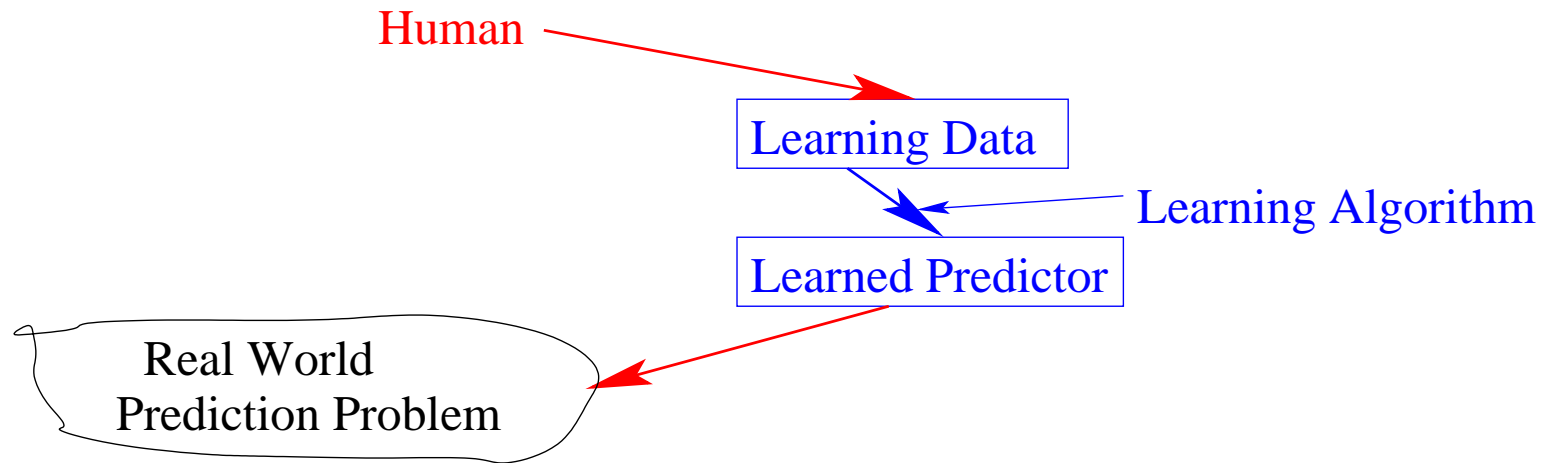


What is Learning Problem Design?

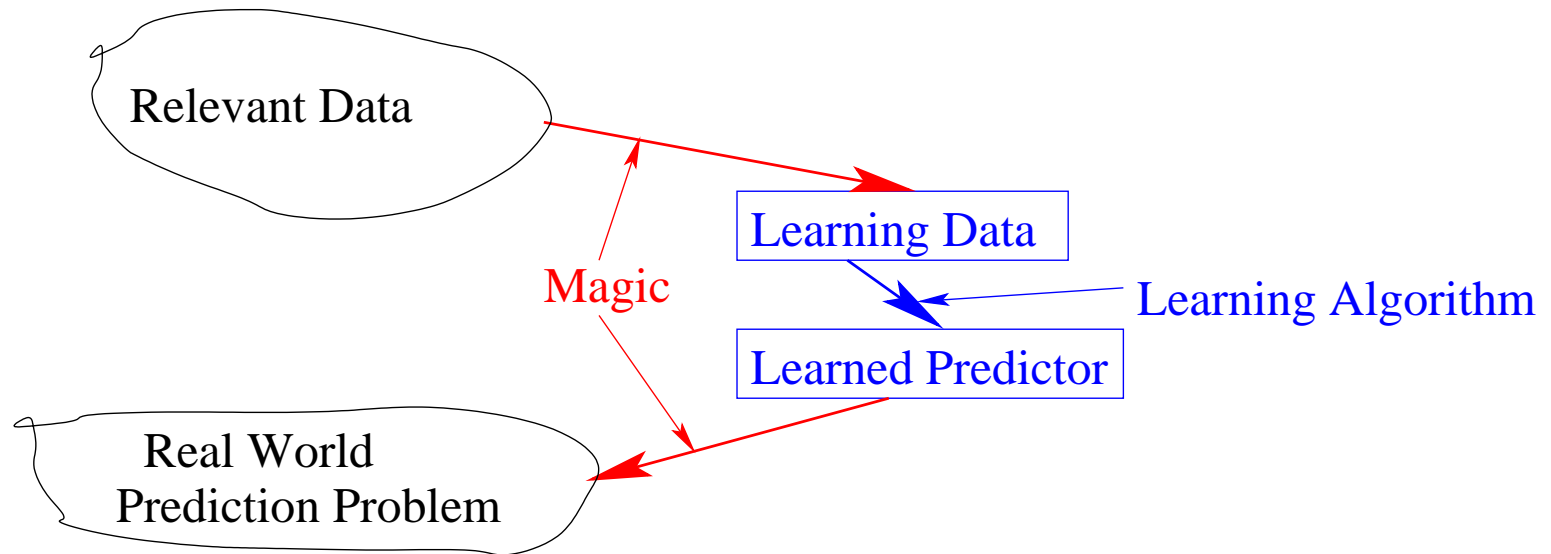
Learning Problem Design Workshop, NIPS 2007

John Langford

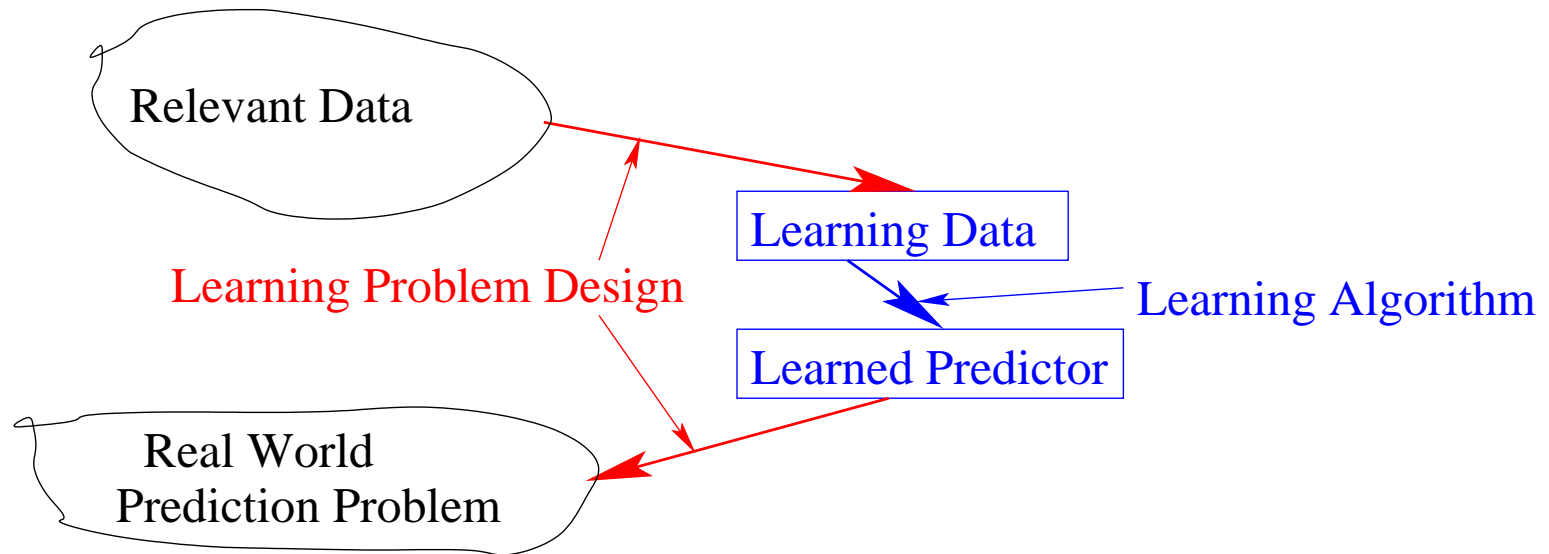
## How Learning Works



## How Learning Sometimes Works



## How we would like learning to Work



# Example: Ad Display Policy

Web | Images | Video | Local | Shopping | more ▾

Flowers Search Options ▾

1-10 of 289,000,000 for **Flowers** (About) - 0.12 sec

SPONSOR RESULTS

Send **Flowers** Same Day For Less 🛒  
Beautiful Fresh **Flowers** Under \$30. Quality & Satisfaction Guaranteed.  
[FromYouFlowers.com](#)

**Flower**  
Get a Free Vase w/ Every **Flower** Bouquet Order at Florist.com Now.  
[www.Florist.com](#)

Free Delivery, FTD **Flowers**  
Free delivery for online orders. FTD florist satisfaction guaranteed.  
[www.ftdfloristsonline.com/flowers](#)

Get **Flowers** at [Justflowers.com](#)  
Same day **flower** delivery by local florist -save \$5 coupon "save5".  
[www.Justflowers.com](#)


Same Day **Flower** Delivery  
Send **Flowers** & Gifts for Less Nationwide. Satisfaction

Also try: [1800 flowers](#), [pictures of flowers](#), [wedding flowers](#), [send flowers](#), [More...](#)

SPONSOR RESULTS

- [FTD® Official Site For Flowers](#) 🛒  
[www.FTD.com](#) - Order Designer **Flowers** & Gifts From \$19.99. Same Day Delivery.
- [Flowers at 1-800-FLOWERS®](#)  
[1800flowers.com](#) - **Flowers** & Gifts from \$24.99. Your Satisfaction Guaranteed.
- [Save \\$10 On Flowers](#)  
[www.teleflora.com](#) - Surprise Them with Fresh **Flowers**. Same Day Delivery by Local Florists.
- [Send Flowers from \\$19.99](#)  
[www.proflowers.com](#) - Send Roses, Lilies & other **Flowers**. Best Value - Wall Street Journal.

**Flowers** - Image Results



[More Flowers images](#)

Yahoo! Shortcut - [About](#)

1. [1-800-FLOWERS.COM](#) - Official Site (Nasdaq: **FLWS**)

You might think this is a ranking or multiclass prediction problem, but...

It's *not* supervised learning. You do not know what would have happened in the past if different ads had been displayed  $\Rightarrow$  you can't evaluate a new policy given data from an arbitrary old policy.

1. What should you predict?
2. And how do you use it to solve the real problem?

## Example: Spam Filtration



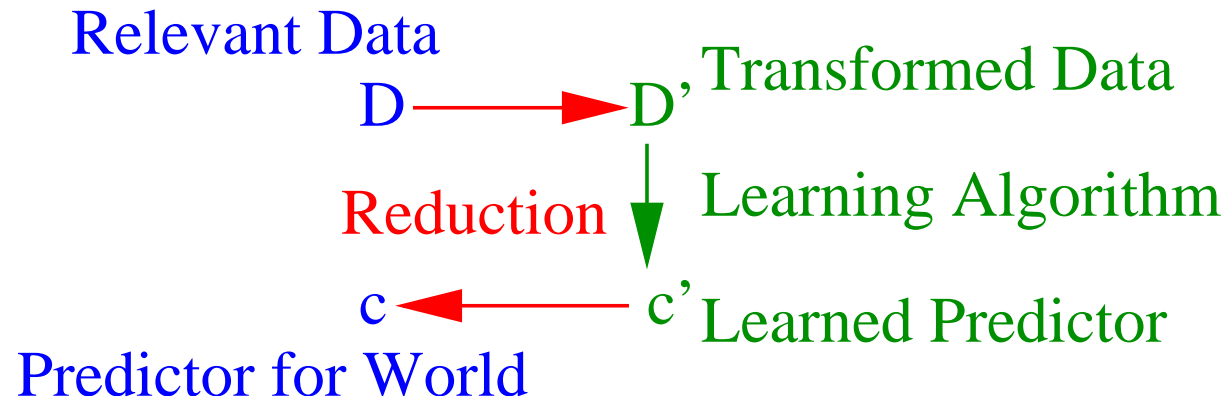
This looks like a standard binary classification problem  
but...

1. There is a shared sense of what spam is, but it's not completely shared. Users are sometimes even adversarial.
2. What is the label of an email? The act of shifting it into a folder? Or the act of leaving it in a folder? Or something else?

## Questions We ask for Learning Problem Design

1. What are relevant data sources?
2. What are good methods for using relevant data sources?
3. How can we best design learning problems taking advantage of relevant data?

## The Role of Reductions



Optimize **Reduction** so that

$$\text{Regret}(c, D) \equiv \text{Loss}(c, D) - \min_h \text{Loss}(h, D)$$

tightly bounded by

$$\text{Regret}(c', D') \equiv \text{Err}(c', D') - \min_{h'} \text{Err}(h', D')$$

## Why Reductions are not a complete answer

1. It isn't clear we can find a reduction for every relevant data source.
2. Reductions assume away the difficulty of the learning step.