}}\mu\log\left(\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right)+1_{\{S_t^{\mathsf{M}}=-\}}1_{\{S_t^{\mathsf{H}}=+\}}\log\left(\frac{\beta^{\mathsf{B}}}{\beta^{\mathsf{W}}}\right).$$

Note that the thresholds in the theorem are determined as the break-even points of the following equations.

$$p\alpha^{\mathsf{B}}\log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right)+\bar{p}\bar{\beta}^{\mathsf{B}}\mu^{\mathsf{B}}\log\left(\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right)=0 \tag{59}$$

$$p\alpha^{\mathsf{W}}\log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right)+\bar{p}\bar{\beta}^{\mathsf{W}}\mu^{\mathsf{W}}\log\left(\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right)=0 \tag{60}$$

$$p\alpha^{\mathsf{H}}\left(\bar{\alpha}^{\mathsf{B}}\mu^{\mathsf{H}}\log\left(\frac{\bar{\alpha}^{\mathsf{B}}}{\bar{\alpha}^{\mathsf{W}}}\right)+\alpha^{\mathsf{B}}\log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right)\right)+\bar{p}\bar{\beta}^{\mathsf{H}}\left(\beta^{\mathsf{B}}\log\left(\frac{\beta^{\mathsf{B}}}{\beta^{\mathsf{W}}}\right)+\bar{\beta}^{\mathsf{B}}\mu^{\mathsf{H}}\log\left(\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right)\right)=0 \tag{61}$$

The left-hand sides of these equations correspond to the mean of $\tilde{L}_{t+1}-\tilde{L}_t$ when evaluated at μ in place of the thresholds at corresponding values of $\tilde{L}_t$. In the remainder of the proof, we explain the sign of the mean of $\tilde{L}_{t+1}-\tilde{L}_t$ given the machine's type and the value of μ. When the sign of this mean is known, the results follow from the proofs of previous theorems on which we elaborate.

- when $\Gamma = \mathsf{B}$,

—$\mu \geq \mu^{\mathsf{B}}$ and $\mu > \mu^{\mathsf{H}}$ implies that the mean of $\tilde{L}_{t+1}-\tilde{L}_t$ for $L_\ell \geq \tilde{L}_t$ for Case 1 ($L_\ell > \tilde{L}_t$ for Case 2) is negative. Further, the mean of $\tilde{L}_{t+1}-\tilde{L}_t$ for $\tilde{L}_t \geq L_h$ for Case 1 ($\tilde{L}_t > L_h$ for Case 2) is nonpositive. Thus, the proof of the first part (for $p \leq p^{\mathsf{W}}$) of Theorem 4 applies to this parameter regime.

—$\mu^{\mathsf{B}} > \mu > \mu^{\mathsf{H}}$ implies that the mean of $\tilde{L}_{t+1}-\tilde{L}_t$ for $L_\ell \geq \tilde{L}_t$ for Case 1 ($L_\ell > \tilde{L}_t$ for Case 2) is positive. Further, the mean of $\tilde{L}_{t+1}-\tilde{L}_t$ for $\tilde{L}_t \geq L_h$ for Case 1 ($\tilde{L}_t > L_h$ for Case 2) is negative. Thus, the proof of the second part (for $p > p^{\mathsf{W}}$) of Theorem 4 applies to this parameter regime.

—$\mu^{\mathsf{H}} \geq \mu \geq \mu^{\mathsf{B}}$ implies that the mean of $\tilde{L}_{t+1}-\tilde{L}_t$ for $L_\ell \geq \tilde{L}_t$ for Case 1 ($L_\ell > \tilde{L}_t$ for Case 2) is nonnegative. Further, the mean of $\tilde{L}_{t+1}-\tilde{L}_t$ for $\tilde{L}_t \geq L_h$ for Case 1 ($\tilde{L}_t > L_h$ for Case 2) is nonpositive. Thus, the proof of the first part (for $p \leq p^{\mathsf{B}}$) of Theorem 3 applies to this parameter regime.

—if $\mu^{\mathsf{B}} > \mu$ and $\mu^{\mathsf{H}} \geq \mu$ implies that the mean of $\tilde{L}_{t+1} - \tilde{L}_t$ for $L_\ell \geq \tilde{L}_t$ for Case 1 ($L_\ell > \tilde{L}_t$ for Case 2) is nonnegative. Further, the mean of $\tilde{L}_{t+1} - \tilde{L}_t$ for $\tilde{L}_t \geq L_h$ for Case 1 ($\tilde{L}_t > L_h$ for Case 2) is positive. Thus, the proof of the second part (for $p > p^{\mathsf{B}}$) of Theorem 3 applies to this parameter regime.

- when $\Gamma = \mathsf{W}$,

—$\mu \geq \mu^{\mathsf{W}}$ implies that the mean of $\tilde{L}_{t+1} - \tilde{L}_t$ for $\tilde{L}_t \geq L_h$ for Case 1 ($\tilde{L}_t > L_h$ for Case 2) is nonpositive. Further, the mean of $\tilde{L}_{t+1} - \tilde{L}_t$ for $L_\ell \geq \tilde{L}_t$ for Case 1 ($L_\ell > \tilde{L}_t$ for Case 2) is negative for all $\mu \geq 1$. Thus, the proof of the first part (for $p \leq p^{\mathsf{W}}$) of Theorem 4 applies to this parameter regime.

—$\mu^{\mathsf{W}} > \mu$ implies that the mean of $\tilde{L}_{t+1} - \tilde{L}_t$ for $\tilde{L}_t \geq L_h$ for Case 1 ($\tilde{L}_t > L_h$ for Case 2) is positive. Further, the mean of $\tilde{L}_{t+1} - \tilde{L}_t$ for $L_\ell \geq \tilde{L}_t$ for Case 1 ($L_\ell > \tilde{L}_t$ for Case 2) is negative for all $\mu \geq 1$. Thus, the proof of the second part (for $p > p^{\mathsf{W}}$) of Theorem 4 applies to this parameter regime.

Finally, we show that $\mu^{\mathsf{H}} > 1$. Note that the following term is always positive when evaluated at $\mu^{\mathsf{H}} = 1$; see Footnote 2.

$$p\alpha^{\mathsf{H}}\left(\bar{\alpha}^{\mathsf{B}}\mu^{\mathsf{H}}\log\left(\frac{\bar{\alpha}^{\mathsf{B}}}{\bar{\alpha}^{\mathsf{W}}}\right) + \alpha^{\mathsf{B}}\log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right)\right) + \bar{p}\bar{\beta}^{\mathsf{H}}\left(\beta^{\mathsf{B}}\log\left(\frac{\beta^{\mathsf{B}}}{\beta^{\mathsf{W}}}\right) + \bar{\beta}^{\mathsf{B}}\mu^{\mathsf{H}}\log\left(\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right)\right)$$

To make this equal to 0, μ^{H} has to be larger than 1 because it is decreasing in μ^{H}. Hence, we conclude the proof. Q.E.D.

Appendix D: Complementarity

Before providing the proof of Theorem 7, we first provide some constructive lemmas and their proofs. The first lemma is analogous to Lemmas 1 and 2 in terms of characterizing the DM's decision rule as a function of her belief.

Lemma 6. *Unique thresholds* b_{C}^- *and* b_{C}^+ *exist such that*

$$\mathbb{P}(\Theta_t = \mathsf{A} \mid S_t^H = +, S_t^M = -, b_{t-1}) \geq r \iff b_{t-1} \leq b_{\mathsf{C}}^-, \tag{62}$$

$$\mathbb{P}(\Theta_t = \mathsf{A} \mid S_t^H = -, S_t^M = +, b_{t-1}) \geq r \iff b_{t-1} \leq b_{\mathsf{C}}^+. \tag{63}$$

Proof of Lemma 6. As mentioned in the proof of Lemma 1, the posterior probabilities in the statement of this lemma are continuous and monotone in b_{t-1} and the boundary values 1 and 0 are on different sides of r due to Complementarity (13)-(14). We next provide the posterior probabilities.

$$\mathbb{P}(\Theta_t = \mathsf{A} \mid S_t^{\mathsf{H}} = +, S_t^{\mathsf{M}} = -, b_{t-1}) = \frac{\alpha^{\mathsf{H}}(b_{t-1}\bar{\alpha}^{\mathsf{C}+} + \bar{b}_{t-1}\bar{\alpha}^{\mathsf{C}-})p}{\alpha^{\mathsf{H}}(b_{t-1}\bar{\alpha}^{\mathsf{C}+} + \bar{b}_{t-1}\bar{\alpha}^{\mathsf{C}-})p + \bar{\beta}^{\mathsf{H}}(b_{t-1}\beta^{\mathsf{C}+} + \bar{b}_{t-1}\beta^{\mathsf{C}-})\bar{p}} \tag{64}$$

$$\mathbb{P}(\Theta_t = \mathsf{A} \mid S_t^{\mathsf{H}} = -, S_t^{\mathsf{M}} = +, b_{t-1}) = \frac{\bar{\alpha}^{\mathsf{H}}(b_{t-1}\alpha^{\mathsf{C}+} + \bar{b}_{t-1}\alpha^{\mathsf{C}-})p}{\bar{\alpha}^{\mathsf{H}}(b_{t-1}\alpha^{\mathsf{C}+} + \bar{b}_{t-1}\alpha^{\mathsf{C}-})p + \beta^{\mathsf{H}}(b_{t-1}\bar{\beta}^{\mathsf{C}+} + \bar{b}_{t-1}\bar{\beta}^{\mathsf{C}-})\bar{p}} \tag{65}$$

We solve the following equations for b_{t-1}

$$\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = +, S_t^{\mathsf{M}} = -, b_{t-1}) = r \text{ and } \mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = -, S_t^{\mathsf{M}} = +, b_{t-1}) = r\,. \tag{66}$$

We then obtain

$$b_{\mathsf{C}}^{-} = \frac{\left(\frac{\bar{r}p\alpha^{\mathsf{H}}}{r\bar{p}\bar{\beta}^{\mathsf{H}}}\right)\bar{\alpha}^{\mathsf{C-}} - \beta^{\mathsf{C-}}}{\left(\frac{\bar{r}p\alpha^{\mathsf{H}}}{r\bar{p}\bar{\beta}^{\mathsf{H}}}\right)[\alpha^{\mathsf{C+}} - \alpha^{\mathsf{C-}}] + \beta^{\mathsf{C+}} - \beta^{\mathsf{C-}}} \text{ and } b_{\mathsf{C}}^{+} = \frac{\bar{\beta}^{\mathsf{C-}} - \left(\frac{\bar{r}p\bar{\alpha}^{\mathsf{H}}}{r\bar{p}\beta^{\mathsf{H}}}\right)\alpha^{\mathsf{C-}}}{\left(\frac{\bar{r}p\bar{\alpha}^{\mathsf{H}}}{r\bar{p}\beta^{\mathsf{H}}}\right)[\alpha^{\mathsf{C+}} - \alpha^{\mathsf{C-}}] + \beta^{\mathsf{C+}} - \beta^{\mathsf{C-}}}\,. \tag{67}$$

The result follows. Q.E.D.

LEMMA 7. *Assume that the accuracy parameters* $(\alpha^{\Gamma}, \beta^{\Gamma})$ *for* $\Gamma \in \{\mathsf{C+}, \mathsf{C-}\}$ *satisfy Complementarity* (13)-(14)*. Then, it follows that*

$$\alpha^{\mathsf{C+}} > \alpha^{\mathsf{C-}} \text{ and } \beta^{\mathsf{C+}} < \beta^{\mathsf{C-}}\,. \tag{68}$$

Proof of Lemma 7. We prove this result by obtaining two subsets of $[0,1] \times [0,1]$ using Complementarity (13)-(14) that contain accuracy parameters $(\alpha^{\mathsf{C+}}, \beta^{\mathsf{C+}})$ and respectively $(\alpha^{\mathsf{C-}}, \beta^{\mathsf{C-}})$. These sets are disjoint besides the lowest sensitivity in the first set is higher than the highest sensitivity in the second set, and vice a versa for specificity.

We first define the following constants.

$$\eta \triangleq \frac{\alpha^{\mathsf{H}}\bar{r}p}{\bar{\beta}^{\mathsf{H}}r\bar{p}} \text{ and } \tilde{\eta} \triangleq \frac{\bar{\alpha}^{\mathsf{H}}\bar{r}p}{\beta^{\mathsf{H}}r\bar{p}} \tag{69}$$

Note that Informativeness (1) implies that $\eta \geq 1$ and $\tilde{\eta} < 1$. Now, using these and (13)-(14), we obtain that

$$\mathcal{G}^{\mathsf{C+}} \triangleq \{(\alpha, \beta) \in [0,1] \times [0,1] : \beta + \eta\alpha > \eta \text{ and } 1 > \beta + \tilde{\eta}\alpha\} \tag{70}$$

$$\mathcal{G}^{\mathsf{C-}} \triangleq \{(\alpha, \beta) \in [0,1] \times [0,1] : \beta + \eta\alpha \leq \eta \text{ and } 1 \leq \beta + \tilde{\eta}\alpha\} \tag{71}$$

First observe that $\mathcal{G}^{\mathsf{C+}} \cap \mathcal{G}^{\mathsf{C-}} = \emptyset$. Next, we define $\hat{\mathcal{G}}^{\mathsf{C+}} \triangleq \{(\alpha, \beta) \in [0,1] \times [0,1] : \beta + \eta\alpha \geq \eta$ and $1 \geq \beta + \tilde{\eta}\alpha\}$. It follows that

$$\min_{(\alpha,\beta) \in \mathcal{G}^{\mathsf{C+}}} \alpha > \min_{(\alpha,\beta) \in \hat{\mathcal{G}}^{\mathsf{C+}}} \alpha$$

because $\mathcal{G}^{\mathsf{C+}} \subset \tilde{\mathcal{G}}^{\mathsf{C+}}$ and the optimal (α, β) pair minimizing α over $\hat{\mathcal{G}}^{\mathsf{C+}}$ is at the corner solution where both inequalities binding; and that point is not in $\mathcal{G}^{\mathsf{C+}}$ since inequalities are weak. In particular, we obtain by solving the system of equations that $\min_{(\alpha,\beta) \in \hat{\mathcal{G}}^{\mathsf{C+}}} \alpha = (\eta - 1)/(\eta - \tilde{\eta})$.

Next, we consider $\max_{(\alpha,\beta) \in \mathcal{G}^{\mathsf{C-}}} \alpha$. Again, the optimal solution is at the corner where both inequalities are binding, and thus it follows that $\max_{(\alpha,\beta) \in \mathcal{G}^{\mathsf{C-}}} \alpha = (\eta - 1)/(\eta - \tilde{\eta})$.

Combining these, we get that $\alpha^{\mathsf{C+}} > \alpha^{\mathsf{C-}}$. Following the same steps to minimize and maximize β over these sets yields $\beta^{\mathsf{C+}} < \beta^{\mathsf{C-}}$. Hence, we conclude the proof. Q.E.D.

We are now ready to prove Theorem 7.

Proof of Theorem 7. We prove this result in 3 steps. Note that the thresholds b_{C}^- and b_{C}^+ defined in Lemma 6 imply the DM's decision rule. In the first step, we provide the log-likelihood ratio process generated by the DM's decision and signal realizations. In the second, we analyze the mean of the random jumps that govern the log-likelihood ratio process. In the last step, we combine our findings to derive the limit result.

Step 1. Assume that[4] $b_{\mathsf{C}}^+ > b_{\mathsf{C}}^-$. Then, we have the following three regimes

- $b_{t-1} > b_{\mathsf{C}}^+$:

$$\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = -, S_t^{\mathsf{M}} = +, b_{t-1}) < r \tag{72}$$

$$\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = +, S_t^{\mathsf{M}} = -, b_{t-1}) < r \tag{73}$$

In this case, $S_t^{\mathsf{M}} = S_t^{\mathsf{H}} = +$ is sufficient and necessary to act, which implies the DM decides as in type C^+.

- $b_{\mathsf{C}}^+ \geq b_{t-1} > b_{\mathsf{C}}^-$:

$$\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = -, S_t^{\mathsf{M}} = +, b_{t-1}) \geq r \tag{74}$$

$$\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = +, S_t^{\mathsf{M}} = -, b_{t-1}) < r \tag{75}$$

In this case, $S_t^{\mathsf{M}} = +$ is necessary and sufficient condition for acting.

- $b_{\mathsf{C}}^- \geq b_{t-1}$:

$$\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = -, S_t^{\mathsf{M}} = +, b_{t-1}) \geq r \tag{76}$$

$$\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = +, S_t^{\mathsf{M}} = -, b_{t-1}) \geq r \tag{77}$$

In this case, the DM decides to act after $S_t^{\mathsf{H}} = +$ or $S_t^{\mathsf{M}} = +$.

Finally, the belief update is as follows.

$$b_t = \begin{cases} \left[1 + \frac{\bar{b}_{t-1}}{b_{t-1}} \left(\frac{\alpha^{\mathsf{C}^-}}{\alpha^{\mathsf{C}^+}}\right)^{1_{\{\Theta_t = \mathsf{A}, S_t^{\mathsf{M}} = +, S_t^{\mathsf{H}} = +\}}} \left(\frac{\bar{\beta}^{\mathsf{C}^-}}{\bar{\beta}^{\mathsf{C}^+}}\right)^{1_{\{\Theta_t = \mathsf{NA}, S_t^{\mathsf{M}} = +, S_t^{\mathsf{H}} = +\}}}\right]^{-1} & \text{if } b_{t-1} > b_{\mathsf{C}}^+ \\ \left[1 + \frac{\bar{b}_{t-1}}{b_{t-1}} \left(\frac{\alpha^{\mathsf{C}^-}}{\alpha^{\mathsf{C}^+}}\right)^{1_{\{\Theta_t = \mathsf{A}, S_t^{\mathsf{M}} = +\}}} \left(\frac{\bar{\beta}^{\mathsf{C}^-}}{\bar{\beta}^{\mathsf{C}^+}}\right)^{1_{\{\Theta_t = \mathsf{NA}, S_t^{\mathsf{M}} = +\}}}\right]^{-1} & \text{if } b_{\mathsf{C}}^+ \geq b_{t-1} > b_{\mathsf{C}}^- \\ \left[1 + \frac{\bar{b}_{t-1}}{b_{t-1}} \tilde{\zeta}\right]^{-1} & \text{if } b_{\mathsf{C}}^- \geq b_{t-1} \end{cases} \tag{78}$$

where $\tilde{\zeta}$ is defined as

$$\tilde{\zeta} \triangleq \left(\frac{\alpha^{\mathsf{C}^-}}{\alpha^{\mathsf{C}^+}}\right)^{1_{\{\Theta_t = \mathsf{A}, S_t^{\mathsf{M}} = +\}}} \left(\frac{\bar{\alpha}^{\mathsf{C}^-}}{\bar{\alpha}^{\mathsf{C}^+}}\right)^{1_{\{\Theta_t = \mathsf{A}, S_t^{\mathsf{M}} = -, S_t^{\mathsf{H}} = +\}}} \left(\frac{\bar{\beta}^{\mathsf{C}^-}}{\bar{\beta}^{\mathsf{C}^+}}\right)^{1_{\{\Theta_t = \mathsf{NA}, S_t^{\mathsf{M}} = +\}}} \left(\frac{\beta^{\mathsf{C}^-}}{\beta^{\mathsf{C}^+}}\right)^{1_{\{\Theta_t = \mathsf{NA}, S_t^{\mathsf{M}} = -, S_t^{\mathsf{H}} = -\}}} \tag{79}$$

---

[4] The case when $b_{\mathsf{C}}^+ \leq b_{\mathsf{C}}^-$ can be analyzed in the same way, and the result still follows as in the case of Theorems 3-4. Thus, we omit it to avoid repetition.

Then, the transition of the log-likelihood ratio process follows from its definition using the recursive expression of b_t. As discussed in the proofs of Theorem 3-4, the transition rule of the log-likelihood ratio process when b_{t-1} is in $(b_{\mathsf{C}}^+, 1)$ and $(0, b_{\mathsf{C}}^-]$ determines the limit of L_t. Thus, in the following steps we analyze the mean of the random jumps that govern L_t when b_{t-1} lies in $(b_{\mathsf{C}}^+, 1)$ and $(0, b_{\mathsf{C}}^-]$.

Step 2. Assume b_t lies in $(b_{\mathsf{C}}^+, 1)$. In this case, the log-likelihood ratio L_t process is governed by the random jumps R_t^{HM} whose mean is given by

$$\mathbb{E}^{\Gamma}[R_t^{\mathsf{HM}}] = p\alpha^{\mathsf{H}}\alpha^{\Gamma}\log\left(\frac{\alpha^{\mathsf{C+}}}{\alpha^{\mathsf{C-}}}\right) + \bar{p}\bar{\beta}^{\mathsf{H}}\bar{\beta}^{\Gamma}\log\left(\frac{\bar{\beta}^{\mathsf{C+}}}{\bar{\beta}^{\mathsf{C-}}}\right)$$

for $\Gamma \in \{\mathsf{C+}, \mathsf{C-}\}$. Since $\alpha^{\mathsf{C+}} > \alpha^{\mathsf{C-}}$ and $\beta^{\mathsf{C+}} < \beta^{\mathsf{C-}}$ (which implies $\bar{\beta}^{\mathsf{C+}} > \bar{\beta}^{\mathsf{C-}}$) as shown in Lemma 7, it follows that $\mathbb{E}^{\Gamma}[R_t^{\mathsf{HM}}] > 0$.

Assume now b_{t-1} lies in $(0, b_{\mathsf{C}}^-]$. In this case, we have

$$\begin{aligned}\mathbb{E}^{\Gamma}[R_t^{\mathsf{HM}}] = & \underbrace{p\alpha^{\mathsf{H}}\left[\alpha^{\Gamma}\log\left(\frac{\alpha^{\mathsf{C+}}}{\alpha^{\mathsf{C-}}}\right) + \bar{\alpha}^{\Gamma}\log\left(\frac{\bar{\alpha}^{\mathsf{C+}}}{\bar{\alpha}^{\mathsf{C-}}}\right)\right]}_{I} + \underbrace{\bar{p}\bar{\beta}^{\mathsf{H}}\left[\bar{\beta}^{\Gamma}\log\left(\frac{\bar{\beta}^{\mathsf{C+}}}{\bar{\beta}^{\mathsf{C-}}}\right) + \beta^{\Gamma}\log\left(\frac{\beta^{\mathsf{C+}}}{\beta^{\mathsf{C-}}}\right)\right]}_{II} \\ & \underbrace{+ p\bar{\alpha}^{\mathsf{H}}\alpha^{\Gamma}\log\left(\frac{\alpha^{\mathsf{C+}}}{\alpha^{\mathsf{C-}}}\right) + \bar{p}\beta^{\mathsf{H}}\bar{\beta}^{\Gamma}\log\left(\frac{\bar{\beta}^{\mathsf{C+}}}{\bar{\beta}^{\mathsf{C-}}}\right)}_{III}\end{aligned}$$

Note that the sign of terms I and II are determined by Γ because the terms inside the square brackets are KL-divergence (see proof of Theorem 1 for more). In particular, terms I and II are positive when $\Gamma = \mathsf{C+}$ and negative when $\Gamma = \mathsf{C-}$. Differently, term III is always positive regardless of Γ and p because $\alpha^{\mathsf{C+}} > \alpha^{\mathsf{C-}}$ and $\beta^{\mathsf{C+}} < \beta^{\mathsf{C-}}$ (which implies $\bar{\beta}^{\mathsf{C+}} > \bar{\beta}^{\mathsf{C-}}$) as shown in Lemma 7. Therefore, when $\Gamma = \mathsf{C+}$, it follows that $\mathbb{E}^{\Gamma}[R_t^{\mathsf{HM}}] > 0$. However, when $\Gamma = \mathsf{C-}$, the value of p determines the sign of $\mathbb{E}^{\Gamma}[R_t^{\mathsf{HM}}]$. Observe that $\mathbb{E}^{\mathsf{C-}}[R_t^{\mathsf{HM}}]$ is linear function of p, and it is positive when $p = 0$ since

$$\bar{\beta}^{\mathsf{C-}}\log\left(\frac{\bar{\beta}^{\mathsf{C+}}}{\bar{\beta}^{\mathsf{C-}}}\right) \geq \beta^{\mathsf{C-}}\log\left(\frac{\beta^{\mathsf{C-}}}{\beta^{\mathsf{C+}}}\right) > \bar{\beta}^{\mathsf{H}}\beta^{\mathsf{C-}}\log\left(\frac{\beta^{\mathsf{C-}}}{\beta^{\mathsf{C+}}}\right)$$

Here, the first inequality follows from the fact that KL-divergence is positive, and the second follows from $1 > \bar{\beta}^{\mathsf{H}}$. Hence, we define p^{C} as follows.

$$p^{\mathsf{C}} \triangleq \begin{cases} \frac{term_b}{term_b - term_a} & \text{if } term_b > term_a \\ 1 & \text{otherwise.} \end{cases} \tag{80}$$

where

$$term_a = \alpha^{\mathsf{C-}}\log\left(\frac{\alpha^{\mathsf{C+}}}{\alpha^{\mathsf{C-}}}\right) + \alpha^{\mathsf{H}}\bar{\alpha}^{\mathsf{C-}}\log\left(\frac{\bar{\alpha}^{\mathsf{C+}}}{\bar{\alpha}^{\mathsf{C-}}}\right)$$

$$term_b = \bar{\beta}^{\mathsf{C-}}\log\left(\frac{\bar{\beta}^{\mathsf{C+}}}{\bar{\beta}^{\mathsf{C-}}}\right) + \bar{\beta}^{\mathsf{H}}\beta^{\mathsf{C-}}\log\left(\frac{\beta^{\mathsf{C+}}}{\beta^{\mathsf{C-}}}\right)$$

Thus, it follows that $\mathbb{E}^{\mathsf{C-}}[R_t^{\mathsf{HM}}] \geq 0$ if $p \leq p^{\mathsf{C}}$; and $\mathbb{E}^{\mathsf{C-}}[R_t^{\mathsf{HM}}] < 0$ if $p > p^{\mathsf{C}}$.

Step 3. We now combine our findings in the previous steps to analyze the limiting behavior of the log-likelihood ratio process. First, we consider $\Gamma = \mathsf{C}^{+}$. In this case, the sign of the mean $\mathbb{E}^{\mathsf{C}^{+}}[R_t^{\mathsf{HM}}]$ is always positive for b_{t-1} in $(b_{\mathsf{C}}^{+}, 1)$ and $(0, b_{\mathsf{C}}^{-}]$. Thus, L_t converges to infinity and b_t converges to 1. Next, consider $\Gamma = \mathsf{C}^{-}$. In this case, for $p \le p^{\mathsf{C}}$, we have the sign of the mean $\mathbb{E}^{\mathsf{C}^{-}}[R_t^{\mathsf{HM}}]$ is positive for b_{t-1} in $(b_{\mathsf{C}}^{+}, 1)$ and nonnegative for b_{t-1} in $(0, b_{\mathsf{C}}^{-}]$. Thus, b_t again converges to 1 because L_t goes to infinity. If $p > p^{\mathsf{C}}$, then $\mathbb{E}^{\mathsf{C}^{-}}[R_t^{\mathsf{HM}}] < 0$ for b_{t-1} in $(0, b_{\mathsf{C}}^{-}]$, while $\mathbb{E}^{\mathsf{C}^{-}}[R_t^{\mathsf{HM}}] > 0$ for b_{t-1} in $(b_{\mathsf{C}}^{+}, 1)$. Therefore, b_t converges to a Bernoulli random variable.[5] Q.E.D.

Appendix E: Relaxation of the Verification Bias

In this section of the appendix, we relax the verification bias by changing the belief updating process in two different ways. Specifically, we first consider a setup where the DM updates her belief about the machine's type based on the signal generated by the machine when the correctness of the machine's prediction is not revealed. Second, we allow the DM to observe the correctness of the machine's prediction regardless of the DM's decision to act.

First, we consider that the DM updates her belief when she decides not to act. In this case, the DM accounts for the machine prescription and her own judgment to update her belief. The modified belief updating rule that replaces equation (6) is given by

$$b_t = \begin{cases} \left[1 + \dfrac{\bar{b}_{t-1}}{b_{t-1}} \dfrac{\mathbb{P}^{\mathsf{W}}(S_t^{\mathsf{M}} = s^{\mathsf{M}}, S_t^{\mathsf{H}} = s^{\mathsf{H}})}{\mathbb{P}^{\mathsf{B}}(S_t^{\mathsf{M}} = s^{\mathsf{M}}, S_t^{\mathsf{H}} = s^{\mathsf{H}})}\right]^{-1} & \text{if } \mathbb{P}(\Theta_t = \mathsf{A} \mid S_t^{\mathsf{H}} = s^{\mathsf{H}},\, S_t^{\mathsf{M}} = s^{\mathsf{M}},\, b_{t-1}) < r \\ \left[1 + \dfrac{\bar{b}_{t-1}}{b_{t-1}} \dfrac{\mathbb{P}^{\mathsf{W}}(S_t^{\mathsf{M}} = s^{\mathsf{M}} \mid \Theta_t = \theta)}{\mathbb{P}^{\mathsf{B}}(S_t^{\mathsf{M}} = s^{\mathsf{M}} \mid \Theta_t = \theta)}\right]^{-1} & \text{if } \mathbb{P}(\Theta_t = \mathsf{A} \mid S_t^{\mathsf{H}} = s^{\mathsf{H}},\, S_t^{\mathsf{M}} = s^{\mathsf{M}},\, b_{t-1}) \ge r. \end{cases} \tag{81}$$

Thus, we directly focus on the proof when the DM updates her belief

Proposition 1. *If the DM updates her belief using* (81)*, then* $b_t \xrightarrow{a.s.} 1_{\{\Gamma = B\}}$.

Proof of Proposition 1. We prove this result by analyzing the no-interaction and no-overriding benchmarks in the first two steps. Finally, we combine our findings to conclude the proof for the main set-up.

No-Interaction Benchmark. The log-likelihood ratio process L_t becomes a random walk such that the mean of the i.i.d. random jumps is given by

$$\begin{aligned} \mathbb{E}^{\Gamma}[R_t] = {} & (p\bar{\alpha}^{\mathsf{H}}\bar{\alpha}^{\Gamma} + \bar{p}\beta^{\mathsf{H}}\beta^{\Gamma}) \log\left(\frac{p\bar{\alpha}^{\mathsf{H}}\bar{\alpha}^{\mathsf{B}} + \bar{p}\beta^{\mathsf{H}}\beta^{\mathsf{B}}}{p\bar{\alpha}^{\mathsf{H}}\bar{\alpha}^{\mathsf{W}} + \bar{p}\beta^{\mathsf{H}}\beta^{\mathsf{W}}}\right) + (p\bar{\alpha}^{\mathsf{H}}\alpha^{\Gamma} + \bar{p}\beta^{\mathsf{H}}\bar{\beta}^{\Gamma}) \log\left(\frac{p\bar{\alpha}^{\mathsf{H}}\alpha^{\mathsf{B}} + \bar{p}\beta^{\mathsf{H}}\bar{\beta}^{\mathsf{B}}}{p\bar{\alpha}^{\mathsf{H}}\alpha^{\mathsf{W}} + \bar{p}\beta^{\mathsf{H}}\bar{\beta}^{\mathsf{W}}}\right) \\ & + p\left[\alpha^{\mathsf{H}}\left(\bar{\alpha}^{\Gamma}\log\left(\frac{\bar{\alpha}^{\mathsf{B}}}{\bar{\alpha}^{\mathsf{W}}}\right) + \alpha^{\Gamma}\log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right)\right)\right] + \bar{p}\left[\bar{\beta}^{\mathsf{H}}\left(\bar{\beta}^{\Gamma}\log\left(\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right) + \beta^{\Gamma}\log\left(\frac{\beta^{\mathsf{B}}}{\beta^{\mathsf{B}}}\right)\right)\right] \end{aligned}$$

---

[5] See proofs of Theorems 3-4 for more about the relation between the sign of the mean of the random jumps and the convergence behavior.

We next show that $\mathbb{E}^{\mathsf{B}}[R_t] > 0$ and $\mathbb{E}^{\mathsf{W}}[R_t] < 0$. The terms inside the square brackets above are in fact the same as in (4). In the proof of Theorem 1, we prove that the sign of those terms are aligned with the true type of the machine. Thus, we focus on the following term.

$$(p\bar{\alpha}^{\mathsf{H}}\bar{\alpha}^{\Gamma}+\bar{p}\beta^{\mathsf{H}}\beta^{\Gamma})\log\left(\frac{p\bar{\alpha}^{\mathsf{H}}\bar{\alpha}^{\mathsf{B}}+\bar{p}\beta^{\mathsf{H}}\beta^{\mathsf{B}}}{p\bar{\alpha}^{\mathsf{H}}\bar{\alpha}^{\mathsf{W}}+\bar{p}\beta^{\mathsf{H}}\beta^{\mathsf{W}}}\right)+(p\bar{\alpha}^{\mathsf{H}}\alpha^{\Gamma}+\bar{p}\beta^{\mathsf{H}}\bar{\beta}^{\Gamma})\log\left(\frac{p\bar{\alpha}^{\mathsf{H}}\alpha^{\mathsf{B}}+\bar{p}\beta^{\mathsf{H}}\bar{\beta}^{\mathsf{B}}}{p\bar{\alpha}^{\mathsf{H}}\alpha^{\mathsf{W}}+\bar{p}\beta^{\mathsf{H}}\bar{\beta}^{\mathsf{W}}}\right)$$

Assume first $\Gamma = \mathsf{B}$. Then, we show that the above term is negative in the following.

$$\begin{aligned}
&-(p\bar{\alpha}^{\mathsf{H}}\bar{\alpha}^{\mathsf{B}}+\bar{p}\beta^{\mathsf{H}}\beta^{\mathsf{B}})\log\left(\frac{p\bar{\alpha}^{\mathsf{H}}\bar{\alpha}^{\mathsf{B}}+\bar{p}\beta^{\mathsf{H}}\beta^{\mathsf{B}}}{p\bar{\alpha}^{\mathsf{H}}\bar{\alpha}^{\mathsf{W}}+\bar{p}\beta^{\mathsf{H}}\beta^{\mathsf{W}}}\right)-(p\bar{\alpha}^{\mathsf{H}}\alpha^{\mathsf{B}}+\bar{p}\beta^{\mathsf{H}}\bar{\beta}^{\mathsf{B}})\log\left(\frac{p\bar{\alpha}^{\mathsf{H}}\alpha^{\mathsf{B}}+\bar{p}\beta^{\mathsf{H}}\bar{\beta}^{\mathsf{B}}}{p\bar{\alpha}^{\mathsf{H}}\alpha^{\mathsf{W}}+\bar{p}\beta^{\mathsf{H}}\bar{\beta}^{\mathsf{W}}}\right)\\
=&(p\bar{\alpha}^{\mathsf{H}}\bar{\alpha}^{\mathsf{B}}+\bar{p}\beta^{\mathsf{H}}\beta^{\mathsf{B}})\log\left(\frac{p\bar{\alpha}^{\mathsf{H}}\bar{\alpha}^{\mathsf{W}}+\bar{p}\beta^{\mathsf{H}}\beta^{\mathsf{W}}}{p\bar{\alpha}^{\mathsf{H}}\bar{\alpha}^{\mathsf{B}}+\bar{p}\beta^{\mathsf{H}}\beta^{\mathsf{B}}}\right)+(p\bar{\alpha}^{\mathsf{H}}\alpha^{\mathsf{B}}+\bar{p}\beta^{\mathsf{H}}\bar{\beta}^{\mathsf{B}})\log\left(\frac{p\bar{\alpha}^{\mathsf{H}}\alpha^{\mathsf{W}}+\bar{p}\beta^{\mathsf{H}}\bar{\beta}^{\mathsf{W}}}{p\bar{\alpha}^{\mathsf{H}}\alpha^{\mathsf{B}}+\bar{p}\beta^{\mathsf{H}}\bar{\beta}^{\mathsf{B}}}\right)\\
\leq&\log\left[p\bar{\alpha}^{\mathsf{H}}\bar{\alpha}^{\mathsf{W}}+\bar{p}\beta^{\mathsf{H}}\beta^{\mathsf{W}}+p\bar{\alpha}^{\mathsf{H}}\alpha^{\mathsf{W}}+\bar{p}\beta^{\mathsf{H}}\bar{\beta}^{\mathsf{W}}\right]=\log\left[p\bar{\alpha}^{\mathsf{H}}+\bar{p}\beta^{\mathsf{H}}\right]<\log(1)=0
\end{aligned}$$

Here, the first inequality follows because $\log(\cdot)$ is a concave function. The second inequality follows because $p\bar{\alpha}^{\mathsf{H}}+\bar{p}\beta^{\mathsf{H}}$ is less than one. Thus, we obtain that $\mathbb{E}^{\mathsf{B}}[R_t] > 0$, and following the same steps imply that $\mathbb{E}^{\mathsf{W}}[R_t] < 0$.

No-Overriding Benchmark. In this case, the mean of the random jumps that govern the log-likelihood ratio process is given by

$$\begin{aligned}
\mathbb{E}^{\Gamma}[R_t]=&(p\bar{\alpha}^{\Gamma}\alpha^{\mathsf{H}}+\bar{p}\beta^{\Gamma}\bar{\beta}^{\mathsf{H}})\log\left(\frac{p\bar{\alpha}^{\mathsf{B}}\alpha^{\mathsf{H}}+\bar{p}\beta^{\mathsf{B}}\bar{\beta}^{\mathsf{H}}}{p\bar{\alpha}^{\mathsf{W}}\alpha^{\mathsf{H}}+\bar{p}\beta^{\mathsf{W}}\bar{\beta}^{\mathsf{H}}}\right)+(p\bar{\alpha}^{\Gamma}\bar{\alpha}^{\mathsf{H}}+\bar{p}\beta^{\Gamma}\beta^{\mathsf{H}})\log\left(\frac{p\bar{\alpha}^{\mathsf{B}}\bar{\alpha}^{\mathsf{H}}+\bar{p}\beta^{\mathsf{B}}\beta^{\mathsf{H}}}{p\bar{\alpha}^{\mathsf{W}}\bar{\alpha}^{\mathsf{H}}+\bar{p}\beta^{\mathsf{W}}\beta^{\mathsf{H}}}\right)\\
&+p\alpha^{\Gamma}\log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right)+\bar{p}\bar{\beta}^{\Gamma}\log\left(\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right).
\end{aligned}$$

Both of these terms are KL-divergence when $\Gamma = \mathsf{B}$ and $-$KL-divergence when $\Gamma = \mathsf{W}$. Thus it is respectively positive and negative which implies that the log-likelihood ratio process converges to infinity and respectively minus infinity. Hence, we obtain $b_t \xrightarrow{\text{a.s.}} 1_{\Gamma=\mathsf{B}\}}$.

Main Set-up. Note that the decision rule characterized in Lemma 1 implies that the DM follows no-interaction and no-overriding decision rules at low and respectively high b_t. As shown in the previous steps, the log-likelihood ratio process under those decision rules converge to ∞ for $\Gamma = \mathsf{B}$, and $-\infty$ for $\Gamma = \mathsf{W}$. Therefore, the DM correctly learns the machine's type. Q.E.D.

Finally, we consider that the DM observes the pair $(S_t^{\mathsf{M}}, \Theta_t)$ in all periods. In this case, the updating rule is given by

$$b_t = \left[1+\frac{\bar{b}_{t-1}}{b_{t-1}}\frac{\mathbb{P}^{\mathsf{W}}(S_t^{\mathsf{M}}\mid\Theta_t=\theta)}{\mathbb{P}^{\mathsf{B}}(S_t^{\mathsf{M}}\mid\Theta_t=\theta)}\right]^{-1} \tag{82}$$

We drop the decision rule in (82) because it does not affect the belief updating there.

In fact, when the DM can observe the pair $(S_t^{\mathsf{M}}, \Theta_t)$ in all periods, the sum of probability of observing all potential outcomes equals one. For those cases, the frequentist consistency of Bayesian updating (Diaconis and Freedman 1986) thus implies that the DM's belief converges almost surely to one when the machine's type is B (and respectively to zero when the machine's type is W).

E.1. Partial Relaxation of the Verification Bias

Note that the modified belief updating rule in Equation (81) is a fully Bayesian belief updating rule in the sense that the DM uses Bayes' rule to update her belief by accounting for all available information. Indeed, the frequentist consistency of Bayesian updating (Diaconis and Freedman 1986) is not violated by the censorship (not observing θ_t in case of no action) because the total probabilities of all potential observations, which are $(S_t^{\mathsf{M}}, S_t^{\mathsf{H}})$ and $(S_t^{\mathsf{M}}, \theta_t)$ pairs add up to one. Hence, the belief updating rules in (6) and (81) represent two extremes such that the DM fully ignores and respectively fully accounts for the information when the correctness of the machine's prediction is not verified. In this section, we analyze the cases in between these two extremes by using the following belief update rule for $0 < \varepsilon < 1$.

For any value of $\varepsilon < 1$, the DM overlooks the information generated by unverified cases. Indeed, Equation (15) is reduced to (6) (to (81)) for $\varepsilon = 0$ (and resp., $\varepsilon = 1$). Hence, Proposition 1 and Theorems 3-4 can be thought of special cases of the following theorem.

Proof of Theorem 8. We prove this result in a way akin to the proof of Proposition 1. Thus, we first start with the benchmark cases even though the statement of the theorem is for the main set-up. Finally, we combine the results for benchmark cases to derive the results in the statement of the theorem.

No-Interaction Benchmark. The log-likelihood ratio process L_t becomes a random walk such that the mean of the i.i.d. random jumps is given by

$$\begin{aligned}\mathbb{E}^{\Gamma}[R_t] =& \varepsilon(p\bar{\alpha}^{\mathsf{H}}\bar{\alpha}^{\Gamma} + \bar{p}\beta^{\mathsf{H}}\beta^{\Gamma})\log\left(\frac{p\bar{\alpha}^{\mathsf{H}}\bar{\alpha}^{\mathsf{B}} + \bar{p}\beta^{\mathsf{H}}\beta^{\mathsf{B}}}{p\bar{\alpha}^{\mathsf{H}}\bar{\alpha}^{\mathsf{W}} + \bar{p}\beta^{\mathsf{H}}\beta^{\mathsf{W}}}\right) + \varepsilon(p\bar{\alpha}^{\mathsf{H}}\alpha^{\Gamma} + \bar{p}\beta^{\mathsf{H}}\bar{\beta}^{\Gamma})\log\left(\frac{p\bar{\alpha}^{\mathsf{H}}\alpha^{\mathsf{B}} + \bar{p}\beta^{\mathsf{H}}\bar{\beta}^{\mathsf{B}}}{p\bar{\alpha}^{\mathsf{H}}\alpha^{\mathsf{W}} + \bar{p}\beta^{\mathsf{H}}\bar{\beta}^{\mathsf{W}}}\right)\\ &+ p\left[\alpha^{\mathsf{H}}\left(\bar{\alpha}^{\Gamma}\log\left(\frac{\bar{\alpha}^{\mathsf{B}}}{\bar{\alpha}^{\mathsf{W}}}\right) + \alpha^{\Gamma}\log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right)\right)\right] + \bar{p}\left[\bar{\beta}^{\mathsf{H}}\left(\bar{\beta}^{\Gamma}\log\left(\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right) + \beta^{\Gamma}\log\left(\frac{\beta^{\mathsf{B}}}{\beta^{\mathsf{B}}}\right)\right)\right]\end{aligned}$$

We next show that $\mathbb{E}^{\mathsf{B}}[R_t] > 0$ and $\mathbb{E}^{\mathsf{W}}[R_t] < 0$. The terms inside the square brackets above are, in fact, the same as in (4). In the proof of Theorem 1, we prove that the sign of those terms are aligned with the true type of the machine. Thus, we focus on the following term.

$$\varepsilon\left\{(p\bar{\alpha}^{\mathsf{H}}\bar{\alpha}^{\Gamma} + \bar{p}\beta^{\mathsf{H}}\beta^{\Gamma})\log\left(\frac{p\bar{\alpha}^{\mathsf{H}}\bar{\alpha}^{\mathsf{B}} + \bar{p}\beta^{\mathsf{H}}\beta^{\mathsf{B}}}{p\bar{\alpha}^{\mathsf{H}}\bar{\alpha}^{\mathsf{W}} + \bar{p}\beta^{\mathsf{H}}\beta^{\mathsf{W}}}\right) + (p\bar{\alpha}^{\mathsf{H}}\alpha^{\Gamma} + \bar{p}\beta^{\mathsf{H}}\bar{\beta}^{\Gamma})\log\left(\frac{p\bar{\alpha}^{\mathsf{H}}\alpha^{\mathsf{B}} + \bar{p}\beta^{\mathsf{H}}\bar{\beta}^{\mathsf{B}}}{p\bar{\alpha}^{\mathsf{H}}\alpha^{\mathsf{W}} + \bar{p}\beta^{\mathsf{H}}\bar{\beta}^{\mathsf{W}}}\right)\right\}$$

In the proof of Proposition 1, we showed that the sign of the term inside the curly brackets above is aligned with the machine's type. Since $\varepsilon > 0$, it does not change it. Hence, the DM's belief converges to $1_{\{\Gamma=\mathsf{B}\}}$ in the no-interaction benchmark when the belief updating rule is as in (15).

No-Overriding Benchmark. In this case, the mean of the random jumps that govern the log-likelihood ratio process is given by

$$\begin{aligned}\mathbb{E}^{\Gamma}[R_t] =& \varepsilon(p\bar{\alpha}^{\Gamma}\alpha^{\mathsf{H}} + \bar{p}\beta^{\Gamma}\bar{\beta}^{\mathsf{H}})\log\left(\frac{p\bar{\alpha}^{\mathsf{B}}\alpha^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\bar{\beta}^{\mathsf{H}}}{p\bar{\alpha}^{\mathsf{W}}\alpha^{\mathsf{H}} + \bar{p}\beta^{\mathsf{W}}\bar{\beta}^{\mathsf{H}}}\right) + \varepsilon(p\bar{\alpha}^{\Gamma}\bar{\alpha}^{\mathsf{H}} + \bar{p}\beta^{\Gamma}\beta^{\mathsf{H}})\log\left(\frac{p\bar{\alpha}^{\mathsf{B}}\bar{\alpha}^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\beta^{\mathsf{H}}}{p\bar{\alpha}^{\mathsf{W}}\bar{\alpha}^{\mathsf{H}} + \bar{p}\beta^{\mathsf{W}}\beta^{\mathsf{H}}}\right)\\ &+ p\alpha^{\Gamma}\log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right) + \bar{p}\bar{\beta}^{\Gamma}\log\left(\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right).\end{aligned}$$

We first consider $\Gamma = \mathsf{B}$. If $p \geq p^{\mathsf{B}}$, then we have the following inequality.

$$p\alpha^{\mathsf{B}} \log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right) + \bar{p}\bar{\beta}^{\mathsf{B}} \log\left(\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right) \geq 0$$

Using this, we obtain the following positive lower bound on $\mathbb{E}^{\mathsf{B}}[R_t]$.

$$\begin{aligned}\mathbb{E}^{\mathsf{B}}[R_t] \geq & \varepsilon(p\bar{\alpha}^{\mathsf{B}}\alpha^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\bar{\beta}^{\mathsf{H}}) \log\left(\frac{p\bar{\alpha}^{\mathsf{B}}\alpha^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\bar{\beta}^{\mathsf{H}}}{p\bar{\alpha}^{\mathsf{W}}\alpha^{\mathsf{H}} + \bar{p}\beta^{\mathsf{W}}\bar{\beta}^{\mathsf{H}}}\right) + \varepsilon(p\bar{\alpha}^{\mathsf{B}}\bar{\alpha}^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\beta^{\mathsf{H}}) \log\left(\frac{p\bar{\alpha}^{\mathsf{B}}\bar{\alpha}^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\beta^{\mathsf{H}}}{p\bar{\alpha}^{\mathsf{W}}\bar{\alpha}^{\mathsf{H}} + \bar{p}\beta^{\mathsf{W}}\beta^{\mathsf{H}}}\right) \\ & + \varepsilon\left(p\alpha^{\mathsf{B}} \log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right) + \bar{p}\bar{\beta}^{\mathsf{B}} \log\left(\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right)\right) > 0\,.\end{aligned}$$

This lower bound is positive because we obtain the KL-divergence as discussed in the proof of Proposition 1 when we use ε as the common factor.

If $p < p^{\mathsf{B}}$, then it follows that

$$p\alpha^{\mathsf{B}} \log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right) + \bar{p}\bar{\beta}^{\mathsf{B}} \log\left(\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right) < 0\,,$$

$$(p\bar{\alpha}^{\mathsf{B}}\alpha^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\bar{\beta}^{\mathsf{H}}) \log\left(\frac{p\bar{\alpha}^{\mathsf{B}}\alpha^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\bar{\beta}^{\mathsf{H}}}{p\bar{\alpha}^{\mathsf{W}}\alpha^{\mathsf{H}} + \bar{p}\beta^{\mathsf{W}}\bar{\beta}^{\mathsf{H}}}\right) + (p\bar{\alpha}^{\mathsf{B}}\bar{\alpha}^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\beta^{\mathsf{H}}) \log\left(\frac{p\bar{\alpha}^{\mathsf{B}}\bar{\alpha}^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\beta^{\mathsf{H}}}{p\bar{\alpha}^{\mathsf{W}}\bar{\alpha}^{\mathsf{H}} + \bar{p}\beta^{\mathsf{W}}\beta^{\mathsf{H}}}\right) > 0\,,$$

$$\begin{aligned}&(p\bar{\alpha}^{\mathsf{B}}\alpha^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\bar{\beta}^{\mathsf{H}}) \log\left(\frac{p\bar{\alpha}^{\mathsf{B}}\alpha^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\bar{\beta}^{\mathsf{H}}}{p\bar{\alpha}^{\mathsf{W}}\alpha^{\mathsf{H}} + \bar{p}\beta^{\mathsf{W}}\bar{\beta}^{\mathsf{H}}}\right) + (p\bar{\alpha}^{\mathsf{B}}\bar{\alpha}^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\beta^{\mathsf{H}}) \log\left(\frac{p\bar{\alpha}^{\mathsf{B}}\bar{\alpha}^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\beta^{\mathsf{H}}}{p\bar{\alpha}^{\mathsf{W}}\bar{\alpha}^{\mathsf{H}} + \bar{p}\beta^{\mathsf{W}}\beta^{\mathsf{H}}}\right) \\ & > p\alpha^{\mathsf{B}} \log\left(\frac{\alpha^{\mathsf{W}}}{\alpha^{\mathsf{B}}}\right) + \bar{p}\bar{\beta}^{\mathsf{B}} \log\left(\frac{\bar{\beta}^{\mathsf{W}}}{\bar{\beta}^{\mathsf{B}}}\right)\end{aligned}$$

Here, the first inequality is implied by the definition of p^{B}. The second and third inequalities follow from the first inequality and the fact that the KL-divergence for the better machine is positive.

Thus, we need to analyze how ε balances these two terms with opposing signs. In particular, a sufficiently high ε ensures that the positive term dominates the negative and hence $\mathbb{E}^{\mathsf{B}}[R_t] > 0$ which implies $b_t \xrightarrow{\text{a.s.}} 1$. Similarly, a sufficiently low ε implies the opposite. These thresholds over ε depend on p. Consider the following break-even point of ε such that $\mathbb{E}^{\mathsf{B}}[R_t] = 0$.

$$\varepsilon^{\mathsf{B}} \triangleq \frac{p\alpha^{\mathsf{B}} \log\left(\frac{\alpha^{\mathsf{W}}}{\alpha^{\mathsf{B}}}\right) + \bar{p}\bar{\beta}^{\mathsf{B}} \log\left(\frac{\bar{\beta}^{\mathsf{W}}}{\bar{\beta}^{\mathsf{B}}}\right)}{(p\bar{\alpha}^{\mathsf{B}}\alpha^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\bar{\beta}^{\mathsf{H}}) \log\left(\frac{p\bar{\alpha}^{\mathsf{B}}\alpha^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\bar{\beta}^{\mathsf{H}}}{p\bar{\alpha}^{\mathsf{W}}\alpha^{\mathsf{H}} + \bar{p}\beta^{\mathsf{W}}\bar{\beta}^{\mathsf{H}}}\right) + (p\bar{\alpha}^{\mathsf{B}}\bar{\alpha}^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\beta^{\mathsf{H}}) \log\left(\frac{p\bar{\alpha}^{\mathsf{B}}\bar{\alpha}^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\beta^{\mathsf{H}}}{p\bar{\alpha}^{\mathsf{W}}\bar{\alpha}^{\mathsf{H}} + \bar{p}\beta^{\mathsf{W}}\beta^{\mathsf{H}}}\right)} \tag{83}$$

If $p < p^{\mathsf{B}}$ and $\varepsilon = \varepsilon^{\mathsf{B}}$, then $\mathbb{E}^{\mathsf{B}}[R_t] = 0$ and hence the log-likelihood ratio process L_t and the DM's belief b_t oscillates (as in the case of Theorem 2 for $p = p^{\mathsf{B}}$). On the other hand, $p < p^{\mathsf{B}}$ and $\varepsilon < \varepsilon^{\mathsf{B}}$ ($\varepsilon > \varepsilon^{\mathsf{B}}$) imply $b_t \xrightarrow{\text{a.s.}} 0$ (and respectively $b_t \xrightarrow{\text{a.s.}} 1$). This break-even point ε^{B} is in $(0,1)$ for $p < p^{\mathsf{B}}$.

Next, we consider $\Gamma = \mathsf{W}$. The same arguments with opposing signs and thresholds p^{W} and ε^{W} follow where

$$\varepsilon^{\mathsf{W}} \triangleq \frac{p\alpha^{\mathsf{W}} \log\left(\frac{\alpha^{\mathsf{B}}}{\alpha^{\mathsf{W}}}\right) + \bar{p}\bar{\beta}^{\mathsf{W}} \log\left(\frac{\bar{\beta}^{\mathsf{B}}}{\bar{\beta}^{\mathsf{W}}}\right)}{(p\bar{\alpha}^{\mathsf{B}}\alpha^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\bar{\beta}^{\mathsf{H}}) \log\left(\frac{p\bar{\alpha}^{\mathsf{B}}\alpha^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\bar{\beta}^{\mathsf{H}}}{p\bar{\alpha}^{\mathsf{W}}\alpha^{\mathsf{H}} + \bar{p}\beta^{\mathsf{W}}\bar{\beta}^{\mathsf{H}}}\right) + (p\bar{\alpha}^{\mathsf{B}}\bar{\alpha}^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\beta^{\mathsf{H}}) \log\left(\frac{p\bar{\alpha}^{\mathsf{B}}\bar{\alpha}^{\mathsf{H}} + \bar{p}\beta^{\mathsf{B}}\beta^{\mathsf{H}}}{p\bar{\alpha}^{\mathsf{W}}\bar{\alpha}^{\mathsf{H}} + \bar{p}\beta^{\mathsf{W}}\beta^{\mathsf{H}}}\right)}\,. \tag{84}$$

The break-even point ε^{W} is in $(0,1)$ for $p < p^{\mathsf{W}}$.

Main Set-up. Indeed, the pairs of p and ε generate the cases that are analogous to Theorems 3-4. This is because, the DM always (regardless the value of ε) correctly learns the machine's type Γ in the no-interaction benchmark. Yet, the belief might wrongly converge to zero or one in the no-overriding benchmark as discussed in the first two steps of this proof.

Note that, the DM's decision rule in the main set-up in equation (6), and that of equation (15) are the same, specifically, $\mathbb{P}(\Theta_t = \mathsf{A} \,|\, S_t^{\mathsf{H}} = s^{\mathsf{H}}, S_t^{\mathsf{M}} = s^{\mathsf{M}}, b_{t-1}) \geq r$. Thus, Lemma 1 characterizes the DM's decision rule again.

Overall, when the machine is better, the DM correctly learns the machine's type in the main set-up if p and ε values imply that the belief at the no-overriding benchmark converges to 1. Otherwise (more specifically $p < p^{\mathsf{B}}$ and $\varepsilon \leq \varepsilon^{\mathsf{B}}$), the DM's belief is recurrent and oscillates as in the proof of Theorem 3.

Similarly, when the machine is worse, the DM correctly learns the machine's type in the main set-up if p and ε values are $p \leq p^{\mathsf{W}}$ or $\varepsilon \geq \varepsilon^{\mathsf{W}}$ such that the belief at the no-overriding benchmark converges to zero or oscillates. Otherwise (more specifically $p > p^{\mathsf{W}}$ and $\varepsilon < \varepsilon^{\mathsf{W}}$), the DM's belief in the main set-up converges to a Bernoulli random variable as in the proof of Theorem 4. Q.E.D.

References

Chung, K. L. (2001), *A course in probability theory*, 3 edn, Academic Press, San Diego, CA.

Csiszar, I. (1975), 'I-divergence geometry of probability distributions and minimization problems', *The Annals of Probability* **3**(1), 146–158.

Diaconis, P. and Freedman, D. (1986), 'On the consistency of bayes estimates', *The Annals of Statistics* **14**(1), 1–26.

Gut, A. (2009), *Stopped Random Walks*, Springer, New York, NY.

Harrison, J. M., Keskin, N. B. and Zeevi, A. (2012), 'Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution', *Management Sci.* **58**(3), 570–586.

Kemperman, J. (1974), 'The oscillating random walk', *Stochastic Processes and their applications* **2**(1), 1–29.

MacKay, D. J. (2003), *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, Cambridge, UK.

Resnick, S. (2014), *A probability path*, Birkhäuser, Boston, MA.

Vatutin, V. A. and Wachtel, V. (2009), 'Local probabilities for random walks conditioned to stay positive', *Probability Theory and Related Fields* **143**(1), 177–217.

Winkler, R. L. (1972), *An introduction to Bayesian inference and decision*, Holt, Rinehart and Winston New York.