Details of the derivation of information gain in Frank et al. (submitted), "Modeling the dynamics of classroom education using teaching games."

As described in the text, we are interested in computing

$$IG(e) = D_{KL}(B_T||B_S) - D_{KL}(B_T||B_{S+e}) \tag{1}$$

where the divergence measure is computed in closed form for e.g., B_T and B_S, as

$$\begin{aligned} D_{KL}(B_T||B_S) =& \log(\frac{B(\alpha_S, \beta_S)}{B(\alpha_T, \beta_T)}) + (\alpha_T - \alpha_S)\psi(\alpha_T) \\ &+ (\beta_T - \beta_S)\psi(\beta_T) + (\alpha_T - \alpha_S + \beta_T - \beta_S)\psi(\alpha_T + \beta_T). \end{aligned} \tag{2}$$

where ψ denotes the digamma function and $B(a, b)$ denotes the beta function. We can substitute Equation 2 into Equation 1 to get

$$\begin{aligned} IG(e) =& \log(\frac{B(\alpha_S, \beta_S)}{B(\alpha_T, \beta_T)}) + (\alpha_T - \alpha_S)\psi(\alpha_T) \\ &+ (\beta_T - \beta_S)\psi(\beta_T) + (\alpha_T - \alpha_S + \beta_T - \beta_S)\psi(\alpha_T + \beta_T) \\ &- \log(\frac{B(\alpha_{S+e}, \beta_{S+e})}{B(\alpha_T, \beta_T)}) - (\alpha_T - \alpha_{S+e})\psi(\alpha_T) \\ &- (\beta_T - \beta_{S+e})\psi(\beta_T) - (\alpha_T - \alpha_{S+e} + \beta_T - \beta_{S+e})\psi(\alpha_T + \beta_T) \end{aligned} \tag{3}$$

Consider the case where e is a single 1 (head). Then $\alpha_{S+e} = \alpha_S + 1$ and $\beta_{S+e} = \beta_S$, so we can simplify Equation 3 to

$$
\begin{aligned}
IG(e) =& \log(\frac{B(\alpha_S, \beta_S)}{B(\alpha_T, \beta_T)}) - \log(\frac{B(\alpha_S + 1, \beta_S)}{B(\alpha_T, \beta_T)}) \\
&+ (\alpha_T - \alpha_S)\psi(\alpha_T) + (\alpha_T - \alpha_S + \beta_T - \beta_S)\psi(\alpha_T + \beta_T) \\
&- (\alpha_T - \alpha_S - 1)\psi(\alpha_T) - (\alpha_T - \alpha_S + \beta_T - \beta_S - 1)\psi(\alpha_T + \beta_T) \\
=& \log(\frac{B(\alpha_S, \beta_S)}{B(\alpha_T, \beta_T)}) - \log(\frac{B(\alpha_S + 1, \beta_S)}{B(\alpha_T, \beta_T)}) \\
&+ \psi(\alpha_T) + \psi(\alpha_T + \beta_T). \\
=& \log(\frac{B(\alpha_S, \beta_S)}{B(\alpha_S + 1, \beta_S)}) \\
&+ \psi(\alpha_T) + \psi(\alpha_T + \beta_T).
\end{aligned} \tag{4}
$$

And, since

$$
B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}, \tag{5}
$$

we can rewrite the first term and reduce:

$$
\begin{aligned}
IG(e) =& \log(\frac{\frac{\Gamma(\alpha_S)\Gamma(\beta_S)}{\Gamma(\alpha_S + \beta_S)}}{\frac{\Gamma(\alpha_S + 1)\Gamma(\beta_S)}{\Gamma(\alpha_S + \beta_S + 1)}}) \\
&+ \psi(\alpha_T) + \psi(\alpha_T + \beta_T). \\
=& \log(\frac{\Gamma(\alpha_S)\Gamma(\beta_S)\Gamma(\alpha_S + \beta_S + 1)}{\Gamma(\alpha_S + \beta_S)\Gamma(\alpha_S + 1)\Gamma(\beta_S)}) \\
&+ \psi(\alpha_T) + \psi(\alpha_T + \beta_T). \\
=& \log(\frac{\Gamma(\alpha_S)\Gamma(\alpha_S + \beta_S + 1)}{\Gamma(\alpha_S + \beta_S)\Gamma(\alpha_S + 1)}) \\
&+ \psi(\alpha_T) + \psi(\alpha_T + \beta_T).
\end{aligned} \tag{6}
$$

Then, since

$$
\frac{\Gamma(x) + 1}{\Gamma(x)} = x, \tag{7}
$$

we can reduce the previous formulation a bit further, to

$$IG(e) = \log(\frac{\alpha_S + \beta_S}{\alpha_S}) + \psi(\alpha_T) + \psi(\alpha_T + \beta_T), \tag{8}$$

which is Equation 5 in the main text.