• 30°C water will feel HOT to a hand that was previously in 10°C water,…
• …and will feel COLD to a hand that was previously in 50°C water.
• The water has a physical property: temperature (as measured in Celsius)
• And a perceived/psychological property: it’s thermal quality (as described by ‘hotness’ or ‘coldness’)

Psychophysics deals with quantifying the relationship between physical stimuli and the perceptions they produce.
Psychophysics is a way of answering questions about our senses. Often just one question:

- What’s the smallest X you can detect/discriminate?
- What’s the smallest letter you can see?
- What’s the quietest tone you can hear?
- What’s the dimmest light you can see?
- What’s the smallest difference between two faces that you can tell apart?
Psychophysics is a way of answering questions about our senses. Often just one question:

- What’s the smallest X you can detect/discriminate?
- What’s the smallest letter you can see?
- What’s the quietest tone you can hear?
- What’s the dimmest light you can see?
- What’s the smallest difference between two faces that you can tell apart?

In psychophysics these are all questions about ‘Thresholds’ (or ‘Limens’).

How to measure a threshold?

First we need to define some (any) task, that will allow us to present different stimulus levels, and see whether the observer sees the stimulus (i.e., responds correctly).

E.g., to measure contrast sensitivity thresholds, we could present targets of different luminance (light intensity), and ask the observer:
- To read the letter aloud (identification paradigm)
- Say ‘yes’ if they see it (yes/no paradigm)
- Say whether the target was on the left or the right (two-alternative forced-choice [2AFC] paradigm)

With an ideal observer we would present each and every possible stimulus level once...

...and find the point where they start responding correctly. This is their threshold. The blue line is their psychometric function, and for the ideal observer is a step function.

Unfortunately, real observers are not ideal. Their data will often look more like this (right). It is not possible to draw a meaningful step function through these data.

Why are the data so messy?

Real observers do not respond consistently because perceptual judgments are intrinsically noisy:
- External noise (e.g., literal noise, passing traffic, jitter)
- Internal noise (e.g., stochastic neural processes, heartbeat, fluctuations in concentration)

This noise means that even if we present the exact same stimulus level multiple times, the internal response in the brain will be slightly different each time.
How to measure a threshold?

Real observers do not respond consistently because perceptual judgments are intrinsically noisy:

- **External noise** (e.g., literal noise, passing traffic, jitter)
- **Internal noise** (e.g., stochastic neural processes, heartbeat, fluctuations in concentration)

This noise means that even if we present the exact same stimulus level multiple times, the internal response in the brain will be slightly different each time.

To deal with this noise, the solution is to present every stimulus more than once, and find the proportion of the correct responses, $P_{\text{Seen}}$.

What you typically obtain is a smooth, sigmoidal psychometric function (where the slope is proportional to the magnitude of internal noise). You can then read across and down to find their threshold.

Any ‘threshold level’ (e.g., 75% correct, 82.5% correct, 93.33% correct) is valid, so be careful when comparing across papers/methods.

Real experiments tend to test fewer stimulus levels (e.g., 5 – 9), for practical reasons.

Real psychometric functions often don’t tend to start at zero, but at a higher value. (In psychophysics language: "the lower asymptote > 0").

This is because the guess rate (chance of a false positive response) is often > 0. E.g., 50% in the case of a 2AFC design.

Higher guess rates also makes the data at all stimulus levels noisier (less reliable/precise) for the same number of trials.

Real psychometric functions often don’t tend to end at 1, but at a lower value. (In psychophysics language: "the upper asymptote < 1").

This is because people get bored, tired, and distracted, and make mistakes (false negative responses).

Again, also adds noise. Plus need to be careful to actually fit the right function to the data, and not just assume that it goes up to 1.
How to measure a threshold?

As well as internal noise, guesses, and lapses, also need to be aware of **response bias** ("criterion effects").

Some people might prefer one response over another, and might differ in how confident they feel is necessary before responding. If you’re not careful this can cause sensitivity to be underestimated.

Can try to minimise this bias by using paradigms like 2AFC, rather than yes/no (e.g., since people tend to have no strong feelings about “1” vs “2”, whereas don’t like saying “yes” and being wrong).

But can never eradicate bias altogether. And, depending on how you plot the data, the bias isn’t always obvious in the psychometric function. Be careful!

**Interim Summary:**

Unlike an **ideal observer** human beings are limited by:

1. Internal noise
2. Lapses (or ‘inattentiveness’)
3. Response Bias

Because of these limitations, we have to treat detection/discrimination as a probabilistic process. Each trial/stimulus is repeated many times, and a psychometric function then fitted to the data ("Method of Constant Stimuli").

What I’ve described so far (the “Method of Constant Stimuli”) is the gold standard. But it doesn’t always work in the real world, for practical reasons:

1. Limited time
2. Non-stationary observers

Is there a faster alternative if we don’t want to characterise the whole psychometric function? (i.e., just want to know the Threshold)

Yes – adaptive methods

**Simple up-down staircase**: Increase the stimulus by +1 unit after a correct response, and by -1 unit after an incorrect response, \( K \).

Compute threshold by averaging the stimulus over the last \( N \) (even no.) of reversals.
Simple up-down staircase: Increase the stimulus by +1 unit after a correct response, \( ✓ \), and by -1 unit after an incorrect response, \( ✗ \). Compute threshold by averaging the stimulus over the last N (even no.) of reversals. This value corresponds to the point on the psychometric function where the probability of responding correctly, \( P(✓) \), is equal to the probability of answering incorrectly, \( P(✗) \). i.e., the “50% detection/discrimination threshold”

2-Up-1-Down (70.7%) transformed staircase (Levitt, 1971). Increase the stimulus by +2 unit after 2 correct responses, \( ✓✓ \), and by -1 unit after 1 incorrect response, \( ✗ \). Compute threshold by averaging the stimulus over the last N (even no.) of reversals.

2-Up-1-Down (70.7%) transformed staircase (Levitt, 1971). Increase the stimulus by +2 unit after 2 correct responses, \( ✓✓ \), and by -1 unit after 1 incorrect response, \( ✗ \). Compute threshold by averaging the stimulus over the last N (even no.) of reversals.

Up-2-Down-1 (33%) weighted staircase (Kaernbach, 1991) Increase the stimulus by +2 unit after 1 correct responses, \( ✓ \), and by -1 unit after 1 incorrect response, \( ✗ \). Compute threshold by averaging the stimulus over the last N (even no.) of reversals.

By varying the step size weighting, \( \omega \), can target any point on the psychometric function: 

\[
\begin{align*}
    \omega &= \frac{1-P(✓)}{P(✓)} \\
    \omega &= \frac{1-P(✗)}{P(✗)} \\
    \omega &= \frac{1-0.333}{0.333} = 2
\end{align*}
\]

In recent years, staircases have started to be replaced by more efficient Maximum Likelihood methods (e.g., QUEST+, psi, qCSF).

These posit a number of hypotheses, and compute the likelihood that each fits the data.
In recent years, staircases have started to be replaced by more efficient Maximum Likelihood methods (e.g., QUEST+, psi, qCSF).

These posit a number of hypotheses, and compute the likelihood that each fits the data.

They can also be used to determine the most informative stimulus to present next.

Can posit any, arbitrarily complex model.
Interim Summary:

Thresholds can be measured by:

1. Using fixed stimuli, and fitting a whole psychometric function (Method of Constant Stimuli)
   - Pros: Gold standard. Can analyse the data post-hoc and find the best fitting model
   - Cons: Slow. Potentially lots of redundant trials

2. Using an adaptive staircase
   - Pros: Fast
   - Cons: Can be affected by bias, lapses, etc.

3. Using a Maximum Likelihood adaptive procedure
   - Pros: Fastest
   - Cons: A complex method that requires lots of mathematical assumptions

Assorted key terms:
- Threshold (limen)
- Ideal observer
- Signal Detection Theory
- Internal noise (intrinsic noise)
- External noise
- Internal response
- Psychometric function
- Slope
- Guess rate (chance rate)
- Lapse rate
- Response bias (criterion effect)
- Method of Constant Stairs
- Stationary
- Adaptive staircase
- Transformed staircase
- Weighted staircase
- Maximum Likelihood
- Prior
- m-Alternative Forced choice paradigm
- yes/no paradigm

Reading: