Technical Appendix

1. Undifferentiated Learning Model

In undifferentiated learning, consumers learn about the overall distribution of an attribute (e.g. quality and price) at a given restaurant across the entire consumer population. Consumers have prior beliefs about the distribution of the attribute in the population, which is also assumed to be the same across consumers. They then update these beliefs based on reviewers' evaluations of the attribute. The consumers then use updated beliefs of the attribute when formulating expected utilities of dining in the reviewed restaurants.

Following the discussions in the paper, we have the following relationships:

$$A_{kj} = A_j + \xi_{kj} \tag{A.1}$$

$$\mathbf{R}_{kj} = A_{kj} \cdot \mathbf{e}_L + \boldsymbol{\varepsilon}_{kj} \tag{A.2}$$

We assume that $\boldsymbol{\varepsilon}_{kj}$ follows a normal distribution with a specific parameterization of the variance-covariance matrix,

$$\boldsymbol{\varepsilon}_{kj} \sim N(\mathbf{0}, \lambda^{-1}\Omega\sigma^2_{\xi,j}) \tag{A.3}$$

where λ is a scalar.

Combining equations (A.2) and (A.3) we have

$$\mathbf{R}_{kj} = A_j \cdot \mathbf{e}_L + \xi_{kj} \cdot \mathbf{e}_L + \boldsymbol{\varepsilon}_{kj} \tag{A.4}$$

and the variance of $\mathbf{R}_{kj}$ is

$$Var\left(\mathbf{R}_{kj}\right) = \left(\mathbf{1}_L + \lambda^{-1}\Omega\right)\sigma^2_{\xi,j} \equiv \tilde{\Omega}\sigma^2_{\xi,j} \tag{A.5}$$

where $\mathbf{1}_L$ is an $L \times L$ matrix with every component being 1.

Our model allows the user to learn about $\sigma^2_{\xi,j}$. Before reading any reviews, we assume that her prior belief is distributed as inverse Gamma,

$$\sigma^2_{\xi,j} \sim IG\left(\frac{a_0}{2}, \frac{b_0}{2}\right)$$

where a_0 and b_0 are the shape and scale parameters, respectively, of the distribution. Conditional on $\sigma^2_{\xi,j}$, the user's prior belief of the mean experience is assumed to be distributed as

$$A_j \mid \sigma^2_{\xi,j} \sim N(A_0, \tau_0^{-1} \cdot \sigma^2_{\xi,j})$$

where A_0 is the prior expectation of A_j, and τ_0 is a scale parameter. The inverse of τ_0 measures the degree of the prior uncertainty.

For the ease of exposition, assume that the user only reads reviews for restaurant j. After reading K reviews, the user's beliefs of $\sigma^2_{\xi,j}$ and A_j are assumed to be updated using the Bayes rule. The marginal distribution of A_j, conditional on the information set $\boldsymbol{I}_{Kj} = \{\boldsymbol{R}_{1j}, \ldots, \boldsymbol{R}_{Kj}\}$ from all reviews, can be derived as a t-distribution with mean $E(A_j \mid I_K) = A_{jK}$ and variance $Var(A_j \mid I_K) = \tau_{jK}^{-1} \cdot \dfrac{b_{jK}}{a_{jK} - 2}$. The derivation of the posterior beliefs and the specification of the updated parameters are in the next section of this document.

The marginal distribution of $\sigma^2_{\xi,j}$ conditional on I_K is an inverse-Gamma distribution with expected value $E(\sigma^2_{\xi,j} \mid I_K) = b_{jK} / \left(a_{jK} - 2\right)$. The expected variance of the user's own experience, $A_{ij} = A_j + \xi_{ij}$, is equal to the sum of the variance of the expectation of the true mean A_j and the variance of ξ_{ij}. Therefore, $\mathrm{var}(A_{ij} \mid I_K) = \left(1 + \tau_{jK}^{-1}\right) \cdot b_{jK} / \left(a_{jK} - 2\right)$.

The expected utility function, after reading K reviews, in equation (1) can be rewritten (by replacing quality Q and price C for the attribute A) as the following:

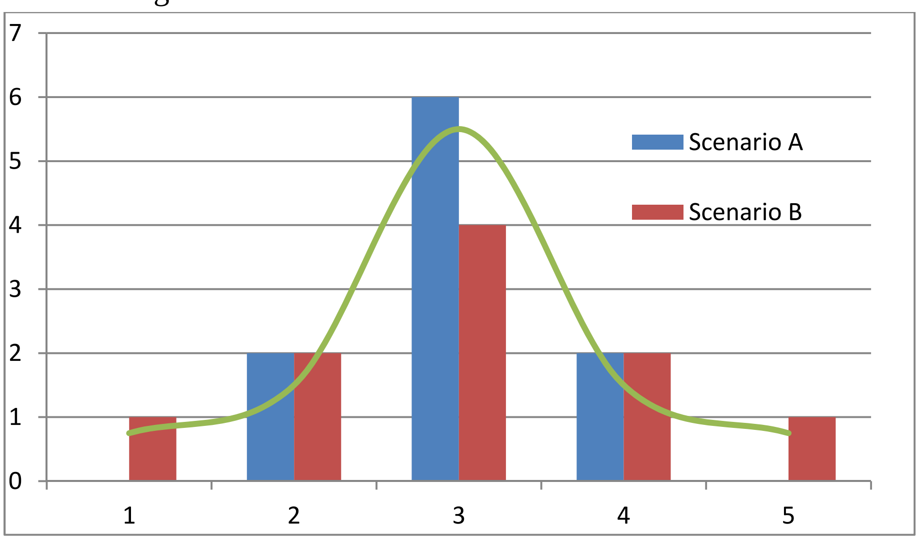
$$E\left[U_{ij} \mid I_K\right] = \alpha_{ij} + w_i^Q \left\{ Q_{jK} + \gamma_i^Q Q_{jK}^2 + \gamma_i^Q \left(1 + \frac{1}{\tau_{jK}^Q}\right) \frac{b_{jK}^Q}{a_{jK}^Q - 2} \right\} + w_i^C C_{jK} + \epsilon_{ij}$$

where Q_{jK} and C_{jK} are the updated means, and $\left(1 + 1/\tau_{jK}^Q\right) \cdot \left(b_{jK}^Q / (a_{jK}^Q - 2)\right)$ is the updated variance for quality.

In a standard learning model in the previous literature, conditional on the average, the diversity of review evaluations does not affect consumers' updating process; in our model, however, it will increase the user's uncertainty of $E(\sigma^2_{\xi,j} \mid I_K)$ and impact the choice probability. We use the following numerical example for an illustration.

Example: Figure A.1 illustrates two scenarios of 10 reported ratings for the quality of a restaurant. The mean ratings in scenarios A and B are equal to 3. However, scenario A is more concentrated around the mean rating (two report a rating 2, six report 3, and two report 4) than scenario B (one reports a rating 1, two report 2, four report 3, two report 4, and one reports 5). The true population distribution of ratings is the curve in the figure with variance 0.9, which is between the variance of ratings in scenarios A (0.4) and B (1.2). If we assume that the variance is known without uncertainty, the updated expected quality $E(Q_j \mid I_K)$ is 3.0, and the associated uncertainty $Var(Q_j \mid I_K)$ is 0.041, exactly the same in both scenarios. If $\sigma^2_{\xi,j}$ has to be learned, however, the posterior expected $\sigma^2_{\xi,j}$ in scenario A is 0.22, and 0.58 in scenario B, and the posterior $Var(Q_j \mid I_K)$ in scenario is 0.037, and in scenario B is 0.098, respectively.[1]

Figure A.1. An Illustration of Variance in Reviews



Without accounting for the learning of variance and how this will impact the consumer choice, we may have a biased measure of the economic value of online reviews, since the above

[1] The calculations are based on the following assumptions of priors: $\lambda = 1.0$, $A_0 = 3.0$, $\tau_0 = 1.0$, $a_0 = 3.0$, and $b_0 = 0.45$. For the model without learning for variance, we assume $\sigma^2_{\xi,j} = 0.45$, the true population variance of quality perception.

example shows that the two sets of reviews in scenarios A and B will be restricted to carry the same information thus have the same economic value. The extent of the bias depends on consumers' risk preference in the utility function. For a risk-averse consumer, reviews in scenario A will reduce her uncertainty more than in B. The reviews will increase the user's expected utility and the probability of choosing the restaurant; therefore, they have a higher value for both the user and the restaurant. We have shown in the reduced-form regression (see Table 3 in the main text) that the variance of reviewers' quality ratings has a significant negative effect on consumers' choice probability, a result that will not exist if consumers are risk-neutral or the variance is common knowledge. This provides indirect evidence supporting the importance of modeling the learning of $\sigma^2_{\xi,j}$ in our context.

2. Undifferentiated Learning Model Posterior Distribution Derivation

Following section A.1, evaluations $\mathbf{R}_{kj}$ are specified as

$$\mathbf{R}_{kj} = A_{kj} \cdot \mathbf{e}_L + \varepsilon_{kj}$$

$$Var(\mathbf{R}_{kj}) = (\mathbf{1}_L + \lambda^{-1}\Omega)\sigma^2_{\xi,j} \equiv \tilde{\Omega}\sigma^2_{\xi,j}$$

where $A_{kj} = A_j + \xi_{kj}$ and $\boldsymbol{\epsilon}_{kj} \sim N(\mathbf{0}, \lambda^{-1}\Omega\sigma^2_{\xi,j})$. A consumer updates the posterior beliefs of parameters of $(A_j, \sigma^2_{\xi,j})$ in this learning framework. Since these two parameters are the sufficient statistics of a normal distribution, we use the natural conjugate distribution of Normal-Inverse Gamma (when both mean and variance are unknown) as the prior distributions for the two parameters. That is:

$$\sigma^2_{\xi,j} \sim IG(\frac{a_0}{2}, \frac{b_0}{2})$$
$$A_j \mid \sigma^2_{\xi,j} \sim N(A_0, \tau_0^{-1} \cdot \sigma^2_{\xi,j})$$

where a_0 , b_0 , A_0 and τ_0 are the prior parameters. Under the nature of conjugate distribution, the updated posterior distribution is also Normal-Inverse Gamma., Conditional on $(A_j, \sigma^2_{\xi,j})$, the probability of K reviewers reporting $(\mathbf{R}_{1j}, \mathbf{R}_{2j}, \cdots, \mathbf{R}_{Kj})$ is:

$$f(\mathbf{R}_{1j}, \cdots, \mathbf{R}_{Kj} \mid A_j, \sigma^2_{\xi,j}) = (2\pi)^{-\frac{K \cdot L}{2}} \mid \tilde{\Omega} \mid^{-\frac{K}{2}} \sigma^{-K}_{\xi,j} \exp[-\frac{1}{2\sigma^2_{\xi,j}} \sum_{k=1}^{K} (\mathbf{R}_{kj} - \mathbf{e}_L \cdot A_j)' \tilde{\Omega}^{-1} (\mathbf{R}_{kj} - \mathbf{e}_L \cdot A_j)]$$ where

$\tilde{\Omega} = \mathbf{1}_L + \lambda^{-1}\Omega$. Based on the assumptions for priors , the prior density of $(A_j, \sigma^2_{\xi,j})$ is:

$$\Pr(A_j, \sigma^2_{\xi,j}) = \frac{\tau_0}{\sqrt{2\pi}\sigma_{\xi,j}} \exp[-\frac{\tau_0}{2\sigma^2_{\xi,j}} (A_j - A_0)^2] \frac{(b_0/2)^{a_0/2}}{\Gamma(a_0/2)} \sigma^{-(a_0+2)}_{\xi,j} \exp(-\frac{b_0}{2\sigma^2_{\xi,j}})$$

According to the Bayes rule, the distribution of the posteriors is:

$$\Pr(A_j, \sigma^2_{\xi,j} \mid \mathbf{R}_{1j}, \mathbf{R}_{2j}, \cdots, \mathbf{R}_{Kj}) \propto \Pr(A_j, \sigma^2_{\xi,j}) \times f(\mathbf{R}_{1j}, \mathbf{R}_{2j}, \cdots, \mathbf{R}_{Kj} \mid A_j, \sigma^2_{\xi,j})$$

Rearranging the equation and remove constant terms, we can derive that:

$$\begin{aligned} &\Pr(A_j, \sigma^2_{\xi,j} \mid \mathbf{R}_{1j}, \mathbf{R}_{2j}, \cdots, \mathbf{R}_{Kj}) \\ &\propto \sigma^{-(K+1+a_0+2)}_{\xi,j} \exp\{-\frac{1}{2\sigma^2_{\xi,j}} [b_0 + \tau_0 (A_j - A_0)^2 + \sum_{k=1}^{K} (\mathbf{R}_{kj} - \mathbf{e}_L \cdot A_j)' \tilde{\Omega}^{-1} (\mathbf{R}_{kj} - \mathbf{e}_L \cdot A_j)]\} \end{aligned}$$

The last term in the above expression can be decomposed as:

$$\begin{aligned} &\sum_{k=1}^{K} (\mathbf{R}_{kj} - \mathbf{e}_L \cdot A_j)' \tilde{\Omega}^{-1} (\mathbf{R}_{kj} - \mathbf{e}_L \cdot A_j) \\ &= \sum_{k=1}^{K} [(\mathbf{R}_{kj} - \mathbf{e}_L \cdot \hat{A}_{jK}) + (\mathbf{e}_L \cdot \hat{A}_{jK} - \mathbf{e}_L \cdot A_j)]' \tilde{\Omega}^{-1} [(\mathbf{R}_{kj} - \mathbf{e}_L \cdot \hat{A}_{jK}) + (\mathbf{e}_L \cdot \hat{A}_{jK} - \mathbf{e}_L \cdot A_j)] \\ &= \sum_{k=1}^{K} (\mathbf{R}_{kj} - \mathbf{e}_L \cdot \hat{A}_{jK})' \tilde{\Omega}^{-1} (\mathbf{R}_{kj} - \mathbf{e}_L \cdot \hat{A}_{jK}) + 2 \sum_{k=1}^{K} (\mathbf{R}_{kj} - \mathbf{e}_L \cdot \hat{A}_{jK})' \tilde{\Omega}^{-1} \mathbf{e}_L \cdot (\hat{A}_{jK} - A_j) + K \cdot \mathbf{e}_L' \tilde{\Omega}^{-1} \mathbf{e}_L \cdot (\hat{A}_{jK} - A_j)^2 \end{aligned}$$

Set $\hat{A}_{jK} = \hat{\delta}_K^{-1} \sum_{k=1}^{K} \mathbf{e}_L' \tilde{\Omega}^{-1} \mathbf{R}_{kj}$ and $\hat{\delta}_K = \sum_{k=1}^{K} \mathbf{e}_L' \tilde{\Omega}^{-1} \mathbf{e}_L$, the second term will cancel out of the equation, because:

$$\begin{aligned} &\sum_{k=1}^{K} \left(\mathbf{R}_{kj} - \mathbf{e}_L \cdot \hat{A}_{jK}\right)' \tilde{\Omega}^{-1} \mathbf{e}_L \cdot \left(\hat{A}_{jK} - A_j\right) \\ &= (\hat{A}_{jK} - A_j) \sum_{k=1}^{K} (\mathbf{R}_{kj} - \mathbf{e}_L \cdot \hat{A}_{jK})' \tilde{\Omega}^{-1} \mathbf{e}_L \\ &= (\hat{A}_{jK} - A_j) [\sum_{k=1}^{K} \mathbf{R}_{kj}' \tilde{\Omega}^{-1} \mathbf{e}_L - \hat{A}_{jK} \sum_{k=1}^{K} \mathbf{e}_L' \tilde{\Omega}^{-1} \mathbf{e}_L] \\ &= (\hat{A}_{jK} - A_j) [\sum_{k=1}^{K} \mathbf{e}_L' \tilde{\Omega}^{-1} \mathbf{R}_{\mathbf{kj}} - \hat{A}_{jK} \hat{\delta}_K] \\ &= 0 \end{aligned}$$

Thus, the previous equation can be re-written as,

$$\begin{aligned}
&\sum\nolimits_{k=1}^{K}(\mathbf{R}_{kj}-\mathbf{e}_L\cdot A_j)'\tilde{\Omega}^{-1}(\mathbf{R}_{kj}-\mathbf{e}_L\cdot A_j)\\
&=\sum\nolimits_{k=1}^{K}(\mathbf{R}_{kj}-\mathbf{e}_L\cdot\hat{A}_{jK})'\tilde{\Omega}^{-1}(\mathbf{R}_{kj}-\mathbf{e}_L\cdot\hat{A}_{jK})+\sum\nolimits_{k=1}^{K}\mathbf{e}_L'\tilde{\Omega}^{-1}\mathbf{e}_L\cdot(\hat{A}_{jK}-A_j)^2\\
&=\hat{\Delta}_{jK}+\hat{\delta}_K(\hat{A}_{jK}-A_j)^2
\end{aligned}$$

where $\hat{\Delta}_{jK}=\sum\nolimits_{k=1}^{K}(\mathbf{R}_{kj}-\mathbf{e}_L\cdot\hat{A}_{jK})'\tilde{\Omega}^{-1}(\mathbf{R}_{kj}-\mathbf{e}_L\cdot\hat{A}_{jK})$. Putting this term back to the posterior density function, we can show that,

$$\begin{aligned}
&\Pr(A_j,\sigma_{\xi,j}^2\,|\,\mathbf{R}_{1j},\mathbf{R}_{2j},\cdots,\mathbf{R}_{Kj})\\
&\propto\sigma_{\xi,j}^{-(K+1+a_0+2)}\exp\{-\frac{1}{2\sigma_{\xi,j}^2}[b_0+\tau_0(A_j-A_0)^2+\hat{\Delta}_{jK}+\hat{\delta}_K(\hat{A}_{jK}-A_j)^2]\}\\
&=\frac{1}{\sigma_{\xi,j}}\exp\{-\frac{1}{2\sigma_{\xi,j}^2}[b_0+(\tau_0+\hat{\delta}_K)(A_j-\frac{\tau_0A_0+\hat{\delta}_K\hat{A}_{jK}}{\tau_0+\hat{\delta}_K})^2+\hat{\Delta}_{jK}+\frac{\tau_0\hat{\delta}_K}{\tau_0+\hat{\delta}_K}(\hat{A}_{jK}-A_0)^2]\}\sigma_{\xi,j}^{-(a_0+K+2)}\\
&\propto\frac{\tau_0+\hat{\delta}_K}{\sqrt{2\pi}\sigma_{\xi,j}}\exp[-\frac{\tau_0+\hat{\delta}_K}{2\sigma_{\xi,j}^2}(A_j-\frac{\tau_0A_0+\hat{\delta}_K\hat{A}_{jK}}{\tau_0+\hat{\delta}_K})^2]\\
&\times\frac{([b_0+\hat{\Delta}_{jK}+\frac{\tau_0\hat{\delta}_K}{\tau_0+\hat{\delta}_K}(\hat{A}_{jK}-A_0)^2]/2)^{(a_0+K)/2}}{\Gamma[(a_0+K)/2]}\sigma_{\xi,j}^{-(a_0+K+2)}\exp\{-\frac{1}{2\sigma_{\xi,j}^2}[b_0+\hat{\Delta}_{jK}+\frac{\tau_0\hat{\delta}_K}{\tau_0+\hat{\delta}_K}(\hat{A}_{jK}-A_0)^2]\}
\end{aligned}$$

which is the product of a Normal density regarding A_j and an Inverse Gamma density about $\sigma_{\xi,j}^2$. Decomposing each parameter component in this Normal-Inverse Gamma distribution, this functional form gives us the updated posterior parameters as

$$\begin{aligned}
&a_{jK}=a_0+K\\
&b_{jK}=b_0+\sum\nolimits_{k=1}^{K}(\mathbf{R}_{kj}-\mathbf{e}_L\cdot\hat{A}_{jK})'\tilde{\Omega}^{-1}(\mathbf{R}_{kj}-\mathbf{e}_L\cdot\hat{A}_{jK})+\frac{\tau_0\hat{\delta}_K}{\tau_0+\hat{\delta}_K}(\hat{A}_{jK}-A_0)^2\\
&A_{jK}=\frac{\tau_0A_0+\hat{\delta}_K\hat{A}_{jK}}{\tau_0+\hat{\delta}_K}\\
&\tau_{jK}=\tau_0+\hat{\delta}_K
\end{aligned}$$

where

$$\hat{\delta}_K = \sum_{k=1}^K \mathbf{e}_L{}' \tilde{\Omega}^{-1} \mathbf{e}_L$$
$$\hat{A}_{jK} = \hat{\lambda}_{jK}^{-1} \sum_{k=1}^K \mathbf{e}_L{}' \tilde{\Omega}^{-1} \mathbf{R}_{kj}$$

In the above equation, a_{jK} and b_{jK} are the posterior shape and posterior scale parameters for the variance distribution. The above posterior density function indicates that, conditional on the information set $\boldsymbol{I}_{Kj} = \{\boldsymbol{R}_{1j}, \ldots, \boldsymbol{R}_{Kj}\}$ from all reviews, the marginal distribution of A_j is a t-distribution with mean $E(A_j \mid I_K) = A_{jK}$ and variance $Var(A_j \mid I_K) = \tau_{jK}^{-1} \cdot \frac{b_{jK}}{a_{jK} - 2}$.

3. Differentiated Learning Model Posterior Distribution Derivation

We assume in Section 3 in the paper that

$$\mathbf{R}_{kj} = (\mathbf{1} - \boldsymbol{\delta}_{ik}) A_0 + \boldsymbol{\delta}_{ik} A_{ij} + \mathbf{u}_{ikj}$$

$$Var(\mathbf{R}_{kj}) = ((1 - \delta_{ik}^2)\tilde{\tau}_0^{-1} \cdot \mathbf{e}_L \mathbf{e}_L{}' + \lambda_k^{-1}\Omega)\sigma_{\xi,j}^2 \equiv \tilde{\Omega}_{ik}\sigma_{\xi,j}^2$$

where $\boldsymbol{\delta}_{ik} \equiv \delta_{ik} \cdot e_L$ and $\mathbf{u}_{ijk} \equiv \sqrt{1-\delta_{ik}^2} e_{ikj} \cdot e_L + \varepsilon_{kj}$. A consumer updates the posterior beliefs of parameters of $(A_{ij}, \sigma_{\xi,j}^2)$ in this learning framework. Similar to the Type I learning model, we use the natural conjugate distribution of Normal-Inverse Gamma as the prior distributions for the two parameters. Specifically, the prior distributions are:

$$\sigma_{\xi,j}^2 \sim IG(\frac{a_0}{2}, \frac{b_0}{2})$$
$$A_{ij} \mid \sigma_{\xi,j}^2 \sim N(A_0, \tilde{\tau}_0^{-1}\sigma_{\xi,j}^2)$$

where a_0, b_0, A_0 and $\tilde{\tau}_0$ are the prior parameters. Under the nature of conjugate distribution, the updated posterior distribution is also Normal-Inverse Gamma. Conditional on $(A_{ij}, \sigma_{\xi,j}^2)$, the probability of K reviewers reporting $(\mathbf{R}_{1j}, \mathbf{R}_{2j}, \cdots, \mathbf{R}_{Kj})$ is:

$$f(\mathbf{R}_{1j}, \mathbf{R}_{2j}, \cdots, \mathbf{R}_{Kj} \mid A_{ij}, \sigma_{\xi,j}^2)$$
$$= (2\pi)^{-\frac{K \cdot L}{2}} \Pi_{k=1}^K |\tilde{\Omega}_{ik}|^{-\frac{1}{2}} \sigma_{\xi,j}^{-K} \exp[-\frac{1}{2\sigma_{\xi,j}^2} \sum_{k=1}^K (\mathbf{R}_{kj} - (1-\boldsymbol{\delta}_{ik})A_0 - \boldsymbol{\delta}_{ik} \cdot A_{ij})' \tilde{\Omega}_{ik}^{-1} (\mathbf{R}_{kj} - (1-\boldsymbol{\delta}_{ik})A_0 - \boldsymbol{\delta}_{ik} \cdot A_{ij})]$$

Based on model assumptions, the prior density of $(A_{ij}, \sigma_{\xi,j}^2)$ is:

$$
\begin{aligned}
&\Pr(A_{ij},\sigma^2_{\xi,j})\\
&=\frac{\tilde{\tau}_0}{\sqrt{2\pi}\sigma_{\xi,j}}\exp[-\frac{\tilde{\tau}_0}{2\sigma^2_{\xi,j}}(A_{ij}-A_0)^2]\frac{(b_0/2)^{a_0/2}}{\Gamma(a_0/2)}\sigma_{\xi,j}^{-(a_0+2)}\exp(-\frac{b_0}{2\sigma^2_{\xi,j}})
\end{aligned}
$$

According to the Bayes rule, we have the distribution of posterior as proportional to the prior multiplied by the sample density:

$$
\Pr(A_{ij},\sigma^2_{\xi,j}\mid \mathbf{R}_{1j},\mathbf{R}_{2j},\cdots,\mathbf{R}_{Kj})\propto \Pr(A_{ij},\sigma^2_{\xi,j})\times f(\mathbf{R}_{1j},\mathbf{R}_{2j},\cdots,\mathbf{R}_{Kj}\mid A_{ij},\sigma^2_{\xi,j})
$$

Rearranging the equation and remove constant terms, we can derive that:

$$
\begin{aligned}
&\Pr(A_{ij},\sigma^2_{\xi,j}\mid \mathbf{R}_{1j},\mathbf{R}_{2j},\cdots,\mathbf{R}_{Kj})\\
&\propto \sigma_{\xi,j}^{-(K+1+a_0+2)}\exp\{-\frac{1}{2\sigma^2_{\xi,j}}[b_0+\tilde{\tau}_0(A_{ij}-A_0)^2\\
&\quad+\textstyle\sum_{k=1}^{K}(\mathbf{R}_{kj}-(1-\boldsymbol{\delta}_{ik})A_0-\boldsymbol{\delta}_{ik}\cdot A_{ij})'\tilde{\Omega}_{ik}^{-1}(\mathbf{R}_{kj}-(1-\boldsymbol{\delta}_{ik})A_0-\boldsymbol{\delta}_{ik}\cdot A_{ij})]\}
\end{aligned}
$$

The last term in the above expression can be decomposed as:

$$
\begin{aligned}
&\sum_{k=1}^{K}(\mathbf{R}_{kj}-(1-\boldsymbol{\delta}_{ik})A_0-\boldsymbol{\delta}_{ik}\cdot A_{ij})'\tilde{\Omega}_{ik}^{-1}(\mathbf{R}_{kj}-(1-\boldsymbol{\delta}_{ik})A_0-\boldsymbol{\delta}_{ik}\cdot A_{ij})\\
&=\sum_{k=1}^{K}[(\mathbf{R}_{kj}-(1-\boldsymbol{\delta}_{ik})A_0-\boldsymbol{\delta}_{ik}\cdot \hat{A}_{iK})+\boldsymbol{\delta}_{ik}\cdot(\hat{A}_{iK}-A_{ij})]'\tilde{\Omega}_{ik}^{-1}[(\mathbf{R}_{kj}-(1-\boldsymbol{\delta}_{ik})A_0-\boldsymbol{\delta}_{ik}\cdot \hat{A}_{iK})+\boldsymbol{\delta}_{ik}\cdot(\hat{A}_{iK}-A_{ij})]\\
&=\sum_{k=1}^{K}(\mathbf{R}_{kj}-(1-\boldsymbol{\delta}_{ik})A_0-\boldsymbol{\delta}_{ik}\cdot \hat{A}_{iK})'\tilde{\Omega}_{ik}^{-1}(\mathbf{R}_{kj}-(1-\boldsymbol{\delta}_{ik})A_0-\boldsymbol{\delta}_{ik}\cdot \hat{A}_{iK})\\
&\quad+2\textstyle\sum_{k=1}^{K}(\mathbf{R}_{kj}-(1-\boldsymbol{\delta}_{ik})A_0-\boldsymbol{\delta}_{ik}\cdot \hat{A}_{iK})'\tilde{\Omega}_{ik}^{-1}\boldsymbol{\delta}_{ik}\cdot(\hat{A}_{iK}-A_{ij})+\sum_{k=1}^{K}\boldsymbol{\delta}'_{ik}\tilde{\Omega}_{ik}^{-1}\boldsymbol{\delta}_{ik}\cdot(\hat{A}_{iK}-A_{ij})^2
\end{aligned}
$$

Set $\hat{A}_{ijK}=\hat{\delta}_{iK}^{-1}\sum_{k=1}^{K}\boldsymbol{\delta}_{ik}{}'\tilde{\Omega}_{ik}^{-1}(\mathbf{R}_{kj}-(1-\boldsymbol{\delta}_{ik})A_0)$ and $\hat{\delta}_{iK}=\sum_{k=1}^{K}\boldsymbol{\delta}'_{ik}\tilde{\Omega}_{ik}^{-1}\boldsymbol{\delta}'_{ik}$, the second term will cancel out of the equation, because:

$$
\begin{aligned}
&\textstyle\sum_{k=1}^{K}(\mathbf{R}_{kj}-(1-\boldsymbol{\delta}_{ik})A_0-\boldsymbol{\delta}_{ik}\cdot \hat{A}_{ijK})'\tilde{\Omega}_{ik}^{-1}\boldsymbol{\delta}_{ik}\cdot(\hat{A}_{ijK}-A_{ij})\\
&=(\hat{A}_{ijK}-A_{ij})\textstyle\sum_{k=1}^{K}(\mathbf{R}_{kj}-(1-\boldsymbol{\delta}_{ik})A_0-\boldsymbol{\delta}_{ik}\cdot \hat{A}_{ijK})'\tilde{\Omega}_{ik}^{-1}\boldsymbol{\delta}_{ik}\\
&=(\hat{A}_{ijK}-A_{ij})[\textstyle\sum_{k=1}^{K}(\mathbf{R}_{kj}-(1-\boldsymbol{\delta}_{ik})A_0)'\tilde{\Omega}_{ik}^{-1}\boldsymbol{\delta}_{ik}-\hat{A}_{ijK}\cdot\sum_{k=1}^{K}\boldsymbol{\delta}'_{ik}\tilde{\Omega}_{ik}^{-1}\boldsymbol{\delta}_{ik}]\\
&=(\hat{A}_{jK}-A_{j})[\textstyle\sum_{k=1}^{K}\boldsymbol{\delta}'_{ik}\tilde{\Omega}_{ik}^{-1}(\mathbf{R}_{\mathbf{kj}}-(1-\boldsymbol{\delta}_{ik})A_0)-\hat{A}_{ijK}\hat{\delta}_{iK}]\\
&=0
\end{aligned}
$$

Thus, the previous equation can be re-written as,

$$
\begin{aligned}
&\sum_{k=1}^{K}(\mathbf{R}_{kj}-(1-\boldsymbol{\delta}_{ik})A_0-\boldsymbol{\delta}_{ik}\cdot A_{ij})'\tilde{\Omega}_{ik}^{-1}(\mathbf{R}_{kj}-(1-\boldsymbol{\delta}_{ik})A_0-\boldsymbol{\delta}_{ik}\cdot A_{ij})\\
&=\sum_{k=1}^{K}(\mathbf{R}_{kj}-(1-\boldsymbol{\delta}_{ik})A_0-\boldsymbol{\delta}_{ik}\cdot \hat{A}_{ijK})'\tilde{\Omega}_{ik}^{-1}(\mathbf{R}_{kj}-(1-\boldsymbol{\delta}_{ik})A_0-\boldsymbol{\delta}_{ik}\cdot \hat{A}_{ijK})+\sum_{k=1}^{K}\boldsymbol{\delta}_{ik}'\tilde{\Omega}_{ik}^{-1}\boldsymbol{\delta}_{ik}\cdot(\hat{A}_{ijK}-A_{ij})^2\\
&=\hat{\Delta}_{ijK}+\hat{\delta}_{iK}(\hat{A}_{ijK}-A_{ij})^2
\end{aligned}
$$

where

$$\hat{\Delta}_{ijK}=\sum_{k=1}^{K}(\mathbf{R}_{kj}-(1-\boldsymbol{\delta}_{ik})A_0-\boldsymbol{\delta}_{ik}\cdot \hat{A}_{ijK})'\tilde{\Omega}_{ik}^{-1}(\mathbf{R}_{kj}-(1-\boldsymbol{\delta}_{ik})A_0-\boldsymbol{\delta}_{ik}\cdot \hat{A}_{ijK})$$

Putting this term back to the posterior density function, we can show that it is proportional to the following,

$$
\begin{aligned}
&\Pr(A_{ij},\sigma_{\xi,j}^2\mid \mathbf{R}_{1j},\mathbf{R}_{2j},\cdots,\mathbf{R}_{Kj})\\
&\propto \sigma_{e,j}^{-(K+1+a_0+2)}\exp\{-\frac{1}{2\sigma_{\xi,j}^2}[b_0+\tilde{\tau}_0(A_{ij}-A_0)^2+\hat{\Delta}_{ijK}+\hat{\delta}_{iK}(\hat{A}_{ijK}-A_{ij})^2]\}\\
&=\frac{1}{\sigma_{\xi,j}}\exp\{-\frac{1}{2\sigma_{\xi,j}^2}[b_0+(\tilde{\tau}_0+\hat{\delta}_{iK})(A_{ij}-\frac{\tilde{\tau}_0A_0+\hat{\delta}_{iK}\hat{A}_{ijK}}{\tilde{\tau}_0+\hat{\delta}_{iK}})^2+\hat{\Delta}_{ijK}+\frac{\tilde{\tau}_0\hat{\delta}_{iK}}{\tilde{\tau}_0+\hat{\delta}_{iK}}(\hat{A}_{ijK}-A_0)^2]\}\sigma_{\xi,j}^{-(a_{i0}+K+2)}\\
&\propto \frac{\tilde{\tau}_0+\hat{\delta}_{iK}}{\sqrt{2\pi}\sigma_{\xi,j}}\exp[-\frac{\tilde{\tau}_0+\hat{\delta}_{iK}}{2\sigma_{\xi,j}^2}(A_{ij}-\frac{\tilde{\tau}_0A_0+\hat{\delta}_{iK}\hat{A}_{ijK}}{\tilde{\tau}_0+\hat{\delta}_{iK}})^2]\\
&\times\frac{([b_{i0}+\hat{\Delta}_{jiK}+\frac{\tilde{\tau}_0\hat{\delta}_{iK}}{\tilde{\tau}_0+\hat{\delta}_{iK}}(\hat{A}_{ijK}-A_0)^2]/2)^{(a_0+K)/2}}{\Gamma[(a_0+K)/2]}\sigma_{\xi,j}^{-(a_0+K+2)}\exp\{-\frac{1}{2\sigma_{\xi,j}^2}[b_0+\hat{\Delta}_{ijK}+\frac{\tilde{\tau}_0\hat{\delta}_{iK}}{\tilde{\tau}_0+\hat{\delta}_{iK}}(\hat{A}_{ijK}-A_0)^2]\}
\end{aligned}
$$

which is the product of a Normal density regarding A_{ij} and an Inverse Gamma density about $\sigma_{\xi,j}^2$. Decomposing each parameter component in this Normal-Inverse Gamma distribution, we have the following learning parameters:

$$
\begin{aligned}
&a_{ijK}=a_{i0}+K\\
&b_{ijK}=b_{i0}+\sum_{k=1}^{K}(\mathbf{R}_{kj}-(1-\boldsymbol{\delta}_{ik})A_0-\boldsymbol{\delta}_{ik}\cdot \hat{A}_{ijK})'\tilde{\Omega}_{ik}^{-1}(\mathbf{R}_{kj}-(1-\boldsymbol{\delta}_{ik})A_0-\boldsymbol{\delta}_{ik}\cdot \hat{A}_{ijK})+\frac{\tilde{\tau}_0\hat{\delta}_{iK}}{\tilde{\tau}_0+\hat{\delta}_{iK}}(\hat{A}_{ijK}-A_0)^2\\
&A_{ijK}=\frac{\tilde{\tau}_0A_0+\hat{\delta}_{iK}\hat{A}_{ijK}}{\tilde{\tau}_0+\hat{\delta}_{iK}}\\
&\tau_{ijK}=\tilde{\tau}_0+\hat{\delta}_{iK}
\end{aligned}
$$

where

$$\hat{\delta}_{iK} = \sum_{k=1}^{K} \boldsymbol{\delta}'_{ik} \tilde{\Omega}_{ik}^{-1} \boldsymbol{\delta}_{ik}$$
$$\hat{A}_{ijK} = \hat{\delta}_{iK}^{-1} \sum_{k=1}^{K} \boldsymbol{\delta}'_{ik} \tilde{\Omega}_{ik}^{-1} (\mathbf{R}_{kj} - (1 - \boldsymbol{\delta}_{ik}) A_0)$$

Based on the above posterior density function, the marginal distribution of A_{ij}, given the information set I_K, is a *t*-distribution with mean and variance as:

$$E(A_{ij} \mid I_K) = A_{ijK}$$
$$Var(A_{ij} \mid I_K) = \frac{1}{\tilde{\tau}_{ijK}} \cdot \frac{b_{ijK}}{a_{ijK} - 2}$$