Signal Detection Theory (SDT)

Joachim Vandekerckhove

Winter 2025

SDT: Core Idea

How do we make categorical judgments based on ambiguous or noisy information?

- Radiologist detecting a tumor
- Witness identifying a suspect
- Participant hearing a faint tone

SDT: Core Idea

How do we make categorical judgments based on ambiguous or noisy information?

- Radiologist detecting a tumor
- Witness identifying a suspect
- Participant hearing a faint tone

The challenge: Distinguishing target signal from background noise.

SDT: Core Idea

How do we make categorical judgments based on ambiguous or noisy information?

- Radiologist detecting a tumor
- Witness identifying a suspect
- Participant hearing a faint tone

The challenge: Distinguishing target signal from background noise. SDT formalizes the decision process into two components:

1. **Sensitivity:** How well can the system separate signal from noise? (Information quality)
2. **Bias/Criterion:** What decision rule is used? (Strategy, goals, expectations)

SDT: Core Idea

How do we make categorical judgments based on ambiguous or noisy information?

- Radiologist detecting a tumor
- Witness identifying a suspect
- Participant hearing a faint tone

The challenge: Distinguishing target signal from background noise. SDT formalizes the decision process into two components:

1. **Sensitivity:** How well can the system separate signal from noise? (Information quality)
2. **Bias/Criterion:** What decision rule is used? (Strategy, goals, expectations)

Allows us to independently quantify sensitivity and bias, unlike simple accuracy.

SDT: History & Significance

Developed in 1950s/60s (Green, Swets, Tanner, Egan, ...).

SDT: History & Significance

Developed in 1950s/60s (Green, Swets, Tanner, Egan, ...). Shift from classical threshold theories (Fechner, Thurstone) which assumed fixed sensory thresholds.

SDT: History & Significance

Developed in 1950s/60s (Green, Swets, Tanner, Egan, ...). Shift from classical threshold theories (Fechner, Thurstone) which assumed fixed sensory thresholds.

SDT introduced a probabilistic view:

- **There is always noise:** Internal and external random fluctuations.
- **Decision is probabilistic:** Based on internal evidence (a random variable).
- **Strategy matters:** Observers adopt criteria based on task demands (payoffs, instructions).

SDT: History & Significance

Developed in 1950s/60s (Green, Swets, Tanner, Egan, ...). Shift from classical threshold theories (Fechner, Thurstone) which assumed fixed sensory thresholds.

SDT introduced a probabilistic view:

- **There is always noise:** Internal and external random fluctuations.
- **Decision is probabilistic:** Based on internal evidence (a random variable).
- **Strategy matters:** Observers adopt criteria based on task demands (payoffs, instructions).

Paradigm shift: From absolute thresholds to decision processes under uncertainty.

SDT: The Standard Model

Common assumptions:

1. **Gaussian Distributions:** Internal evidence for Noise (N) and Signal+Noise (S+N) follows Normal distributions.

$$\text{Evidence}|\text{Noise} \sim \mathcal{N}(\mu_N, \sigma_N^2)$$

$$\text{Evidence}|\text{Signal} \sim \mathcal{N}(\mu_{S+N}, \sigma_{S+N}^2)$$

SDT: The Standard Model

Common assumptions:

1. **Gaussian Distributions:** Internal evidence for Noise (N) and Signal+Noise (S+N) follows Normal distributions.

   $$\text{Evidence}|\text{Noise} \sim \mathcal{N}(\mu_N, \sigma_N^2)$$

   $$\text{Evidence}|\text{Signal} \sim \mathcal{N}(\mu_{S+N}, \sigma_{S+N}^2)$$

2. **Equal Variances:** $\sigma_N = \sigma_{S+N} = \sigma$.

SDT: The Standard Model

Common assumptions:

1. **Gaussian Distributions:** Internal evidence for Noise (N) and Signal+Noise (S+N) follows Normal distributions.

   $$\text{Evidence}|\text{Noise} \sim \mathcal{N}(\mu_N, \sigma_N^2)$$

   $$\text{Evidence}|\text{Signal} \sim \mathcal{N}(\mu_{S+N}, \sigma_{S+N}^2)$$

2. **Equal Variances:** $\sigma_N = \sigma_{S+N} = \sigma$.
3. **Additivity:** Signal shifts the mean ($\mu_{S+N} > \mu_N$) but not the shape.

SDT: The Standard Model

Common assumptions:

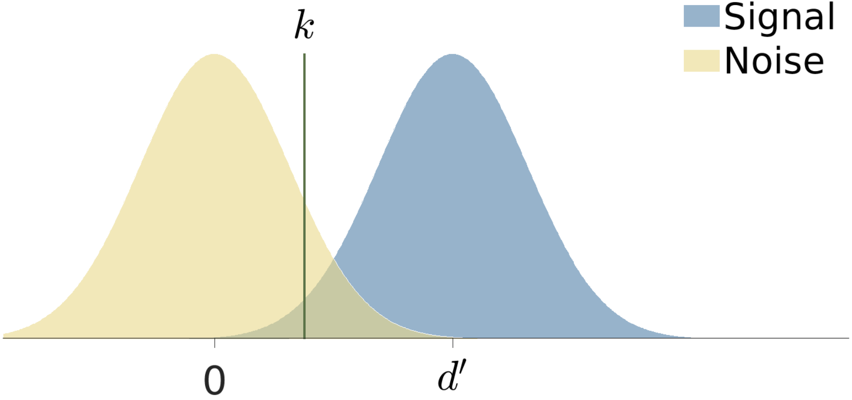
1. **Gaussian Distributions:** Internal evidence for Noise (N) and Signal+Noise (S+N) follows Normal distributions.

   $$\text{Evidence}|\text{Noise} \sim \mathcal{N}(\mu_N, \sigma_N^2)$$

   $$\text{Evidence}|\text{Signal} \sim \mathcal{N}(\mu_{S+N}, \sigma_{S+N}^2)$$

2. **Equal Variances:** $\sigma_N = \sigma_{S+N} = \sigma$.
3. **Additivity:** Signal shifts the mean ($\mu_{S+N} > \mu_N$) but not the shape.
4. **Stable Criterion:** Observer uses a fixed decision criterion k. Respond "Yes" if Evidence $> k$.

Standard Model: Illustration



SDT: Model Parameters (d', c)

Key parameters under standard assumptions:

- **Sensitivity (d'):** Standardized difference between means.

$$d' = \frac{\mu_{S+N} - \mu_N}{\sigma}$$

Measures distribution separation (signal-to-noise ratio).

SDT: Model Parameters (d', c)

Key parameters under standard assumptions:

- **Sensitivity (d'):** Standardized difference between means.

$$d' = \frac{\mu_{S+N} - \mu_N}{\sigma}$$

  Measures distribution separation (signal-to-noise ratio).

- **Criterion (c):** Position relative to the midpoint between means.

$$c = \frac{k - (\mu_N + \mu_{S+N})/2}{\sigma}$$

  Reflects response bias:

  - $c = 0$: Neutral bias (relative to midpoint).
  - $c > 0$: Conservative bias (need more evidence for "Yes").
  - $c < 0$: Liberal bias (need less evidence for "Yes").

SDT: Model Parameters (d', c)

- **Convention:** Often set $\mu_N = 0$ and $\sigma = 1$. Then:

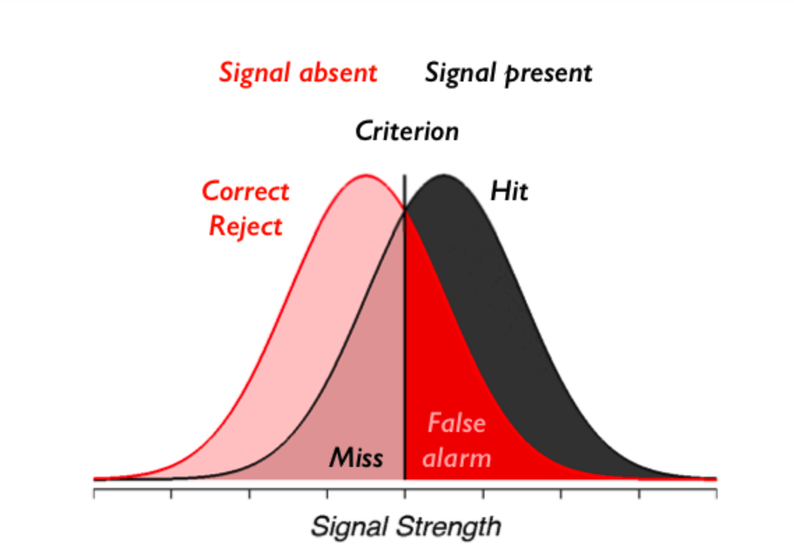
$$\text{Noise} \sim \mathcal{N}(0, 1)$$

$$\text{Signal} \sim \mathcal{N}(d', 1)$$

$$c = k - d'/2$$

$$k = c + d'/2 \text{ (Absolute criterion location)}$$

Probabilities are areas



SDT: Linking Model to Data (HR, FAR)

The model links latent parameters (d', c) to observable data via Hit Rate (HR) and False Alarm Rate (FAR), using the standard normal CDF $\Phi(z) = P(Z \leq z)$:

SDT: Linking Model to Data (HR, FAR)

The model links latent parameters (d', c) to observable data via Hit Rate (HR) and False Alarm Rate (FAR), using the standard normal CDF $\Phi(z) = P(Z \leq z)$:

- **Hit Rate (HR):** $P(\text{"Yes"}|\text{Signal})$

$$
\begin{aligned}
HR &= P(\text{Evidence} > k|\text{Signal}) \\
&= P(Z > \frac{k - \mu_{S+N}}{\sigma}) \quad \text{(Standardize)} \\
&= P(Z > k - d') \quad \text{(Using convention } \mu_{S+N} = d', \sigma = 1) \\
&= 1 - \Phi\left(k - d'\right) = \Phi\left(d' - k\right) \\
&= \Phi\left(d' - (c + d'/2)\right) = \Phi\left(d'/2 - c\right)
\end{aligned}
$$

SDT: Linking Model to Data (HR, FAR)

The model links latent parameters (d', c) to observable data via Hit Rate (HR) and False Alarm Rate (FAR), using the standard normal CDF $\Phi(z) = P(Z \leq z)$:

- **False Alarm Rate (FAR):** $P(\text{"Yes"}|\text{Noise})$

$$
\begin{aligned}
FAR &= P(\text{Evidence} > k|\text{Noise}) \\
&= P(Z > \frac{k - \mu_N}{\sigma}) \quad \text{(Standardize)} \\
&= P(Z > k) \quad \text{(Using convention } \mu_N = 0, \sigma = 1) \\
&= 1 - \Phi(k) = \Phi(-k) \\
&= \Phi\left(-(c + d'/2)\right) = \Phi\left(-d'/2 - c\right)
\end{aligned}
$$

SDT: Calculating d' and c

Equations can be inverted to estimate d' and c from observed HR and FAR.

SDT: Calculating d' and c

Equations can be inverted to estimate d' and c from observed HR and FAR. Let $z(p) = \Phi^{-1}(p)$ be the inverse standard normal CDF (z-score or quantile function).

SDT: Calculating d' and c

Equations can be inverted to estimate d' and c from observed HR and FAR. Let $z(p) = \Phi^{-1}(p)$ be the inverse standard normal CDF (z-score or quantile function).

From:

- $HR = \Phi(d'/2 - c) \implies \Phi^{-1}(HR) = d'/2 - c$
- $FAR = \Phi(-d'/2 - c) \implies \Phi^{-1}(FAR) = -d'/2 - c$

SDT: Calculating d' and c

Equations can be inverted to estimate d' and c from observed HR and FAR. Let $z(p) = \Phi^{-1}(p)$ be the inverse standard normal CDF (z-score or quantile function).

From:

- $HR = \Phi(d'/2 - c) \implies \Phi^{-1}(HR) = d'/2 - c$
- $FAR = \Phi(-d'/2 - c) \implies \Phi^{-1}(FAR) = -d'/2 - c$

Solving this system yields:

$$d' = \Phi^{-1}(HR) - \Phi^{-1}(FAR)$$

$$c = -\tfrac{1}{2}[\Phi^{-1}(HR) + \Phi^{-1}(FAR)]$$

SDT: ROC Curves

Receiver Operating Characteristic (ROC) Curve:

- Plots HR vs. FAR.
- Each point represents a (HR, FAR) pair achievable for a fixed sensitivity (d') by varying the criterion (c).
- Illustrates the trade-off: Increasing HR means increasing FAR for a given d'.

SDT: ROC Curves

Receiver Operating Characteristic (ROC) Curve:

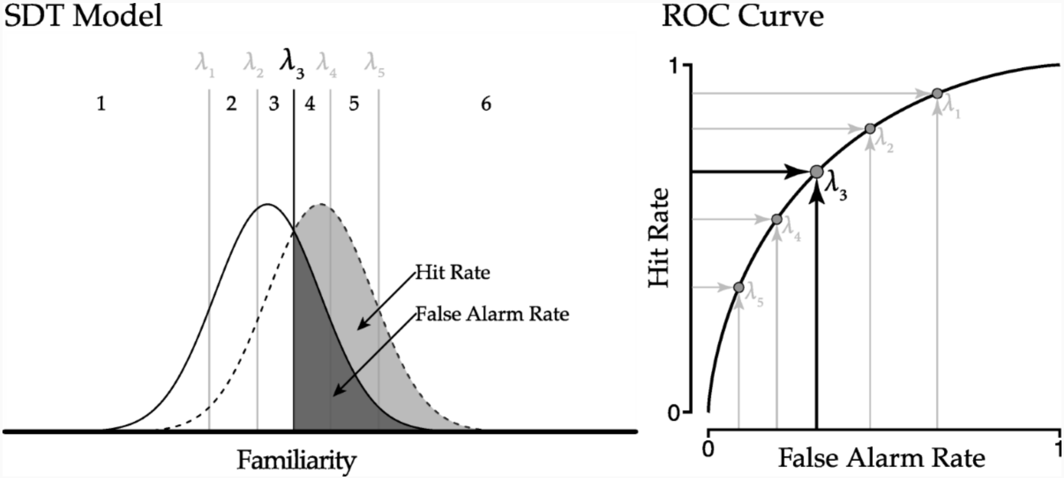
- Plots HR vs. FAR.
- Each point represents a (HR, FAR) pair achievable for a fixed sensitivity (d') by varying the criterion (c).
- Illustrates the trade-off: Increasing HR means increasing FAR for a given d'.
- **Properties:**
  - Diagonal (HR = FAR) $\implies$ Chance performance ($d' = 0$).
  - Bows towards upper-left corner (HR=1, FAR=0 = Perfect).
  - Degree of bowing reflects sensitivity (higher d' curves closer to top-left).

SDT: ROC Curves

Receiver Operating Characteristic (ROC) Curve:

- Plots HR vs. FAR.
- Each point represents a (HR, FAR) pair achievable for a fixed sensitivity (d') by varying the criterion (c).
- Illustrates the trade-off: Increasing HR means increasing FAR for a given d'.
- **Properties:**
    - Diagonal (HR = FAR) $\implies$ Chance performance ($d' = 0$).
    - Bows towards upper-left corner (HR=1, FAR=0 = Perfect).
    - Degree of bowing reflects sensitivity (higher d' curves closer to top-left).
- **Area Under the ROC Curve (AUC):**
    - Non-parametric measure of sensitivity, independent of bias.
    - AUC $= 0.5 \implies$ Chance.
    - AUC $= 1.0 \implies$ Perfect.
    - For equal-variance Gaussian model: $AUC = \Phi\left(d'/\sqrt{2}\right)$.

ROC Curve Illustration



SDT: Assumptions & Extensions

Standard model assumptions are often simplifications:

- **Unequal Variances (UVSDT):**
  - Often more realistic, but harder to estimate (needs more data, e.g., ratings).
- **Non-Gaussian Distributions:** Other evidence distributions (Poisson, Exponential) lead to different ROC shapes.
- **Criterion Variance:** Criterion might fluctuate across trials.
- **Task Dependence:** Model applies most directly to Yes/No tasks.
  - Rating Scales: Collect confidence ratings (e.g., 1-6). Allows plotting empirical ROC by collapsing ratings above different thresholds.
  - Forced-Choice (e.g., 2AFC): Choose interval/location with signal. Bias minimized. $PC = \Phi\left(d'/\sqrt{2}\right)$ for 2AFC.

SDT: Assumptions & Extensions

Limitations: SDT is a descriptive, static model. It doesn't typically explain:

- Underlying mechanisms generating distributions.
- Time course of decision (see DDM, etc.).
- Dynamic changes (learning).

First assignment: PyMC Implementation (Basic SDT)

The concepts discussed in this lecture (estimating d' and c for one observer/condition) are implemented using PyMC in a Python script in the class repository:

`0-introduction/src/sdt/basic.py`

This script demonstrates how to:

- Define priors for d' and c.
- Link parameters to observed Hits and False Alarms using a Binomial likelihood.
- Sample from the posterior distribution using MCMC.
- Analyze and visualize the results (parameter estimates, uncertainty).

Your assignment is to start the class container and run this script. Then, explore the code to study a practical application of Bayesian SDT modeling for the basic case.

Signal Detection Theory (SDT)

Joachim Vandekerckhove

Winter 2025