Eq. (27) is the classical *replicator* (or logit) gradient. Define the simplex $\Delta^{k-1} := \{\pi \in (0,1]^k \mid \sum_i \pi_i = 1\}$ and write $\pi_\theta = (\pi_\theta(N_1), \dots, \pi_\theta(N_k))$.

Letting $\eta \to 0$ yields the ODE

$$\dot{\pi}_i \;=\; \pi_i\big(A(N_i) - \langle \pi, A\rangle\big), \qquad i = 1, \dots, k, \tag{28}$$

with $\langle \pi, A\rangle = \sum_j \pi_j A(N_j)$. Eq. (28) is the **replicator dynamics** for a fitness landscape A on Δ^{k-1}.

Consider the Kullback–Leibler divergence to the optimal pure strategy $\mathbf{e}_1 = (1, 0, \dots, 0)$,

$$V(\pi) \;=\; \sum_{i=1}^{k} \pi_i \ln\!\Big(\tfrac{\pi_i}{e_{1,i}}\Big) \;=\; -\ln \pi_1.$$

V is non-negative on Δ^{k-1} and $V(\pi) = 0$ iff $\pi = \mathbf{e}_1$.

Taking the time derivative along Eq. (28) gives

$$\frac{dV}{dt} = -\frac{\dot{\pi}_1}{\pi_1} = -\big(A(N_1) - \langle \pi, A\rangle\big) \;\le\; 0,$$

with equality iff $\pi_1 = 1$ *or* $A(N_1) = \langle \pi, A\rangle$. The latter can only happen if $\pi_1 = 1$ because $A(N_1) > A(N_j)$ for $j > 1$. Hence V is a strict Lyapunov function, and $\mathbf{e}_1$ is the *unique* asymptotically stable equilibrium of Eq. (28). All other stationary points (mixtures over sub-optimal arms) are unstable.

For sufficiently small but fixed η (choose $\eta < \frac{1}{A^*}$, which always exists), projected gradient ascent is a *perturbed* discretisation of Eq. (28). Standard results for primal-space mirror descent imply that the discrete iterates $\pi^{(t)} \equiv \pi_{\theta^{(t)}}$ converge almost surely to the set of asymptotically stable equilibria of the ODE, i.e. to $\{\mathbf{e}_1\}$. Therefore

$$\lim_{t\to\infty} \pi_{\theta^{(t)}}(N_i) = \begin{cases} 1, & \text{if } i = \arg\max_j A(N_j), \\ 0, & \text{otherwise.} \end{cases}$$

Because A may attain its maximum at several arms, the limit is a deterministic policy that places all probability on *some* maximiser of A.

Thus gradient ascent on Eq. (25) converges to a deterministic policy that always selects an optimal CoT length $N^* = \arg\max_{N\in\mathcal{A}} A(N)$, completing the proof. $\square$

G.7 Technical Lemmas

Lemma G.1 (test point)**.** *Let $F(x)$ be defined as*

$$F(x) = \ln\left(1 - \frac{T}{Mx}\right) + \frac{T}{Mx\left(1 - \frac{T}{Mx}\right)} + \ln\left(1 - \frac{T}{C}\right),$$

where $T, M, C \in \mathbb{R}^+$ satisfy the conditions:

- $0 < \frac{T}{C} < 0.9$*,*
- $0 < \frac{T}{Mx} < 1$*.*

Define x_0 as

$$x_0 = \frac{\sqrt{T(C-T)} + T}{M}.$$

Then, we have

$$F(x_0) < 0.$$

Proof. At $x = x_0$, note that

$$Mx_0 = \sqrt{T(C-T)} + T.$$

Thus,

$$1 - \frac{T}{Mx_0} = 1 - \frac{T}{T + \sqrt{T(C-T)}} = \frac{\sqrt{T(C-T)}}{T + \sqrt{T(C-T)}}.$$

Therefore,

$$\ln\Big(1-\frac{T}{Mx_0}\Big) = \ln\left(\frac{\sqrt{T(C-T)}}{T+\sqrt{T(C-T)}}\right) = \ln\sqrt{T(C-T)} - \ln\big(T+\sqrt{T(C-T)}\big).$$

Also, observe that

$$\frac{T}{Mx_0\Big(1-\frac{T}{Mx_0}\Big)} = \frac{T}{(T+\sqrt{T(C-T)})\Big(\frac{\sqrt{T(C-T)}}{T+\sqrt{T(C-T)}}\Big)} = \frac{T}{\sqrt{T(C-T)}} = \sqrt{\frac{T}{C-T}}.$$

It is convenient to introduce the change of variable

$$u = \sqrt{\frac{T}{C-T}},$$

so that

$$T = u^2(C-T), \quad \sqrt{T(C-T)} = u(C-T).$$

Then we have

$$T + \sqrt{T(C-T)} = u^2(C-T) + u(C-T) = u(C-T)(u+1).$$

In these terms we have:

$$\ln\sqrt{T(C-T)} = \ln\big[u(C-T)\big] = \ln u + \ln(C-T),$$

$$\ln\big(T+\sqrt{T(C-T)}\big) = \ln\big[u(C-T)(u+1)\big] = \ln u + \ln(C-T) + \ln(u+1),$$

and

$$\sqrt{\frac{T}{C-T}} = u.$$

Finally, we have

$$\ln\Big(1-\frac{T}{C}\Big) = -\ln(\frac{C}{C-T}) = -\ln(u^2+1)$$

Thus, the function $F(x_0)$ becomes

$$\begin{aligned} F(x_0) &= \ln u + \ln(C-T) - \big(\ln u + \ln(C-T) + \ln(u+1)\big) + u - \ln(u^2+1) && (29)\\ &= -\ln(u+1) + u - \ln(u^2+1)\,, && (30)\end{aligned}$$

where $u = \sqrt{\frac{T}{C-T}} \in (0,3)$. It is easy to show $F(x_0) < 0$ when $u \in (0,3)$. □

Lemma G.2 (Estimation of the n-th Moment of the Beta Distribution)**.** *Let* $x \sim \mathrm{Beta}(\alpha,\beta)$. *Then*

$$\mathbb{E}[(1-x)^n] \le \left(1-\frac{\alpha}{\alpha+\beta+n-1}\right)^n.$$

Proof.

$$
\begin{aligned}
\mathbb{E}[(1-x)^n] &= \frac{1}{B(\alpha,\beta)} \int_0^1 (1-x)^n x^{\alpha-1}(1-x)^{\beta-1}\,dx \\
&= \frac{1}{B(\alpha,\beta)} \int_0^1 x^{\alpha-1}(1-x)^{\beta+n-1}\,dx \\
&= \frac{B(\alpha,\beta+n)}{B(\alpha,\beta)} \\
&= \frac{\Gamma(\alpha)\Gamma(\beta+n)}{\Gamma(\alpha+\beta+n)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\
&= \frac{\Gamma(\beta+n)}{\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+n)} \\
&= \prod_{i=0}^{n-1} \frac{\beta+i}{\alpha+\beta+i} \\
&\leq \left(\frac{\beta+n-1}{\alpha+\beta+n-1}\right)^n \\
&= \left(1-\frac{\alpha}{\alpha+\beta+n-1}\right)^n.
\end{aligned}
$$

□

H Pseudo-code of Length-filtered Vote

Algorithm 1 Length-filtered Vote

Input: Model f_θ, Question q, Space of All Possible Answers A, Number of Total Groups M, Number of Selected Groups K, Group Width D

Output: Final Answer $\hat{a}$

Sample candidates $c_1, \ldots, c_n \overset{i.i.d.}{\sim} f_\theta(q)$

Define $\mathcal{A}(c)$ as the corresponding answer of candidates c.

Define $p_j \in [0,1]^{|\mathcal{A}|}$ as the frequency of each answer in length group L_j.

for $j = 1$ to m **do**

  $L_j = \{c_i \mid \ell(c_i) \in [D*(j-1), D*j)\,, i = 1, \cdots, n\}$

  for $a \in \mathcal{A}$ **do**

$$p_j[a] = \frac{\sum_{c \in L_j} \mathbb{I}(\mathcal{A}(c) = a)}{|L_j|}$$

  end for

end for

$\{s_1, \ldots, s_K\} = \arg\min_{S \subseteq \{1,\ldots,M\}, |S|=K} \sum_{s \in S} H(p_s)$

$\hat{a} = \arg\max_{a \in A} \sum_{c \in L_{s_1} \cup \cdots \cup L_{s_K}} \mathbb{I}(\mathcal{A}(c) = a)$

return $\hat{a}$