D. Claims and Proofs in Rebuttal

We take the cross-entropy loss as an example. Suppose we have two ordinal feedbacks Z and Z', taking values in $\mathcal{Z}$ and $\mathcal{Z}'$, respectively. Suppose Z is a hierarchical expectation of Z', of which the coupling is formulated as (W, W'). We make further assumptions here besides the "wisdom of the crowd" assumption (Assumption 3.1).

Assumption D.1. Assume that each Z_i (or Z_i') is independent. Furthermore, assume that each $z_j \in \mathcal{Z}$ is bounded away from 0 and 1 by

$$\gamma < z_j < 1 - \gamma.$$

The above assumption is not restrictive: the independence naturally holds if each annotation is independent. The boundedness can be satisfied by clipping the feedback options, and is assumed to prevent those infinite values attained in the supremum step due to logarithms (if $\varepsilon = +1$) while making sure those maxima can be obtained if $\varepsilon = -1$.

Assumption D.2. Assume that the function class $\mathcal{H}$ is rich enough (w.r.t. γ) with high probability ($\geq 1 - \delta$). That is, $\exists \delta \geq 0$, s.t.

$$E \coloneqq \left\{[\gamma, 1-\gamma]^n \in \left\{\left(h(x_1, y_{1,1}, y_{1,2}), \ldots, h(x_n, y_{n,1}, y_{n,2})\right) \middle| h \in \mathcal{H}\right\}\right\}$$

holds with probability at least $1 - \delta$.

We denote the number of positive Rademacher variables (that is, those $\varepsilon_i = +1$) by τ (which is a random variable itself). From the definition of τ, we know

$$\mathbb{E}_\varepsilon[\tau] = \frac{n}{2}. \tag{8}$$

Without loss of generality, we assume that $\varepsilon_1 = \cdots = \varepsilon_\tau = +1$, and $\varepsilon_{\tau+1} = \cdots = \varepsilon_n = -1$. Due to the assumption that each $(x_i, y_{i,1}, y_{i,2}, Z_i)$ is i.i.d. (or Z_i'), such a τ summarizes all the dependence on ε. The two Rademacher complexities are (omitting the dependence on x, y_1, y_2 for notational simplicity and denoting $h_i \coloneqq h(x_i, y_{i,1}, y_{i,2})$)

$$\begin{aligned}\mathrm{Rad}_{\mathcal{Z},n}(\ell \circ \mathcal{H}) = \frac{1}{n}\mathbb{E}_{\tau,W}\Bigg[\sup_{h\in\mathcal{H}} &\sum_{i=1}^{\tau} -W_i \cdot \log(h_i) - (1 - W_i) \cdot \log(1 - h_i) \\ &+ \sum_{i=\tau+1}^{n} W_i \cdot \log(h_i) + (1 - W_i) \cdot \log(1 - h_i)\Bigg],\end{aligned}$$

and

$$\begin{aligned}\mathrm{Rad}_{\mathcal{Z}',n}(\ell \circ \mathcal{H}) = \frac{1}{n}\mathbb{E}_{\tau,W}\Bigg[\mathbb{E}_{W'}\Bigg[\sup_{h\in\mathcal{H}} &\sum_{i=1}^{\tau} -W_i' \cdot \log(h_i) - (1 - W_i') \cdot \log(1 - h_i) \\ &+ \sum_{i=\tau+1}^{n} W_i' \cdot \log(h_i) + (1 - W_i') \cdot \log(1 - h_i)\Bigg| W\Bigg]\Bigg].\end{aligned}$$

Lemma D.3. *Under Assumptions 4.8 and D.1, if E holds, then the following holds (here, q_i's are real numbers independent of h_i's):*

$$\begin{aligned}&\sup_{h\in\mathcal{H}} \sum_{i=1}^{\tau} -W_i \cdot \log(h_i) - (1 - W_i) \cdot \log(1 - h_i) + \sum_{i=\tau+1}^{n} W_i \cdot \log(h_i) + (1 - W_i) \cdot \log(1 - h_i) \\ &= \sup_{h\in\mathcal{H}} \sum_{i=1}^{\tau} -W_i \cdot \log(h_i) - (1 - W_i) \cdot \log(1 - h_i) + \sup_{q_i\in[\gamma,1-\gamma]} \sum_{i=\tau+1}^{n} W_i \cdot \log(q_i) + (1 - W_i) \cdot \log(1 - q_i),\end{aligned}$$

and

$$\begin{aligned}&\sup_{h\in\mathcal{H}} \sum_{i=1}^{\tau} -W_i' \cdot \log(h_i) - (1 - W_i') \cdot \log(1 - h_i) + \sum_{i=\tau+1}^{n} W_i' \cdot \log(h_i) + (1 - W_i') \cdot \log(1 - h_i) \\ &\geq \sup_{h\in\mathcal{H}} \sum_{i=1}^{\tau} -W_i' \cdot \log(h_i) - (1 - W_i') \cdot \log(1 - h_i) + \sup_{q_i\in[\gamma,1-\gamma]} \sum_{i=\tau+1}^{n} W_i' \cdot \log(q_i) + (1 - W_i') \cdot \log(1 - q_i).\end{aligned}$$

Proof. From the fact that $\{(h_1, \dots, h_\tau, q_{\tau+1}, \dots, q_n) | h \in \mathcal{H}\} \subset \{(h_1, \dots, h_n) | h \in \mathcal{H}\}$ under E, we know that

$$\begin{aligned}
&\sup_{h \in \mathcal{H}} \sum_{i=1}^{\tau} -W_i \cdot \log(h_i) - (1 - W_i) \cdot \log(1 - h_i) + \sum_{i=\tau+1}^{n} W_i \cdot \log(h_i) + (1 - W_i) \cdot \log(1 - h_i) \\
&\geq \sup_{h \in \mathcal{H}, q_i \in [\gamma, 1-\gamma]} \sum_{i=1}^{\tau} -W_i \cdot \log(h_i) - (1 - W_i) \cdot \log(1 - h_i) + \sum_{i=\tau+1}^{n} W_i \cdot \log(q_i) + (1 - W_i) \cdot \log(1 - q_i) \\
&= \sup_{h \in \mathcal{H}} \sum_{i=1}^{\tau} -W_i \cdot \log(h_i) - (1 - W_i) \cdot \log(1 - h_i) + \sup_{q_i \in [\gamma, 1-\gamma]} \sum_{i=\tau+1}^{n} W_i \cdot \log(q_i) + (1 - W_i) \cdot \log(1 - q_i),
\end{aligned}$$

and

$$\begin{aligned}
&\sup_{h \in \mathcal{H}} \sum_{i=1}^{\tau} -W_i' \cdot \log(h_i) - (1 - W_i') \cdot \log(1 - h_i) + \sum_{i=\tau+1}^{n} W_i' \cdot \log(h_i) + (1 - W_i') \cdot \log(1 - h_i) \\
&\geq \sup_{h \in \mathcal{H}, q_i \in [\gamma, 1-\gamma]} \sum_{i=1}^{\tau} -W_i' \cdot \log(h_i) - (1 - W_i') \cdot \log(1 - h_i) + \sum_{i=\tau+1}^{n} W_i' \cdot \log(q_i) + (1 - W_i') \cdot \log(1 - q_i) \\
&= \sup_{h \in \mathcal{H}} \sum_{i=1}^{\tau} -W_i' \cdot \log(h_i) - (1 - W_i') \cdot \log(1 - h_i) + \sup_{q_i \in [\gamma, 1-\gamma]} \sum_{i=\tau+1}^{n} W_i' \cdot \log(q_i) + (1 - W_i') \cdot \log(1 - q_i).
\end{aligned}$$

From Assumption D.1, we know that $\forall h \in \mathcal{H}$,

$$\sum_{i=\tau+1}^{n} W_i \cdot \log(h_i) + (1 - W_i) \cdot \log(1 - h_i) \leq \sum_{i=\tau+1}^{n} -\mathrm{Ent}(W_i) = \sup_{q_i \in [\gamma, 1-\gamma]} \sum_{i=\tau+1}^{n} W_i \cdot \log(q_i) + (1 - W_i) \cdot \log(1 - q_i),$$

which implies that

$$\begin{aligned}
&\sup_{h \in \mathcal{H}} \sum_{i=1}^{\tau} -W_i \cdot \log(h_i) - (1 - W_i) \cdot \log(1 - h_i) + \sum_{i=\tau+1}^{n} W_i \cdot \log(h_i) + (1 - W_i) \cdot \log(1 - h_i) \\
&\leq \sup_{h \in \mathcal{H}} \sum_{i=1}^{\tau} -W_i \cdot \log(h_i) - (1 - W_i) \cdot \log(1 - h_i) + \sup_{h \in \mathcal{H}} \sum_{i=\tau+1}^{n} W_i \cdot \log(h_i) + (1 - W_i) \cdot \log(1 - h_i) \\
&= \sup_{h \in \mathcal{H}} \sum_{i=1}^{\tau} -W_i \cdot \log(h_i) - (1 - W_i) \cdot \log(1 - h_i) + \sup_{q_i \in [\gamma, 1-\gamma]} \sum_{i=\tau+1}^{n} W_i \cdot \log(q_i) + (1 - W_i) \cdot \log(1 - q_i).
\end{aligned}$$

Combining the above, we conclude the proof. □

Lemma D.4. *Denote the supremum over the cross-entropy as*

$$f(z) := \sup_{q \in [\gamma, 1-\gamma]} z \cdot \log(q) + (1 - z) \cdot \log(1 - q).$$

Then

$$\mathrm{Rad}_{\mathcal{Z}',n}(\ell \circ \mathcal{H}) - \mathrm{Rad}_{\mathcal{Z},n}(\ell \circ \mathcal{H}) \geq \frac{1}{n} \mathbb{E}_{\tau,W} \left[\sum_{i=\tau+1}^{n} \mathbb{E}_{W'} \left[f(W_i') \middle| W_i \right] \right] - \frac{1}{n} \mathbb{E}_{\tau,W} \left[\sum_{i=\tau+1}^{n} f(W_i) \right].$$

Proof. Due to the property of supremum, the supremum taken over the entire $(q_{\tau+1}, \dots, q_n) \in [\gamma, 1-\gamma]^{n-\tau}$ is equivalent to the supremum taken step by step (that is, by taking each q_i's supremum one by one). Therefore,

$$\sup_{q_i \in [\gamma, 1-\gamma]} \sum_{i=\tau+1}^{n} W_i \cdot \log(q_i) + (1 - W_i) \cdot \log(1 - q_i) = \sum_{i=\tau+1}^{n} \sup_{q_i \in [\gamma, 1-\gamma]} W_i \cdot \log(q_i) + (1 - W_i) \cdot \log(1 - q_i) = \sum_{i=\tau+1}^{n} f(W_i).$$

Similar conclusions also hold for W_i''s. We conclude the proof by combining the facts with Lemma D.3. □

Lemma D.5. *Suppose $w \in [\gamma, 1-\gamma]$ and $w' \in [0,1]$. Denote the clipped w' by*

$$w'_\gamma \coloneqq \max\{\gamma, \min\{w', 1-\gamma\}\}.$$

Then

$$f(w') \geq f(w) + \frac{\mathrm{d}f}{\mathrm{d}z}\bigg|_{z=w} \cdot (w'-w) + 2(w'_\gamma - w)^2.$$

Proof. For those $w' \in [\gamma, 1-\gamma]$, $w'_\gamma = w'$. The function f of w' is the entropy, of which the second-order derivative is

$$\frac{\mathrm{d}^2 f}{\mathrm{d}z^2}\bigg|_{z=w'} = \frac{1}{w'(1-w')} \geq 4.$$

Then by Taylor's formula, we know that $\exists \theta \in [0,1]$, s.t.

$$\begin{aligned}
f(w') &= f(w) + \frac{\mathrm{d}f}{\mathrm{d}z}\bigg|_{z=w} \cdot (w'-w) + \frac{1}{2} \cdot \frac{\mathrm{d}^2 f}{\mathrm{d}z^2}\bigg|_{z=\theta w + (1-\theta) w'} \cdot (w'-w)^2 \\
&\geq f(w) + \frac{\mathrm{d}f}{\mathrm{d}z}\bigg|_{z=w} \cdot (w'-w) + 2(w'-w)^2.
\end{aligned}$$

For those $w' > 1-\gamma$, $w'_\gamma = 1-\gamma$. The function f is linear in the interval $[1-\gamma, 1]$, of which the first-order derivative is $\frac{\mathrm{d}f}{\mathrm{d}z}|_{z=1-\gamma}$ (easy to check f is continuously differentiable at $1-\gamma$). Thus,

$$\begin{aligned}
f(w') &= f(w'_\gamma) + \frac{\mathrm{d}f}{\mathrm{d}z}\bigg|_{z=w'_\gamma} \cdot (w' - w'_\gamma) \\
&\geq f(w'_\gamma) + \frac{\mathrm{d}f}{\mathrm{d}z}\bigg|_{z=w} \cdot (w' - w'_\gamma) \\
&\geq f(w) + \frac{\mathrm{d}f}{\mathrm{d}z}\bigg|_{z=w} \cdot (w' - w'_\gamma + w'_\gamma - w) + (w'_\gamma - w)^2,
\end{aligned}$$

which verifies the proof.

For those $w' < \gamma$, the conclusion holds similarly. $\square$

Proposition D.6. *Under Assumptions 3.1, D.1, and D.2, we have*

$$\mathrm{Rad}_{\mathcal{Z}',n}(\ell \circ \mathcal{H}) - \mathrm{Rad}_{\mathcal{Z},n}(\ell \circ \mathcal{H}) = \Omega\left(\mathbb{E}_W[\mathrm{Var}(W'|W)]\right).$$

Proof. Denote the conditional distribution of W'_i on observing W_i by $P_{W'_i|W_i}$. From Lemmas D.4 and D.5, we obtain that

$$\begin{aligned}
&\mathrm{Rad}_{\mathcal{Z}',n}(\ell \circ \mathcal{H}) - \mathrm{Rad}_{\mathcal{Z},n}(\ell \circ \mathcal{H}) \\
\geq\ &\frac{1}{n}\mathbb{E}_{\tau,W}\left[\sum_{i=\tau+1}^{n} \mathbb{E}_{W'}\left[f(W'_i)|W_i\right]\right] - \frac{1}{n}\mathbb{E}_{\tau,W}\left[\sum_{i=\tau+1}^{n} f(W_i)\right] \\
=\ &\frac{1}{n}\mathbb{E}_{\tau,W}\left[\sum_{i=\tau+1}^{n} \int f(W'_i) - f(W_i)\,\mathrm{d}P_{W'_i|W_i}\right] \\
\geq\ &\frac{1}{n}\mathbb{E}_{\tau,W}\left[\sum_{i=\tau+1}^{n} \int f(W_i) + \frac{\mathrm{d}f}{\mathrm{d}z}\bigg|_{z=W_i} \cdot (W'_i - W_i) + 2(W'_{i,\gamma} - W_i)^2 - f(W_i)\,\mathrm{d}P_{W'_i|W_i}\right] \\
=\ &\frac{1}{n}\mathbb{E}_{\tau,W}\left[\sum_{i=\tau+1}^{n} \int 2(W'_{i,\gamma} - W_i)^2\,\mathrm{d}P_{W'_i|W_i}\right].
\end{aligned}$$

By Assumption D.1, we know that

$$\frac{(W'_{i,\gamma} - W_i)^2}{(W'_i - W_i)^2} = \Omega(1),$$

which implies that

$$\int 2(W'_{i,\gamma} - W_i)^2 \, \mathrm{d}P_{W'_i|W_i} \geq \Omega(1) \cdot \int (W'_i - W_i)^2 \, \mathrm{d}P_{W'_i|W_i} = \Omega\big(\mathrm{Var}(W'_i|W_i)\big).$$

Since $\mathbb{E}[\tau] = \frac{n}{2}$, we know that

$$\mathrm{Rad}_{\mathcal{Z}',n}(\ell \circ \mathcal{H}) - \mathrm{Rad}_{\mathcal{Z},n}(\ell \circ \mathcal{H}) = \Omega\big(\mathrm{Var}(W'_i|W_i)\big).$$

$\square$