More Detailed Derivation of Eq.(2) in the IGL Paper [Xie et al., 2021]

For any policy π, we have

$$\begin{aligned}
&\mathbb{E}_{(x,a)\sim d_0\times\pi}[\psi(y)\mathbb{1}(r=1)] \\
&= \sum_{x,a} d_0(x)\pi(a|x)\sum_{y,r}\Pr(y,r|x,a)\psi(y)\mathbb{1}(r=1) \\
&= \sum_{x,a} d_0(x)\pi(a|x)\sum_{y,r}\underbrace{\Pr(y|r,x,a)}_{=\Pr(y|r)\text{ by } x,a\perp\!\!\!\perp y|r}\Pr(r|x,a)\psi(y)\mathbb{1}(r=1) \\
&= \sum_{x,a} d_0(x)\pi(a|x)\sum_{r}\Pr(r|x,a)\mathbb{1}(r=1)\sum_{y}\Pr(y|r)\psi(y) \\
&= \sum_{x,a} d_0(x)\pi(a|x)\Pr(r=1|x,a)\sum_{y}\Pr(y|r=1)\psi(y) \\
&= V(\pi)\mathbb{E}\left[\psi(y)|r=1\right],
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}_{(x,a)\sim d_0\times\pi}[\psi(y)\mathbb{1}(r=0)] \\
&= \sum_{x,a} d_0(x)\pi(a|x)\sum_{y,r}\Pr(y,r|x,a)\psi(y)\mathbb{1}(r=0) \\
&= \sum_{x,a} d_0(x)\pi(a|x)\sum_{y,r}\underbrace{\Pr(y|r,x,a)}_{=\Pr(y|r)\text{ by } x,a\perp\!\!\!\perp y|r}\Pr(r|x,a)\psi(y)\mathbb{1}(r=0) \\
&= \sum_{x,a} d_0(x)\pi(a|x)\sum_{r}\Pr(r|x,a)\mathbb{1}(r=0)\sum_{y}\Pr(y|r)\psi(y) \\
&= \sum_{x,a} d_0(x)\pi(a|x)\Pr(r=0|x,a)\sum_{y}\Pr(y|r=0)\psi(y) \\
&= (1-V(\pi))\mathbb{E}\left[\psi(y)|r=0\right].
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathbb{E}_{(x,a)\sim d_0\times\pi}[\psi(y)\mathbb{1}(r=1)+\psi(y)\mathbb{1}(r=0)] - \mathbb{E}_{(x,a)\sim d_0\times\pi_{\mathsf{bad}}}[\psi(y)\mathbb{1}(r=1)+\psi(y)\mathbb{1}(r=0)] \\
&= \Big[V(\pi)\underbrace{\mathbb{E}\left[\psi(y)|r=1\right]}_{=:\psi_1} + (1-V(\pi))\underbrace{\mathbb{E}\left[\psi(y)|r=0\right]}_{=:\psi_0}\Big] \\
&\quad - \Big[V(\pi_{\mathsf{bad}})\mathbb{E}\left[\psi(y)|r=1\right] + (1-V(\pi_{\mathsf{bad}}))\mathbb{E}\left[\psi(y)|r=0\right]\Big] \\
&= V(\pi)\underbrace{(\psi_1-\psi_0)}_{=:\Delta\psi} + \psi_0 - \Big[V(\pi_{\mathsf{bad}})(\psi_1-\psi_0)+\psi_0\Big] \\
&= (V(\pi)-V(\pi_{\mathsf{bad}}))\Delta\psi.
\end{aligned}$$

References

Tengyang Xie, John Langford, Paul Mineiro, and Ida Momennejad. Interaction-grounded learning. In *International Conference on Machine Learning*, pages 11414–11423. PMLR, 2021.