

# Getting a Grip on Consciousness

Manuel Blum

Carnegie Mellon University

# Studying Brains in The Neurophysiology Laboratory of McCulloch and Pitts

- 1958, I apprenticed myself to the eminent neurophysiologist Warren S. McCulloch and his colleague Walter Pitts.

- From **W. S. McCulloch** I learned about **brains**

**Walter Pitts**

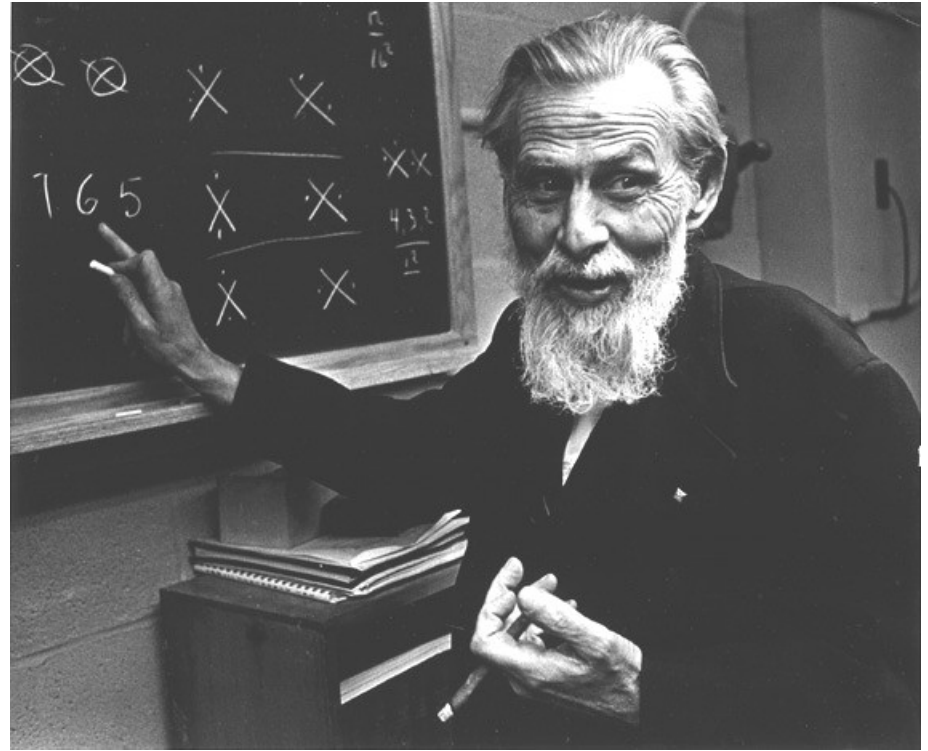
**books**

**Jerome Lettvin**


**eyes**

**Patrick Wall**

**pain**



# 1969 I got tenure at Berkeley. Took a week off to think about consciousness. What I learned?

- Consciousness is like a flashlight. 
- Wherever you look, whatever you choose to think about, ideas bubble up.
- I don't know where those ideas bubble up from.
- Visually, we see a lot less than what we think we see.
- What we think we see is what we can see if we choose to look.
  - Can focus on whatever detail we want to see.
  - Can't do that with dream images.
  - Can't do that with pain. **Pain is atomic to consciousness.**

# Why look at consciousness?

Several reasons:

- Interesting Philosophical questions.
- Curiosity about how the brain achieves it.
- Potent applications to **high-level planning**:
  - learning to think.
  - building computers/robots to understand stuff.

# Bernard Baars: **Global Workspace** **Theory** of Consciousness

- Views brain as an auditorium of unconscious processors.
- Only what's on stage (in the floodlights) is conscious.
- What's on stage is broadcast broadly to the unconscious processors (in the dark).



# **There is broad consensus that Global Workspace Theory is essentially correct**

- Rodolfo Llinas: Neurophysiologist
- David Chalmers: Philosopher who maintains online bibliography
- Antonio Damasio: Neuroscientist
- Susan Blackmore disagrees (with aspects)

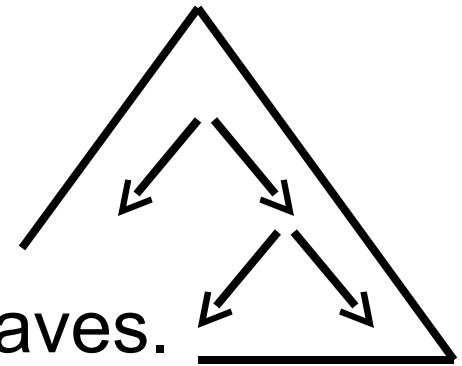
**I like Global Workspace Theory.  
It agrees with what I discovered  
for myself... and it goes much  
further.**

- It says something about what is conscious.
- Explains why consciousness is a serial bottleneck.
- Suggests experiments.
- Has broad backing.

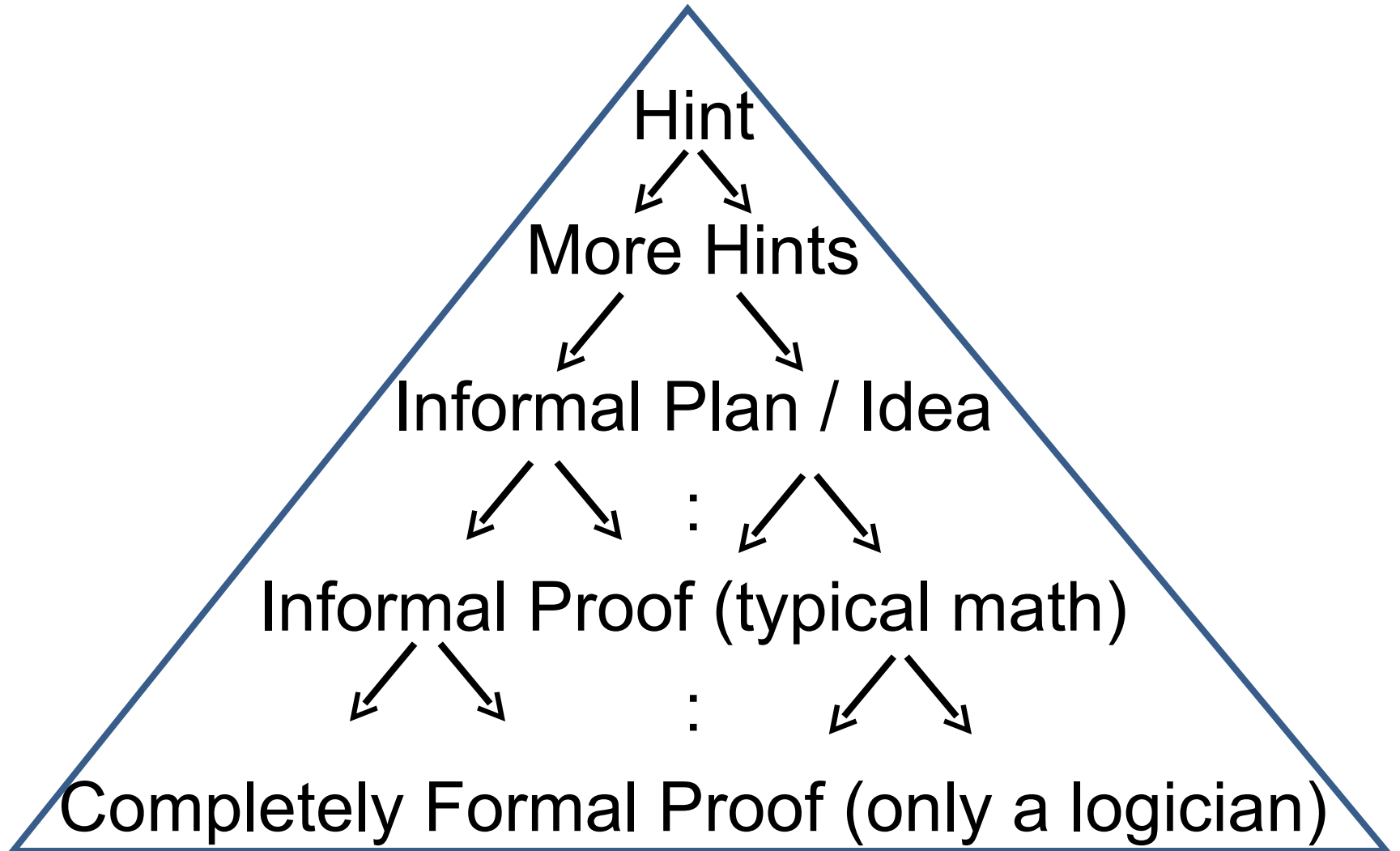


# What's in it (Global Workspace Theory) for us computer scientists?

- Global Workspace Theory lends itself to a theory of high-level planning.
- I view the proof of a theorem as a Triangle/Template/Tree
- It has an idea/ hint at the top/root.  
an informal proof at mid-level.  
a formal proof at the bottom/leaves.
- The nodes correspond to what is conscious (when we mull over that point in the proof).



# Template for a proof



# **Leslie Lamport corroborates this idea: “How to Write a Proof” 1993**

- Calls it a hierarchical proof.
- Creates such a hierarchical proof for a theorem in Kelley.
- Explains how he (Lamport) writes proofs.

# From where do we get these triangles/templates/trees?

- From examples, eg dynamic programming.
- Templates are reinforced with homework problems.
- New templates arise when those that don't quite work get fixed with ideas from other templates (Terence Tao).
- My own shtick: New templates arise from problem modification: if the problem is hard to solve, modify it to get simpler problem.

# To create new templates, try **Problem Mutation**

- If the problem is hard to solve, modify it to get one you can solve.
- Use solution to modified problem as a template to solve the original problem.
- Repeat as needed.

## **The END**

# How triangles get created

- Example: To prove the **Theorem**: There exists an irrational  $\alpha$  such that  $\alpha^\alpha$  is rational.
- **Hint** anyone ?!

Theorem: There exists an irrational  $\alpha$  such that  $\alpha^\alpha = \beta$  is rational.

**Hint:**

Try  $\alpha = \text{sqrt}(2)$ .

$\text{Sqrt}(2)^{\text{sqrt}(2)} = ?$

**Doesn't seem to work : Ouch !!**

Theorem: There exists an irrational  $\alpha$  such that  $\alpha^\alpha = \beta$  is rational.

**Hint:**

Try  $\beta = 2$ .

**Plan for a Proof:**

Try  $\beta = 2$ . Prove  $\alpha$  is irrational the way one proves that  $\sqrt{2}$  is irrational.