



Scientific Method by, and of, the Dummies

Indranil Gupta
January 27, 2010




If a Magical Genie appeared before you, and asked you to make one wish, what would you ask for?




- As a graduate student, you would ask for your...

2


You would ask for your...



Thesis!



Unfortunately, there is the Real World



Real World

Experiments



Isaac Newton was struck by an apple, and discovered gravity.

What's missing?

Thesis

4

Unfortunately, there is the Real World

Real World

Experiments


What's missing?

- Before being apple-struck Newton was already thinking about puzzles to which gravity was the solution.
- After being struck by gravity, he formed a hypothesis.
- He extrapolated, experimented, and finally
- He formed the three laws of motion

Thesis

5

You are as good a scientist as Newton!



Real World

Problem Statement(s)

New Hypothesis

Experiments

Observations

Results

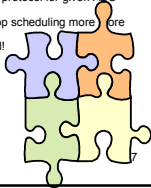
Thesis

Refine your Hypothesis

6

Let's Break it down

- **Real world:** everything around you and at your fingertips (including literature!)
- **Problem Statement:** anything that is unsolved or not solved satisfactorily. Examples:
 - increase throughput of a TCP version under congestion
 - make a file system fault-tolerant
 - make a multimedia protocol more scalable
 - find out if P=NP
 - solve an open algorithmic problem
 - something you don't know!
- **Hypothesis:** could be an **idea for a technique/algorithm and how it might perform**, or an **explanation** for a particular phenomenon (which is also an idea). Examples:
 - using a maxflow algorithm will help me design a better multimedia protocol for given ABC setting.
 - I have a new idea for a scheduling algorithm that can make Hadoop scheduling more fault-tolerant!
 - I have a series of ways that explain why Web traffic is heavy-tailed!
- Experiment
- Observation
- Results
- Thesis: the golden egg



Example



- **Example Problem Statement:**
 - You are bored.
 - You have to design a scalable, efficient, and fault-tolerant scheduling algorithm for the Dryad execution engine in Clouds.

8

Example



- **Example Problem Statement:**
 - You are bored: **you need something to solve your boredom**
 - You have to design a scalable, efficient, and fault-tolerant scheduling algorithm for the Dryad execution engine in Clouds: **you need to find a new protocol and implement a new system (note “find” may mean find from literature, or if not present there, invent).**

9

Example



- **Example Hypothesis:**
 - You are bored: **listening to music might help**
 - You have to design a scalable, efficient, and fault-tolerant scheduling algorithm for the Dryad execution engine in Clouds: **<Your idea here>**

10

Experiment: Prep for



- **Example new hypothesis:** **“there is a radio station broadcasting music at 91.2 FM, that I can listen to now”**
- How do you verify this?
- **Independent Variables:** these are variables that you have control over: knobs you can turn
 - The frequency dial on the radio, the volume control
- **Dependent Variables:** these are variables that you do not directly control, but that are affected by your independent variables
 - The sound that you hear from the radio
- **Control Variables:** these are variables that you typically do not vary.
 - The radio set you are using

11

Performing the Experiment



- You pick up your old radio set, turn the dial to 91.2 FM, and listen for the sound
- **Observation:** Two possibilities:
 1. You hear music or someone talking on the air: your hypothesis is proved! You have a result!
 - You may or may not have a **thesis!** (if the radio solves your boredom you do, otherwise you may need a new hypothesis)
 2. You hear static: **you conclude there is no radio station broadcasting now at 91.2 FM. Your hypothesis is disproved.**
- What is the fallacy with the above reasoning?

12

Performing the Experiment



- You pick up your old radio set, turn the dial to 91.2 FM, and listen for the sound
- Observation: Two possibilities:
 - You hear music or someone talking on the air: your hypothesis is proved! You have a result!
 - You may or may not have a **thesis!** (if the radio solves your boredom you do, otherwise you may need a new hypothesis)
 - You hear static: you conclude there is no radio station broadcasting now at 91.2 FM. Your hypothesis is disproved.
- What is the fallacy with the above reasoning?
 - What if the radio is faulty? (What if your control variables are misplaced?)
 - What if the tuner button does not work? (What if the independent variable is actually not getting changed?)
 - What if the speakers are busted? (What if the dependent variables are not being influenced at all?)
 - What if the radio station is off the air now? (Hmm)

13

So what do you do?

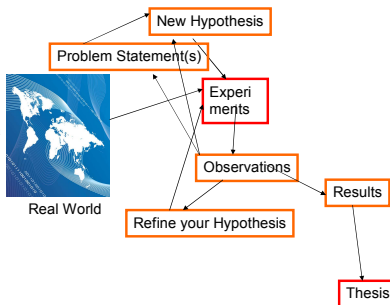


- New hypothesis: **this radio might be faulty**
 - Try a couple of other channels you know (say 90.9 FM and 92.5 FM)
 - If new channels come up with static, then you have a better confidence that your new hypothesis is true: **this radio might be faulty** → try a different radio, repeat above experiment.
 - Another new hypothesis: **fixing this radio is fun, and it will solve my boredom!** (your original problem was to solve boredom; you are free to change that problem statement and go off and fix the radio) → you go off on a new set of observations, and experiments
 - If new channels come up with music, then you have another new hypothesis that **91.2 FM either is not on the air or does not exist anymore**
 - How do you verify this hypothesis? You can try to find the 91.2 FM online, or drive over to the radio station and see if it is still open!
 - May also solve your boredom!
 - On the other hand, you may have a new hypothesis: **I am not bored anymore**

14

Summary so far...

- Careful how you interpret observations/data!
- Don't be afraid to form new hypotheses!
- The cycle of the scientific method is a never-ending process.
- Like it or not, this is what you/I/we all do all the time.
 - So we might as well do it consciously and systematically!



15

Important things to Remember

- Hypotheses are of different types: some examples
 - Inductive:** generalization based on specific observations.
 - Deductive:** derived from theory and supports, expands, or contradicts the theory.
- A good hypothesis is **refutable**
- Finding proofs supporting a hypothesis is tricky
 - If someone hypothesizes that **every Ford Focus car is blue**,
 - You see a Ford Focus that is red: what do you conclude?
 - You see a 100 Ford Focuses and they are all blue: what do you conclude?
 - If someone hypothesizes that **there will never be a green Ford Focus**,
 - Even if you see all the Ford Focuses in the world and none is green, what do you conclude?
 - Not a good hypothesis



16

Important things to Remember

- Experiments have to be **repeatable!**
 - You ran an experiment and then say it cannot be repeated by any other scientist: it has no value to the world
 - E.g., Bell Labs scientist who claimed to create a 1-atom transistor
 - E.g., Korean scientist who claimed to be the first to clone an animal
- You have to **repeat your trials** for each data point
 - A single or two trials are never enough
 - The world is mostly probabilistic (with high probability)
 - More on this later in CS591FSN!
- You need **patience** during experimentation
 - I drive k miles away from Champaign in a random direction. I repeat experiment 300 times.
 - If k=100 (low patience while driving), I can conclude "world is flat"
 - If k higher, maybe I will get a different result
- Don't forge your results!



17

So we solved the boredom problem; now you can go off and solve the other problem...

- Example **Hypothesis**:
 - ✓ You are bored: **listening to music might help**
 - You have to design a scalable, efficient, and fault-tolerant scheduling algorithm for the Dryad execution engine in Clouds: **<Your idea here>**
- The scientific method is your most reliable approach to obtain fast and impactful research results**

18