Random effect shrinkage

The following simple example illustrates (from a Bayesian perspective) why and how random effects are shrunk to a common value.
Suppose that $y_1, \ldots, y_n$ satisfy

$$y_j \mid \theta_j \overset{\text{ind}}{\sim} N(\theta_j, v_j), \quad \theta_1, \ldots, \theta_n \mid \mu \overset{\text{iid}}{\sim} N(\mu, \sigma^2), \quad \mu \sim N(\mu_0, \tau^2),$$

where $v_1, \ldots, v_n$, σ^2, μ_0 and τ^2 are assumed known here. Then, the usual posterior calculations give us

$$\mathrm{E}(\mu \mid y) = \frac{\mu_0/\tau^2 + \sum y_j/(\sigma^2 + v_j)}{1/\tau^2 + \sum 1/(\sigma^2 + v_j)}, \ \mathrm{var}(\mu \mid y) = \frac{1}{1/\tau^2 + \sum 1/(\sigma^2 + v_j)},$$

and

$$\mathrm{E}(\theta_j \mid y) = (1 - w)\mathrm{E}(\mu \mid y) + w y_j,$$

where

$$w = \frac{\sigma^2}{\sigma^2 + v_j}.$$