Appendix

Proof of Theorem 1:

The multivariate Gaussian probability density function for random vector $\mathbf{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_N)^T$ is:

$p(\mathbf{x}) = \frac{1}{(2\pi)^{N/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)$. The marginal distribution of X_i is Gaussian with mean μ_i and variance equal to the i^{th} diagonal term of $\boldsymbol{\Sigma}$, denoted σ_i^2.

The decision maker is risk-neutral and therefore s/he maximizes the total expected profits. A priori, the decision maker chooses the projects with positive expected value. If we denote the set of a priori profitable projects as $\mathbb{P}$, then the prior value $PV = \sum_i \max(0, \mu_i) = \sum_{i \in \mathbb{P}} \mu_i$.

Suppose there is an information source $\mathbf{Y}$ about the projects; $\mathbf{Y} = \mathbf{X} + \mathbf{E}$, $\mathbf{E} \sim N(\mathbf{0}, \mathbf{T})$. Let $\mathbb{K}$ denote a subset of the projects. If the decision maker is able to observe information $\mathbf{Y}_{\mathbb{K}}$ about the subset (for free), then the value with this information is:

$$\int_{\mathbf{y}_{\mathbb{K}}} \sum_{i=1}^{N} \left[\max\left(E(\mathrm{X}_i \mid \mathbf{y}_{\mathbb{K}}), 0 \right) \right] p(\mathbf{y}_{\mathbb{K}}) \, d\mathbf{y}_{\mathbb{K}} = \sum_{i=1}^{N} \left[\int_{\mathbf{y}_{\mathbb{K}}} \max\left(E(\mathrm{X}_i \mid \mathbf{y}_{\mathbb{K}}), 0 \right) p(\mathbf{y}_{\mathbb{K}}) \, d\mathbf{y}_{\mathbb{K}} \right].$$

From the properties of the multivariate Gaussian distribution, see e.g. Anderson (2003),

$E[\mathrm{X}_i \mid \mathbf{y}_{\mathbb{K}}] = \mu_i + \boldsymbol{\Sigma}_{i\mathbb{K}} (\boldsymbol{\Sigma}_{\mathbb{K}} + \mathbf{T}_{\mathbb{K}})^{-1} (\mathbf{y}_{\mathbb{K}} - \boldsymbol{\mu}_{\mathbb{K}})$. Let us denote $E[\mathrm{X}_i \mid \mathbf{y}_{\mathbb{K}}]$ as W_i.

W_i is a linear combination of Gaussian random variables and therefore has marginal distribution $p(\mathrm{w}_i) = N(\mu_i, s_i^2)$, where $s_i^2 = \boldsymbol{\Sigma}_{i\mathbb{K}} (\boldsymbol{\Sigma}_{\mathbb{K}} + \mathbf{T}_{\mathbb{K}})^{-1} \boldsymbol{\Sigma}_{\mathbb{K}i}$.

This linear combination is the uni-variate variable of interest in the multi-dimensional integral required for the value of information computation. Marginalizing out the remaining dimensions of information, the value with information $\mathbf{Y}_{\mathbb{K}}$ can be simplified to: $\sum_{i=1}^{N}\left[\int_{\mathrm{w}_i}\max\left(\mathrm{w}_i,0\right)p(\mathrm{w}_i)d\mathrm{w}_i\right]=\sum_{i=1}^{N}E\left[\max\left\{\mathrm{W}_i,0\right\}\right]$.

The following result is from Schlaiffer (1959) and Bickel (2008):

For a Gaussian variable W with mean m and variance s^2, $E\left[\max\left\{\mathrm{W},0\right\}\right]=m\Phi(m/s)+s\phi\left(m/s\right)$.

Using this result and subtracting the prior value from the value with information:

$$
\begin{aligned}
\mathrm{VOI}(\mathbf{Y}_{\mathbb{K}}) &= \sum_{i=1}^{N}\left[\mu_i\Phi\left(\mu_i/s_i\right)+s_i\phi\left(\mu_i/s_i\right)\right]-\sum_{i\in\mathbb{P}}\mu_i \\
&= \sum_{i\in\mathbb{P}}\left[\mu_i\left(\Phi\left(\mu_i/s_i\right)-1\right)+s_i\phi\left(\mu_i/s_i\right)\right]+\sum_{i\in\mathbb{P}}\left[\mu_i\Phi\left(\mu_i/s_i\right)+s_i\phi\left(\mu_i/s_i\right)\right].
\end{aligned}
$$

Since $1-\Phi\left(\mu_i/\sigma_i\right)=\Phi\left(-\mu_i/\sigma_i\right)$, we get the required result. □

Proof of Corollary 1:

For TVOI, there is information regarding all projects, therefore $\mathbb{K}=\mathbb{N}$.

The result follows by substituting $\mathbf{\Sigma}_{\mathbb{K}}=\mathbf{\Sigma}$ and $\mathbf{T}_{\mathbb{K}}=\mathbf{T}$, and recognizing that $s_i^2=e_i^{\,T}\mathbf{\Sigma}\left(\mathbf{\Sigma}+\mathbf{T}\right)^{-1}\mathbf{\Sigma}e_i$. □

Proof of Theorem 2:

The proof of Theorem 1 shows that the $\mathrm{VOI}(\mathbf{Y}_{\mathbb{K}})$ can be written as:

$$\mathrm{VOI}(\mathbf{Y}_{\mathbb{K}})=\sum_{i=1}^{N}\left[\mu_i\Phi\left(\mu_i/s_i\right)+s_i\phi\left(\mu_i/s_i\right)\right]-\sum_{i\in\mathbb{P}}\mu_i\ .$$

Consider again a Gaussian variable W with mean m and variance s^2. We will find partial derivatives of the expression $m\Phi(m/s)+s\phi\left(m/s\right)$ with respect to any parameter θ.

$$\frac{d\left[m\Phi\left(m/s\right)+s\phi\left(m/s\right)\right]}{d\theta}=\frac{dm}{d\theta}\Phi\left(m/s\right)+m\frac{d\Phi\left(m/s\right)}{dm/s}\frac{dm/s}{d\theta}+\frac{ds}{d\theta}\phi\left(m/s\right)+s\frac{d\phi\left(m/s\right)}{dm/s}\frac{dm/s}{d\theta}.$$

Since $\frac{d\Phi(x)}{dx}=\phi(x)$ and $\frac{d\phi(x)}{dx}=\frac{d}{dx}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)=-x\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)=-x\phi(x)$,

the derivative simplifies to:

$$\frac{d\left[m\Phi\left(m/s\right)+s\phi\left(m/s\right)\right]}{d\theta}=\frac{dm}{d\theta}\Phi\left(m/s\right)+m\phi\left(m/s\right)\frac{dm/s}{d\theta}+\frac{ds}{d\theta}\phi\left(m/s\right)-s\left(\frac{m}{s}\right)\phi\left(m/s\right)\frac{dm/s}{d\theta}$$
$$=\frac{dm}{d\theta}\Phi\left(m/s\right)+\frac{ds}{d\theta}\phi\left(m/s\right)$$

- Sensitivity to the mean: If $\theta=m$, $\frac{d\left[m\Phi\left(m/s\right)+s\phi\left(m/s\right)\right]}{dm}=\Phi\left(m/s\right)$.
- Sensitivity to the standard deviation: If $\theta=s$, $\frac{d\left[m\Phi\left(m/s\right)+s\phi\left(m/s\right)\right]}{ds}=\phi\left(m/s\right)$.

(i) Sensitivity to μ_i:

From the result above (sensitivity to the mean), the derivative with respect to the mean of any project is:

$$\frac{d\mathrm{VOI}(\mathbf{Y}_{\mathbb{K}})}{d\mu_i}=\frac{d\left[\sum_{i=1}^{N}\left[\mu_i\Phi\left(\mu_i/s_i\right)+s_i\phi\left(\mu_i/s_i\right)\right]-\sum_{i\in\mathbb{P}}\mu_i\right]}{d\mu_i}=\left\{\begin{array}{lll}\Phi\left(\mu_i/s_i\right)-1, & i\in\mathbb{P}, & \mu_i>0\\ \Phi\left(\mu_i/s_i\right) & i\notin\mathbb{P} & \mu_i\le 0\end{array}\right\}.$$

$\mathrm{VOI}(\mathbf{Y}_{\mathbb{K}})$ is maximum at $\mu_i=0$ and decreases as the mean gets further from 0.

(ii) Sensitivity to τ_i:

The measurement noise τ_i affects the equivalent standard deviations s_i but does not affect any mean parameter μ_i. For a parameter θ that only affects the equivalent standard deviations,

$$\frac{d\mathrm{VOI}(\mathbf{Y}_{\mathbb{K}})}{d\theta} = \frac{d\left[\sum_{i=1}^{N}\left[\mu_i\Phi\left(\mu_i/s_i\right)+s_i\phi\left(\mu_i/s_i\right)\right]-\sum_{i\in\mathbb{P}}\mu_i\right]}{d\theta} = \sum_{i=1}^{N}\frac{1}{2s_i}\frac{ds_i^{\,2}}{d\theta}\phi\left(\mu_i/s_i\right).$$

For the last step in the computation above, we have used the following result, which we proved before part (i) (sensitivity to the standard deviation): $\frac{d\left[\mu_i\Phi\left(\mu_i/s_i\right)+s_i\phi\left(\mu_i/s_i\right)\right]}{ds_i} = \phi\left(\mu_i/s_i\right)$.

For the remainder of the proof, we will compute the derivative of the equivalent variance with respect to q, i.e. $\frac{ds_i^{\,2}}{d\theta}$, and then replace it back in the expression $\sum_{i=1}^{N}\frac{1}{2s_i}\frac{ds_i^{\,2}}{d\theta}\phi\left(\mu_i/s_i\right)$.

The derivative of the inverse of an invertible matrix $\mathbf{M}$ is $\frac{d\mathbf{M}^{-1}}{d\theta} = -\mathbf{M}^{-1}\frac{d\mathbf{M}}{d\theta}\mathbf{M}^{-1}$, where the matrix derivative works for every entry of the matrix. Here, $s_i^2 = \mathbf{\Sigma}_{i\mathbb{K}}\left(\mathbf{\Sigma}_{\mathbb{K}}+\mathbf{T}_{\mathbb{K}}\right)^{-1}\mathbf{\Sigma}_{\mathbb{K}i}$, therefore, for some parameter θ in $\mathbf{T}_{\mathbb{K}}$, $\frac{ds_i^2}{d\theta} = \frac{d\left(\mathbf{\Sigma}_{i\mathbb{K}}\left(\mathbf{\Sigma}_{\mathbb{K}}+\mathbf{T}_{\mathbb{K}}\right)^{-1}\mathbf{\Sigma}_{\mathbb{K}i}\right)}{d\theta} = -\mathbf{\Sigma}_{i\mathbb{K}}\left(\mathbf{\Sigma}_{\mathbb{K}}+\mathbf{T}_{\mathbb{K}}\right)^{-1}\frac{d\mathbf{T}_{\mathbb{K}}}{d\theta}\left(\mathbf{\Sigma}_{\mathbb{K}}+\mathbf{T}_{\mathbb{K}}\right)^{-1}\mathbf{\Sigma}_{\mathbb{K}i}$.

If $q = t_i^2$ for $i \in \mathbb{K}$, then $\frac{d\mathbf{T}_{\mathbb{K}}}{d\theta}$ is a matrix of 0s except the i^th diagonal element, which equals 1. Matrices $\mathbf{\Sigma}_{\mathbb{K}}$ and $\mathbf{\Sigma}_{\mathbb{K}}+\mathbf{T}_{\mathbb{K}}$ are both symmetric. $\left(\mathbf{\Sigma}_{\mathbb{K}}+\mathbf{T}_{\mathbb{K}}\right)^{-1}$ is also symmetric since matrix inversion retains the symmetric property. Thus, by replacing $\mathbf{b} = \left(\mathbf{\Sigma}_{\mathbb{K}}+\mathbf{T}_{\mathbb{K}}\right)^{-1}\mathbf{\Sigma}_{\mathbb{K}}\boldsymbol{e}_i$, $\frac{ds_i^2}{d\theta} = -\mathbf{b}^T\frac{d\mathbf{T}_{\mathbb{K}}}{d\tau_i^2}\mathbf{b} = -b_i^2 < 0$.

In summary, we get negative effects and the derivative is:

$$\frac{d\mathrm{VOI}(\mathbf{Y}_{\mathbb{K}})}{d\tau_i^2} = -\sum_i\left[\frac{1}{2s_i}\phi\left(\mu_i/s_i\right)b_i^2\right].$$

(iii) Sensitivity to σ_i :

The prior standard deviation σ_i affects the equivalent standard deviations s_i but does not affect any mean parameter μ_i. Similar to (ii), for a parameter θ that only affects the equivalent standard deviations,

$$\frac{d\mathrm{VOI}(\mathbf{Y}_{\mathbb{K}})}{d\theta} = \frac{d\left[\sum_{i=1}^{N}\left[\mu_i\Phi\left(\mu_i/s_i\right) + s_i\phi\left(\mu_i/s_i\right)\right] - \sum_{i\in\mathbb{P}}\mu_i\right]}{d\theta} = \sum_{i=1}^{N}\frac{1}{2s_i}\frac{ds_i^{\,2}}{d\theta}\phi\left(\mu_i/s_i\right).$$

Consider first the special case of TVOI. The derivative of s_i^2 is again the crucial part. The following matrix identities hold by the Sherman-Woodbury-Morrison formula, see e.g. Henderson and Searle (1981): $\left(\mathbf{\Sigma}+\mathbf{T}\right)^{-1} = \mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\left(\mathbf{\Sigma}^{-1}+\mathbf{T}^{-1}\right)^{-1}\mathbf{\Sigma}^{-1}$ and $\left(\mathbf{\Sigma}^{-1}+\mathbf{T}^{-1}\right)^{-1} = \mathbf{T} - \mathbf{T}\left(\mathbf{\Sigma}+\mathbf{T}\right)^{-1}\mathbf{T}$.

Therefore, $\boldsymbol{S} = \mathbf{\Sigma}\left(\mathbf{\Sigma}+\mathbf{T}\right)^{-1}\mathbf{\Sigma} = \mathbf{\Sigma} - \mathbf{T} + \mathbf{T}\left(\mathbf{\Sigma}+\mathbf{T}\right)^{-1}\mathbf{T}$. For the derivative of s_i^2 :

$$\frac{d\boldsymbol{e}_i^T\left(\mathbf{\Sigma}-\mathbf{T}+\mathbf{T}\left(\mathbf{\Sigma}+\mathbf{T}\right)^{-1}\mathbf{T}\right)\boldsymbol{e}_i}{d\theta} = \boldsymbol{e}_i^T\frac{d\mathbf{\Sigma}}{d\theta}\boldsymbol{e}_i - \boldsymbol{e}_i^T\mathbf{T}\left(\mathbf{\Sigma}+\mathbf{T}\right)^{-1}\frac{d\mathbf{\Sigma}}{d\theta}\left(\mathbf{\Sigma}+\mathbf{T}\right)^{-1}\mathbf{T}\boldsymbol{e}_i.$$

If $q = s_i^{\,2}$, $\frac{d\mathbf{\Sigma}}{d\theta}$ is a matrix which is nonzero only for the elements in row i and column i. Defining $\boldsymbol{B} = \left(\mathbf{\Sigma}+\mathbf{T}\right)^{-1}\frac{d\mathbf{\Sigma}}{d\sigma_i^2}\left(\mathbf{\Sigma}+\mathbf{T}\right)^{-1}$, which is non-negative definite, the above expression can be written as:

$$\frac{d\boldsymbol{e}_i^T\left(\mathbf{\Sigma}-\mathbf{T}+\mathbf{T}\left(\mathbf{\Sigma}+\mathbf{T}\right)^{-1}\mathbf{T}\right)\boldsymbol{e}_i}{d\sigma_i^2} = \boldsymbol{e}_i^T\left(\mathbf{\Sigma}+\mathbf{T}\right)\boldsymbol{B}\left(\mathbf{\Sigma}+\mathbf{T}\right)\boldsymbol{e}_i - \boldsymbol{e}_i^T\mathbf{T}\boldsymbol{B}\mathbf{T}\boldsymbol{e}_i.$$

Since both $\mathbf{\Sigma}$ and $\mathbf{T}$ are positive definite, this derivative will be positive, and TVOI increases as a function of the prior uncertainty.

The case with PVOI can be viewed as a special case of total information where some t_is go to infinity. The same proof therefore holds.