# AI, Headquarter and Guijie

August 14, 2025

## 1 Model

There is a principal $P$ (say, a headquarter) and an agent $A$ (a local manager/"guijie").

The principal randomly chooses one experiment which needs the agent to conduct. The quality of the experiment is $q \sim N(0, \sigma_P^2)$.

The agent's type is $\theta \sim N(0, \sigma_A^2)$, which is independent of $q$.

Let $z$ be the outcome of the experiment. The data generation process of $z$ is $z = q + \theta$.

The principal observes an outcome of the experiment $\tilde{z}$. She cannot observe the actual quality of the experiment $q$, nor the agent's type $\theta$.

## 2 Analysis

### 2.1 Estimation of the Uninformed Principal

After observing $\tilde{z}$, the principal estimates the experiment quality $q$ and the agent's type $\theta$. Formally, the posterior distributions of $q$ and $\theta$, conditional on $z = \tilde{z}$, are

$$q \mid z = \tilde{z} \sim N\left(\frac{\sigma_P^2}{\sigma_P^2 + \sigma_A^2}\tilde{z}, \frac{\sigma_A^2 \sigma_P^2}{\sigma_A^2 + \sigma_P^2}\right),$$

and

$$\theta \mid z = \tilde{z} \sim N\left(\frac{\sigma_A^2}{\sigma_P^2 + \sigma_A^2}\tilde{z}, \frac{\sigma_A^2 \sigma_P^2}{\sigma_A^2 + \sigma_P^2}\right).$$

As a result, the best point estimates are $\hat{q} = \frac{\sigma_P^2}{\sigma_P^2+\sigma_A^2}\tilde{z}$ and $\hat{\theta} = \frac{\sigma_A^2}{\sigma_P^2+\sigma_A^2}\tilde{z}$.

Suppose that $\tilde{z} < 0$, which we interpret it as a bad experiment outcome. If $\sigma_A^2 \gg \sigma_P^2$, then $\hat{q} \approx 0$ and $\hat{\theta} \approx \tilde{z}$. This means that the uninformed principal would almost attribute the bad outcome all to the agent's type.

Consider another extreme case where $\tilde{z} < 0$ and AI enables the principal to perfectly observe the agent's type. Then the uninformed principal will understand that the bad outcome is due to the experiment quality.

## 2.2 Discussion

1. If $\tilde{z} > 0$ and $\sigma_A^2 \gg \sigma_P^2$, then without AI, the uninformed principal will attribute the good result to having had a good agent, which is a bit weird to me. It would be fantastic to know whether you share the same view, or find it natural enough.

2. Instead of focusing on an agent's type, I can model the agent's action which will then involve strategic behavior of the agent, but I am not sure whether it's a good thing to do at present. Could you please advise what your choice would be if you were developing this model?

3. What is your general opinion on the current model?