This formulation seems promising.

$$A_b(k) = \underbrace{\beta_v \hat{A}_{b,v}(k)}_{\circ BE_v} + \underbrace{\beta_0 \hat{A}_{b,0}(k)}_{\substack{\circ BE_0 \\ \text{Fixed to } \beta_0 = 1}}, \qquad P_b(k) = \frac{e^{A_b(k)}}{\sum_{b'} e^{A_{b'}(k)}} \quad (1)$$

with

$\hat{A}_{b,v}(k)$ = squashed value for the agent of behaviour b (estimated with or without filtering), in the $[-1, 1]$ range.

and

$$\hat{A}_{b,0}(k) = \Big(1 - \underbrace{\frac{\Delta t}{T_{0f}}}_{\substack{\text{forgetting} \\ \text{time constant}}}\Big) \hat{A}_{b,0}(k-1) + \underbrace{\frac{\Delta t}{T_{01}}}_{\text{"time for one observation"}} \ln p[\tilde{x}(k) | b] \quad (2)$$

The introduction of T_{01} is to account for the fact that our simulation time step might be quite different from the "rate" at which the agent samples the world.

$$\boxed{\text{Probability of a "bad" decision} \longrightarrow \text{setting } \beta_v}$$

Consider a case where one behaviour b_+ the value is $V_+ = -1$, and all others it is $V_{\neg +} = 0$. Assuming just $\circ BE_v$, not $\circ BE_0$:

$$P_{b_+} = \frac{e^{-\beta_v}}{e^{-\beta_v} + (N_{beh} - 1) e^0} = \frac{1}{1 + (N_{beh} - 1) e^{\beta_v}}$$

So if we set a probability P_+ of how likely it should be for the agent to make that really bad decision (e.g. $P_+ = 0.01$ or $P_+ = 0.001$), we get:

$$P_+ = \frac{1}{1+(N_{beh}-1)e^{\beta_V}} \Leftrightarrow 1+(N_{beh}-1)e^{\beta_V} = \frac{1}{P_+}$$

$$\Leftrightarrow (N_{beh}-1)e^{\beta_V} = \frac{1}{P_+} - 1 = \frac{1-P_+}{P_+}$$

$$\Leftrightarrow \boxed{\beta_V = \ln\left(\frac{1}{N_{beh}-1}\,\frac{1-P_+}{P_+}\right)} \qquad (3)$$

Which we can use to fix β_V sensibly (or use P_+ as a more interpretable free parameter).

OBE$_0$ evidence building without forgetting

With $T_{of} \to \infty$ we get:

$$\hat{A}_{b,0}(k) = \hat{A}_{b,0}(k-1) + \frac{\Delta t}{T_{o1}} \ln p[\tilde{x}(k)|b]$$

so if $p[\tilde{x}(k)|b]$ is constant, p_b we get:

$$\boxed{\hat{A}_{b,0}(k) = \underbrace{\hat{A}_{b,0}(0)}_{\substack{\text{Assumed}\\ =0}} + k\frac{\Delta t}{T_{o1}} \ln p_b = k\frac{\Delta t}{T_{o1}} \ln p_b} \qquad (4)$$

bBE$_0$ time to confidence threshold, without forgetting

Now assume for one behaviour b_*, $p_{b^*} = p_{max}$

and for all the others $p_{\neg b^*} = \alpha_0 p_{max} < p_{max}$.
$= p_{min}$ $\iff \alpha_0 = \frac{p_{min}}{p_{max}}$

This gives:

$$\boxed{P_{b^*}(k) = \frac{e^{k\frac{\Delta t}{T_{01}}\ln p_{max}}}{e^{k\frac{\Delta t}{T_{01}}\ln p_{max}} + (N_{beh}-1)e^{k\frac{\Delta t}{T_{01}}\underbrace{\ln(\alpha_0 p_{max})}_{=\ln\alpha_0 + \ln p_{max}}}}}$$

$$= \frac{e^{k\frac{\Delta t}{T_{01}}\ln p_{max}}}{e^{k\frac{\Delta t}{T_{01}}\ln p_{max}} + (N_{beh}-1)e^{k\frac{\Delta t}{T_{01}}\ln\alpha_0}\,e^{k\frac{\Delta t}{T_{01}}\ln p_{max}}}$$

$$= \boxed{\frac{1}{1 + (N_{beh}-1)e^{k\frac{\Delta t}{T_{01}}\ln\alpha_0}}} \qquad (5)$$

So the time to reach a given confidence threshold $P_{b^*} = P_{b,th}$ can be obtained, similarly to (3) above:

$$k\frac{\Delta t}{T_{01}}\ln\alpha_0 = \ln\left(\frac{1}{N_{beh}-1}\,\frac{1-P_{b,th}}{P_{b,th}}\right)$$

$$\Rightarrow \boxed{k\Delta t = \frac{T_{01}}{\ln\alpha_0}\ln\left(\frac{1}{N_{beh}-1}\,\frac{1-P_{b,th}}{P_{b,th}}\right)} \qquad (6)$$

eller

$$\boxed{k\Delta t = T_{01}\,\frac{\ln\left[(N_{beh}-1)\frac{P_{b,th}}{1-P_{b,th}}\right]}{\ln\frac{p_{max}}{p_{min}}}}$$

$\boxed{\text{oBEo evidence buildup, with forgetting}}$

With constant observation likelihood $p[\tilde{x}(k)|b] = p_b$

$$\hat{A}_{b,0}(k) = \left(1 - \frac{\Delta t}{T_{0f}}\right) \hat{A}_{b,0}(k-1) + \frac{\Delta t}{T_{01}} \ln p_b \qquad (7)$$

Which converges when:

$$\hat{A}_{b,0}(k) = \hat{A}_{b,0}(k-1)$$

$$\Leftrightarrow \cancel{\hat{A}_{b,0}(k-1)} = (\cancel{1} - \frac{\Delta t}{T_{0f}}) \hat{A}_{b,0}(k-1) + \frac{\Delta t}{T_{01}} \ln p_b$$

$$\Leftrightarrow 0 = - \frac{\Delta t}{T_{0f}} \hat{A}_{b,0}(k-1) + \frac{\Delta t}{T_{01}} \ln p_b$$

$$\Rightarrow \boxed{\hat{A}_{b,0}(k-1) = \frac{T_{0f}}{T_{01}} \ln p_b} \qquad (8)$$

To get an expression for $\hat{A}_{b,0}$ a function of time, rewrite (7) as a continuous-time differential eq^{n}:

$$\frac{d\hat{A}_{b,0}}{dt} = - \frac{\hat{A}_{b,0}}{T_{0f}} + \frac{\ln p_b}{T_{01}} \qquad (9)$$

This should have a solution on the form:

$$A(t) = \frac{T_{0f}}{T_{01}} \ln p_b \left(1 - e^{-wt}\right) \qquad \left(\text{exponential convergence to (8)}\right)$$

$$\Rightarrow \frac{dA}{dt} = w \frac{T_{0f}}{T_{01}} \ln p_b \, e^{-wt}$$

Insert in (9):

$$\omega \frac{T_{of}}{T_{o1}} \ln p_b e^{-\omega t} = -\frac{1}{T_{of}}\left[\frac{T_{of}}{T_{o1}} \ln p_b \left(1-e^{-\omega t}\right)\right] + \frac{\ln p_b}{T_{o1}}$$

$$= -\frac{1}{T_{o1}} \ln p_b \left(\cancel{1}-e^{-\omega t}\right) + \cancel{\frac{\ln p_b}{T_{o1}}}$$

$$= \frac{1}{T_{o1}} \ln p_b e^{-\omega t}$$

Assuming $p_b < 1$, we get:

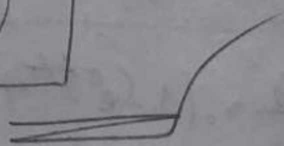
$$\omega T_{of} = 1$$

$$\Leftrightarrow \omega = \frac{1}{T_{of}}$$, i.e.

$$\boxed{\hat{A}_{b,0}(t) = \frac{T_{of}}{T_{o1}} \ln p_b \left(1-e^{-t/T_{of}}\right)} \quad (10)$$

or in discrete-time form:

$$\hat{A}_{b,0}(k) = \frac{T_{of}}{T_{o1}} \ln p_b \left(1-e^{-k\Delta t/T_{of}}\right)$$



oBEo time to confidence threshold, with forgetting

With the same assumptions as those that led to (6) above, but for a non-infinite T_{of}, we can now use (10) to get:

$$\boxed{P_{b^*}(t) = \frac{e^{\frac{T_{of}}{T_{o1}} \ln p_{max}(1-e^{-t/T_{of}})}}{e^{\frac{T_{of}}{T_{o1}} \ln p_{max}(1-e^{-t/T_{of}})} + (N_{bch}-1)\, e^{\frac{T_{of}}{T_{o1}} \underbrace{\ln(\alpha_0 p_{max})}_{=(\ln \alpha_0 + \ln p_{max})} (1-e^{-t/T_{of}})}}}$$

$$= \left\{ \text{similarly to the derivation of (5)} \right\}$$

$$= \boxed{\frac{1}{1 + (N_{bch}-1)\, e^{\frac{T_{of}}{T_{o1}} \ln \alpha_0 (1-e^{-t/T_{of}})}}} \qquad (11)$$

So now, the time to reach a confidence threshold $P_{b,th}$ can be obtained as:

$$P_{b,th} = \frac{1}{1 + (N_{bch}-1)\, e^{\frac{T_{of}}{T_{o1}} \ln \alpha_0 (1-e^{-t/T_{of}})}}$$

$$\Leftrightarrow \left\{ \text{similarly to (3) and (6) above} \right\}$$

$$\Leftrightarrow \frac{T_{of}}{T_{o1}} \ln \alpha_0 \left(1-e^{-t/T_{of}}\right) = \ln\left(\frac{1}{N_{bch}-1} \frac{1-P_{b,th}}{P_{b,th}}\right)$$

$$\Leftrightarrow 1-e^{-t/T_{of}} = \frac{T_{o1} \ln\left(\frac{1}{N_{bch}-1} \frac{1-P_{b,th}}{P_{b,th}}\right)}{T_{of} \ln \alpha_0}$$

$$\Leftrightarrow e^{-t/T_{of}} = 1 - \frac{T_{o1} \ln\left(\frac{1}{N_{bch}-1} \frac{1-P_{b,th}}{P_{b,th}}\right)}{T_{of} \ln \alpha_0}$$

$$\Leftrightarrow \left[t = -T_{of} \ln\left(1 - \frac{T_{o1} \ln\left(\frac{1}{N_{bch}-1} \frac{1-P_{b,th}}{P_{b,th}}\right)}{T_{of} \ln \alpha_0}\right)\right.$$

$$= -T_{of} \ln\left(1 - \frac{-T_{o1} \ln\left[(N_{bch}-1)\frac{P_{b,th}}{1-P_{b,th}}\right]}{-T_{of} \ln \frac{p_{max}}{p_{min}}}\right)$$

$$\boxed{= T_{of} \ln\left(\frac{1}{1 - \frac{T_{o1} \ln\left[(N_{bch}-1)\frac{P_{b,th}}{1-P_{b,th}}\right]}{T_{of} \ln \frac{p_{max}}{p_{min}}}}\right)} \quad (12)$$

Note that (11) shows that $\alpha_0 = \frac{p_{min}}{p_{max}}$ limits how high $P_{b,t}$ can maximally get, and this is reflected also in (12), which can go to infinity if $\frac{p_{max}}{p_{min}}$ is too small for the $P_{b,th}$, N_{bch}, T_{o1}, and T_{of} at hand.

oBEo + oBEv evidence buildup, without forgetting

Similarly to the derivation of (4), if we assume $p(\tilde{x}(k)|b] = p_b$ and the estimated value of b also constant, $\hat{A}_{b,v}(k) = V_b$, we get

$$\boxed{A_b(k) = \beta_V V_b + \beta_0 k \frac{\Delta t}{T_{o1}} \ln p_b} \quad (13)$$

(Note that I am including β_0 for completeness, even though I think we should fix $\beta_0 = 1$)

oBEo + oBEv, observing a "bad" decision, without forgetting

Now combine the assumptions which gave (3) with those which gave (5), with $b_* = b_+$:

$$P_{b_+}(k) = \frac{e^{-\beta_v + \beta_0 k \frac{\Delta t}{T_{01}} \ln p_{max}}}{e^{-\beta_v + \beta_0 k \frac{\Delta t}{T_{01}} \ln p_{max}} + (N_{beh}-1) e^{\beta_0 k \frac{\Delta t}{T_{01}} \underbrace{\ln(\alpha_0 p_{max})}_{= (\ln \alpha_0 + \ln p_{max})}}}$$

$= \{$similarly to previous derivations$\}$

$$= \frac{e^{-\beta_v}}{e^{-\beta_v} + (N_{beh}-1) e^{\beta_0 k \frac{\Delta t}{T_{01}} \ln \alpha_0}}$$

$$= \boxed{\frac{1}{1 + (N_{beh}-1) e^{\beta_v - \beta_0 k \frac{\Delta t}{T_{01}} \ln \frac{p_{max}}{p_{min}}}}} \qquad (14)$$

This reaches threshold as (similarly to above):

$$\beta_v - \beta_0 k \frac{\Delta t}{T_{01}} \ln \frac{p_{max}}{p_{min}} = \ln \left(\frac{1}{N_{beh}-1} \frac{1-P_{b,th}}{P_{b,th}} \right)$$

$$\Leftrightarrow \beta_0 k \frac{\Delta t}{T_{01}} \ln \frac{p_{max}}{p_{min}} = \beta_v + \ln \left[(N_{beh}-1) \frac{P_{b,th}}{1-P_{b,th}} \right]$$

$$\Leftrightarrow k\Delta t = T_{01} \frac{\beta_v + \ln \left[(N_{beh}-1) \frac{P_{b,th}}{1-P_{b,th}} \right]}{\beta_0 \ln \frac{p_{max}}{p_{min}}} = \{(3)\}$$

$$= T_{01} \frac{\ln \left(\frac{1}{N_{beh}-1} \frac{1-P_+}{P_+} \right) + \ln \left[(N_{beh}-1) \frac{P_{b,th}}{1-P_{b,th}} \right]}{\beta_0 \ln \frac{p_{max}}{p_{min}}}$$

$$\therefore \boxed{k\Delta t = T_{o1} \frac{\ln\left(\frac{1-P_{+}}{P_{+}} \cdot \frac{P_{b,th}}{1-P_{b,th}}\right)}{\beta_o \ln \frac{p_{max}}{p_{min}}}} \quad (15)$$

It's sort of clear ~~from~~ (15) that setting $\beta_o \neq 1$ would be a little artifizial/atheoretizal.

$\boxed{\text{oBE}_o + \text{oBE}_v \text{ evidence buildup, with forgetting}}$

As for (13) above, but with non-infinite T_{of}, we get, similarly to (7) etc:

$$A_b(k) = \beta_v V_b + \beta_o \left[\left(1 - \frac{\Delta t}{T_{of}}\right) \hat{A}_{b,o}(k-1) + \frac{\Delta t}{T_{o1}} \ln p_b \right]$$

The convergence for $\hat{A}_{b,o}$ will still be as in (8) so we get the overall convergence.

$$\boxed{A_b \to \beta_v V_b + \beta_o \frac{T_{of}}{T_{o1}} \ln p_b, \text{ when } t \to \infty} \quad (16)$$

Similarly to (10), we thus get:

$$\boxed{A_b(k) = \beta_v V_b + \beta_o \frac{T_{of}}{T_{o1}} \ln p_b \left(1 - e^{-k\Delta t / T_{of}}\right)} \quad (17)$$

$\boxed{\text{oBE}_o + \text{oBE}_v\text{, observing a "bad" decision, with forgetting}}$

As in the derivation of (15), but now with non-infinite T_{of}, we can now use (17):

$$P_{b_+}(k) = \frac{e^{-\beta_v + \beta_o \frac{T_{of}}{T_{o1}} \ln p_{max} (1 - e^{-k\Delta t / T_{of}})}}{e^{-\beta_v + \beta_o \frac{T_{of}}{T_{o1}} \ln p_{max} (1 - e^{-k\Delta t / T_{of}})} + (N_{beh} - 1) e^{0 + \beta_o \frac{T_{of}}{T_{o1}} \underbrace{\ln \alpha_o p_{max}}_{= (\ln \alpha_o + \ln p_{max})} (1 - e^{-k\Delta t / T_{of}})}}$$

Similarly to previous derivations:

$$\boxed{P_{t_f}(h) = \frac{e^{-\beta_v}}{e^{-\beta_v} + (N_{beh}-1)\, e^{\beta_0 \frac{T_{of}}{T_{o1}} \ln \alpha_0 \left(1-e^{-k\Delta t/T_{of}}\right)}}}$$

$$\boxed{= \frac{1}{1 + (N_{beh}-1)\, e^{\beta_v - \beta_0 \frac{T_{of}}{T_{o1}} \ln \frac{R_{max}}{R_{min}} \left(1-e^{-k\Delta t/T_{of}}\right)}}} \quad (18)$$

So, once more, we can get the time to confidence threshold $P_{b,th}$ from

$$\beta_v - \beta_0 \frac{T_{of}}{T_{o1}} \ln \frac{R_{max}}{R_{min}} \left(1 - e^{-k\Delta t/T_{of}}\right) = \ln\left(\frac{1}{N_{beh}-1} \, \frac{1-P_{b,th}}{P_{b,th}}\right)$$

$$\Leftrightarrow \beta_0 \frac{T_{of}}{T_{o1}} \ln \frac{R_{max}}{R_{min}} \left(1 - e^{-k\Delta t/T_{of}}\right) = \beta_v + \ln\left[(N_{beh}-1)\frac{P_{b,th}}{1-P_{b,th}}\right]$$

$$\Leftrightarrow 1 - e^{-k\Delta t/T_{of}} = \frac{T_{o1}\left\{\beta_v + \ln\left[(N_{beh}-1)\frac{P_{b,th}}{1-P_{b,th}}\right]\right\}}{\beta_0 T_{of} \ln \frac{R_{max}}{R_{min}}}$$

$$\Leftrightarrow e^{-k\Delta t/T_{of}} = 1 - \left(\text{———} \;''\; \text{———} \right)$$

$$\Leftrightarrow -\frac{k\Delta t}{T_{of}} = \ln\left(1 - \left(\text{———} \;''\; \text{———} \right)\right)$$

$$\Leftrightarrow k\Delta t = -T_{of} \ln\left(1 - \left(\text{———} \;''\; \text{———} \right)\right)$$

$$= T_{of} \ln \frac{1}{1 - \frac{T_{o1}\left\{\beta_v + \ln\left[(N_{beh}-1)\frac{P_{b,th}}{1-P_{b,th}}\right]\right\}}{\beta_0 T_{of} \ln \frac{R_{max}}{R_{min}}}}$$

$$\therefore \boxed{k\Delta t} = \{\text{as for } (15)\}$$

$$= T_{of} \ln \frac{1}{1 - \frac{T_{o1} \ln\left(\frac{1-P_t}{P_t} \cdot \frac{P_{b,th}}{1-P_{b,th}}\right)}{\beta_0 T_{of} \ln \frac{p_{max}}{p_{min}}}} \qquad (19)$$