COGS 303

Gary Neels

UBC

Oct 27, 2023

1. Assignments
   - Problem Set
   - Literature Review
2. Replication and why it matters
3. Sources of the Problem
   - Systemic problems
   - Questionable Research Practices

Problem Set Question 1

The key here is to see that Larry's claim $P(D|V) = P(D|\neg V)$ reduces to the simpler claim that $P(V) = P(\neg V)$:

- There's more than one way to get there. Here's one way

Problem Set Question 1

The key here is to see that Larry's claim $P(D|V) = P(D|\neg V)$ reduces to the simpler claim that $P(V) = P(\neg V)$:

- There's more than one way to get there. Here's one way
- Larry's claim is equivalent to

$$\frac{P(D \wedge V)}{P(V)} = \frac{P(D \wedge \neg V)}{P(\neg V)}$$

Problem Set Question 1

The key here is to see that Larry's claim $P(D|V) = P(D|\neg V)$ reduces to the simpler claim that $P(V) = P(\neg V)$:

- There's more than one way to get there. Here's one way
- Larry's claim is equivalent to

$$\frac{P(D \wedge V)}{P(V)} = \frac{P(D \wedge \neg V)}{P(\neg V)}$$

- The numerators are equivalent (they are both equal to $\frac{50}{totalpopulation}$)

Problem Set Question 1

The key here is to see that Larry's claim $P(D|V) = P(D|\neg V)$ reduces to the simpler claim that $P(V) = P(\neg V)$:

- There's more than one way to get there. Here's one way
- Larry's claim is equivalent to

$$\frac{P(D \wedge V)}{P(V)} = \frac{P(D \wedge \neg V)}{P(\neg V)}$$

- The numerators are equivalent (they are both equal to $\frac{50}{totalpopulation}$)
- So, his claim amounts to the claim that the denominators are equivalent

Problem Set Question 1

The key here is to see that Larry's claim $P(D|V) = P(D|\neg V)$ reduces to the simpler claim that $P(V) = P(\neg V)$:

- There's more than one way to get there. Here's one way
- Larry's claim is equivalent to

$$\frac{P(D \wedge V)}{P(V)} = \frac{P(D \wedge \neg V)}{P(\neg V)}$$

- The numerators are equivalent (they are both equal to $\frac{50}{totalpopulation}$)
- So, his claim amounts to the claim that the denominators are equivalent
- So, his inference would only be good if he knew that to be the case (doubtful)

Problem Set Question 2

Here, the key is to notice the implication that $P(A) = 1$ has for $P(\neg A)$ and then to see what that means for $P(A|B)$

- Using one of our versions of Bayes' theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)}$$

Problem Set Question 2

Here, the key is to notice the implication that $P(A) = 1$ has for $P(\neg A)$ and then to see what that means for $P(A|B)$

- Using one of our versions of Bayes' theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)}$$

- Since, $P(A) = 1$, we know that $P(\neg A) = 0$

Problem Set Question 2

Here, the key is to notice the implication that $P(A) = 1$ has for $P(\neg A)$ and then to see what that means for $P(A|B)$

- Using one of our versions of Bayes' theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)}$$

- Since, $P(A) = 1$, we know that $P(\neg A) = 0$
- This means that $P(B|\neg A)P(\neg A) = 0$

Problem Set Question 2

Here, the key is to notice the implication that $P(A) = 1$ has for $P(\neg A)$ and then to see what that means for $P(A|B)$

- Using one of our versions of Bayes' theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)}$$

- Since, $P(A) = 1$, we know that $P(\neg A) = 0$
- This means that $P(B|\neg A)P(\neg A) = 0$
- So,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + 0} = 1$$

Problem Set Question 2

Here, the key is to notice the implication that $P(A) = 1$ has for $P(\neg A)$ and then to see what that means for $P(A|B)$

- Using one of our versions of Bayes' theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)}$$

- Since, $P(A) = 1$, we know that $P(\neg A) = 0$
- This means that $P(B|\neg A)P(\neg A) = 0$
- So,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + 0} = 1$$

- Whatever B is, you will continue to be dogmatic about A

Problem Set Question 3

Here we want to compare two conditional probabilities, $P(1|h)$ and $P(2|h)$ (h=the host reveals that the prize is not behind door 3)

- Door 1

$$P(1|h) = \frac{P(h|1)P(1)}{P(h|1)P(1) + P(h|2)P(2)}$$

Problem Set Question 3

Here we want to compare two conditional probabilities, $P(1|h)$ and $P(2|h)$ (h=the host reveals that the prize is not behind door 3)

- Door 1

$$P(1|h) = \frac{P(h|1)P(1)}{P(h|1)P(1) + P(h|2)P(2)}$$

- Door 2

$$P(2|h) = \frac{P(h|2)P(2)}{P(h|1)P(1) + P(h|2)P(2)}$$

Problem Set Question 3

Here we want to compare two conditional probabilities, $P(1|h)$ and $P(2|h)$ (h=the host reveals that the prize is not behind door 3)

- Door 1

$$P(1|h) = \frac{P(h|1)P(1)}{P(h|1)P(1) + P(h|2)P(2)}$$

- Door 2

$$P(2|h) = \frac{P(h|2)P(2)}{P(h|1)P(1) + P(h|2)P(2)}$$

- The denominators are equivalent, and $P(1) = P(2)$, so it comes down to the likelihoods

Problem Set Question 3

Here we want to compare two conditional probabilities, $P(1|h)$ and $P(2|h)$ (h=the host reveals that the prize is not behind door 3)

- Door 1

$$P(1|h) = \frac{P(h|1)P(1)}{P(h|1)P(1) + P(h|2)P(2)}$$

- Door 2

$$P(2|h) = \frac{P(h|2)P(2)}{P(h|1)P(1) + P(h|2)P(2)}$$

- The denominators are equivalent, and $P(1) = P(2)$, so it comes down to the likelihoods
- If the prize is behind door 1, which you have chosen, then the host can pick between revealing door 2 or 3, so $P(h|1) = 0.5$

Problem Set Question 3

Here we want to compare two conditional probabilities, $P(1|h)$ and $P(2|h)$ (h=the host reveals that the prize is not behind door 3)

- Door 1

$$P(1|h) = \frac{P(h|1)P(1)}{P(h|1)P(1) + P(h|2)P(2)}$$

- Door 2

$$P(2|h) = \frac{P(h|2)P(2)}{P(h|1)P(1) + P(h|2)P(2)}$$

- The denominators are equivalent, and $P(1) = P(2)$, so it comes down to the likelihoods
- If the prize is behind door 1, which you have chosen, then the host can pick between revealing door 2 or 3, so $P(h|1) = 0.5$
- If the prize is behind door 2, which you haven't chosen, then the host has to reveal door 3, so $P(h|2) = 1$

Problem Set Question 3

Here we want to compare two conditional probabilities, $P(1|h)$ and $P(2|h)$ (h=the host reveals that the prize is not behind door 3)

- Door 1

$$P(1|h) = \frac{P(h|1)P(1)}{P(h|1)P(1) + P(h|2)P(2)}$$

- Door 2

$$P(2|h) = \frac{P(h|2)P(2)}{P(h|1)P(1) + P(h|2)P(2)}$$

- The denominators are equivalent, and $P(1) = P(2)$, so it comes down to the likelihoods
- If the prize is behind door 1, which you have chosen, then the host can pick between revealing door 2 or 3, so $P(h|1) = 0.5$
- If the prize is behind door 2, which you haven't chosen, then the host has to reveal door 3, so $P(h|2) = 1$
- The upshot: if you switch, you're twice as likely to win the prize

Problem Set Question 4

Here, it's a matter of calculating the EMV for taking the bet with the Canucks winning, and figuring out the value where the other bet would have break even (in terms of EMV)

- EMV if Canucks win$= (0.5 \cdot \$5) + (0.5 \cdot -\$2) = \$1.50$

Problem Set Question 4

Here, it's a matter of calculating the EMV for taking the bet with the Canucks winning, and figuring out the value where the other bet would have break even (in terms of EMV)

- EMV if Canucks win$= (0.5 \cdot \$5) + (0.5 \cdot -\$2) = \$1.50$
- EMV if Kraken win$=(0.5x) + (0.5 \cdot -\$10) = 0.5x - \$5$

Problem Set Question 4

Here, it's a matter of calculating the EMV for taking the bet with the Canucks winning, and figuring out the value where the other bet would have break even (in terms of EMV)

- EMV if Canucks win$= (0.5 \cdot \$5) + (0.5 \cdot -\$2) = \$1.50$
- EMV if Kraken win$=(0.5x) + (0.5 \cdot -\$10) = 0.5x - \$5$
- Solve for x where EMV$= \$1.50$

$$\$1.50 = 0.5x - \$5$$

$$\$6.50 = 0.5x$$

$$x = \$13$$

Problem Set Question 4

Here, it's a matter of calculating the EMV for taking the bet with the Canucks winning, and figuring out the value where the other bet would have break even (in terms of EMV)

- EMV if Canucks win$= (0.5 \cdot \$5) + (0.5 \cdot -\$2) = \$1.50$
- EMV if Kraken win$=(0.5x) + (0.5 \cdot -\$10) = 0.5x - \$5$
- Solve for x where EMV$= \$1.50$

$$\$1.50 = 0.5x - \$5$$

$$\$6.50 = 0.5x$$

$$x = \$13$$

- At this value, she would break even, so this is the minimum payoff that she would need to take this bet

Problem Set Question 4

Here, it's a matter of calculating the EMV for taking the bet with the Canucks winning, and figuring out the value where the other bet would have break even (in terms of EMV)

- EMV if Canucks win$= (0.5 \cdot \$5) + (0.5 \cdot -\$2) = \$1.50$
- EMV if Kraken win$=(0.5x) + (0.5 \cdot -\$10) = 0.5x - \$5$
- Solve for x where EMV$= \$1.50$

$$\$1.50 = 0.5x - \$5$$

$$\$6.50 = 0.5x$$

$$x = \$13$$

- At this value, she would break even, so this is the minimum payoff that she would need to take this bet
- Most common error on this one is to miss that we know her fair betting ratio is 0.5

Problem Set Question 5

This is a lot like an example we worked through in class:

- $$P(S|P) = \frac{P(P|S)P(S)}{P(P|S)P(S) + P(P|\neg S)P(\neg S)}$$

Problem Set Question 5

This is a lot like an example we worked through in class:

- $$P(S|P) = \frac{P(P|S)P(S)}{P(P|S)P(S) + P(P|\neg S)P(\neg S)}$$
- $P(S) = \frac{1}{20000} = 0.00005$

Problem Set Question 5

This is a lot like an example we worked through in class:

- $$P(S|P) = \frac{P(P|S)P(S)}{P(P|S)P(S) + P(P|\neg S)P(\neg S)}$$
- $P(S) = \frac{1}{20000} = 0.00005$
- $P(\neg S) = 1 - 0.00005 = 0.99995$

Problem Set Question 5

This is a lot like an example we worked through in class:

- $$P(S|P) = \frac{P(P|S)P(S)}{P(P|S)P(S) + P(P|\neg S)P(\neg S)}$$
- $P(S) = \frac{1}{20000} = 0.00005$
- $P(\neg S) = 1 - 0.00005 = 0.99995$
- $P(P|S) = 0.99$

Problem Set Question 5

This is a lot like an example we worked through in class:

- $$P(S|P) = \frac{P(P|S)P(S)}{P(P|S)P(S) + P(P|\neg S)P(\neg S)}$$
- $P(S) = \frac{1}{20000} = 0.00005$
- $P(\neg S) = 1 - 0.00005 = 0.99995$
- $P(P|S) = 0.99$
- $P(P|\neg S) = 0.005$

Problem Set Question 5

This is a lot like an example we worked through in class:

- $$P(S|P) = \frac{P(P|S)P(S)}{P(P|S)P(S) + P(P|\neg S)P(\neg S)}$$
- $P(S) = \frac{1}{20000} = 0.00005$
- $P(\neg S) = 1 - 0.00005 = 0.99995$
- $P(P|S) = 0.99$
- $P(P|\neg S) = 0.005$
- So,

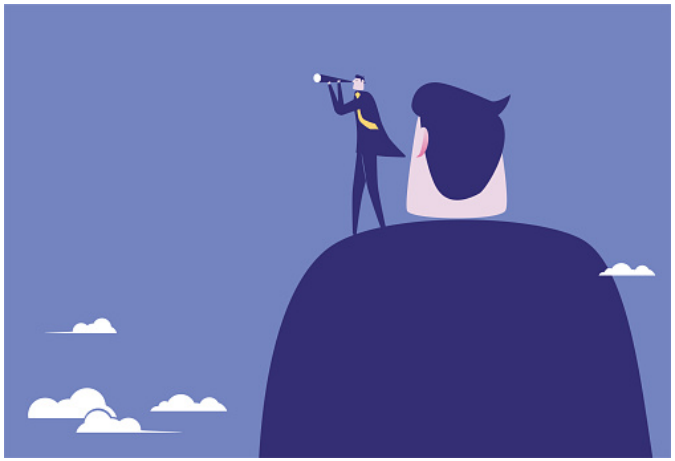
$$P(S|P) = \frac{0.99 \cdot 0.00005}{(0.99 \cdot 0.00005) + (0.005 \cdot 0.99995)}$$

$$= \frac{0.0000495}{0.0000495 + 0.00499975} \approx 0.0098$$

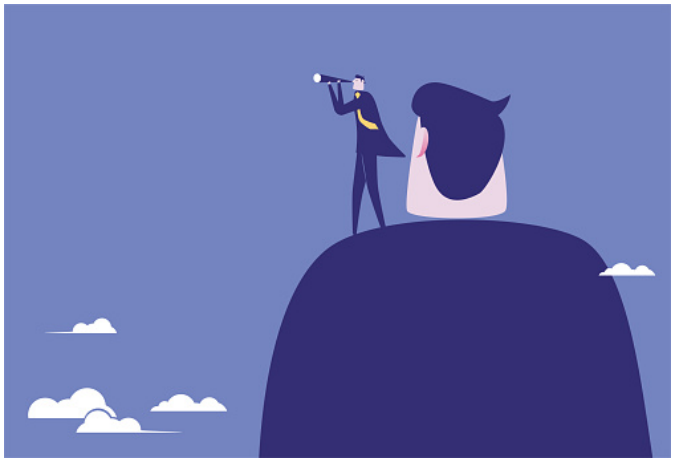
Literature Review

- A crucial aspect of research is learning what others have already discovered or said about your topic

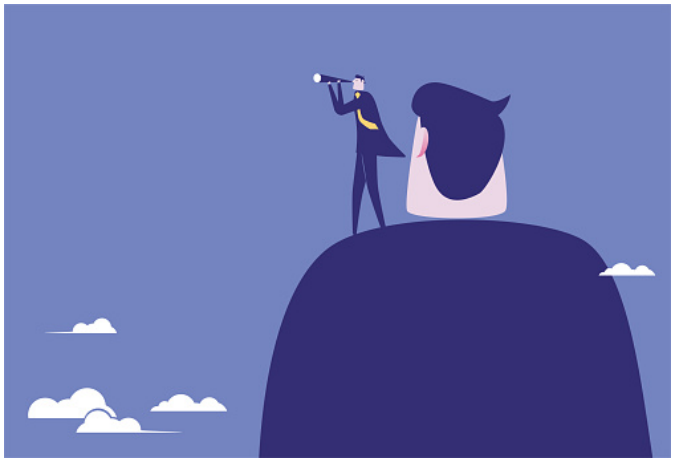
Literature Review

- A crucial aspect of research is learning what others have already discovered or said about your topic
- 

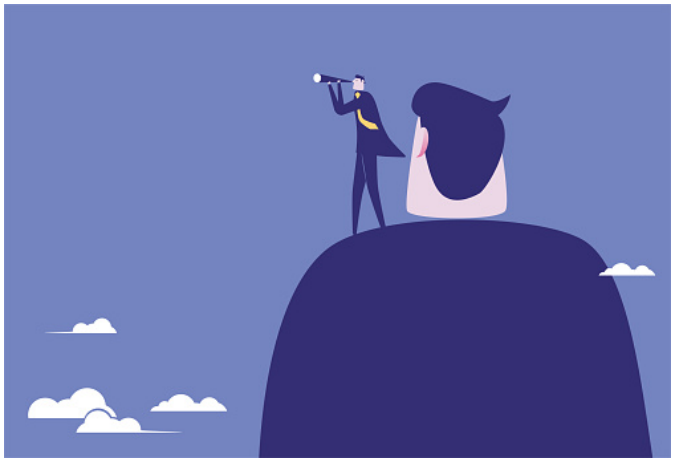
Literature Review

- A crucial aspect of research is learning what others have already discovered or said about your topic
- 
- There's no need to re-invent the wheel; build on the work that's been done

Literature Review

- A crucial aspect of research is learning what others have already discovered or said about your topic
- 
- There's no need to re-invent the wheel; build on the work that's been done
- Being able to accurately and succinctly summarize what has been said about a topic is a key skill for any researcher

Literature Review

- A crucial aspect of research is learning what others have already discovered or said about your topic
- 
- There's no need to re-invent the wheel; build on the work that's been done
- Being able to accurately and succinctly summarize what has been said about a topic is a key skill for any researcher
- The lit review assignment is an exercise to help you develop this skill

Literature Review

- Aim for clarity, accuracy, charity

Literature Review

- Aim for clarity, accuracy, charity
- **Synthesize** the material into a coherent whole (organize by content, not by source)

Literature Review

- Aim for clarity, accuracy, charity
- **Synthesize** the material into a coherent whole (organize by content, not by source)
- Note that the material from your lit review will comprise part (roughly half) of your final paper

Literature Review

- Aim for clarity, accuracy, charity
- **Synthesize** the material into a coherent whole (organize by content, not by source)
- Note that the material from your lit review will comprise part (roughly half) of your final paper
- In your final paper, you will be presenting your view on the subject, in interaction with the other views that are out there

General Writing Tips

- Science papers and arts papers have very different formats

General Writing Tips

- Science papers and arts papers have very different formats
- Science papers often have a standard set of headings/sections: context, methods, results, comment, etc.

General Writing Tips

- Science papers and arts papers have very different formats
- Science papers often have a standard set of headings/sections: context, methods, results, comment, etc.
- Your work in this course should not have that format

General Writing Tips

- Science papers and arts papers have very different formats
- Science papers often have a standard set of headings/sections: context, methods, results, comment, etc.
- Your work in this course should not have that format
- Overall, your work in this course should more closely resemble an arts paper

General Writing Tips

- Science papers and arts papers have very different formats
- Science papers often have a standard set of headings/sections: context, methods, results, comment, etc.
- Your work in this course should not have that format
- Overall, your work in this course should more closely resemble an arts paper
- But don't be intimidated if you're less familiar with that format

General Writing Tips

- Science papers and arts papers have very different formats
- Science papers often have a standard set of headings/sections: context, methods, results, comment, etc.
- Your work in this course should not have that format
- Overall, your work in this course should more closely resemble an arts paper
- But don't be intimidated if you're less familiar with that format
- Just think of your lit review as the "context" section, and your final paper as a combination of the "context" and "comment" sections

General Writing Tips

- Editing: look to avoid

General Writing Tips

- Editing: look to avoid
  - Run-on sentences

General Writing Tips

- Editing: look to avoid
  - Run-on sentences
  - Fragments

General Writing Tips

- Editing: look to avoid
  - Run-on sentences
  - Fragments
  - Punctuation errors

General Writing Tips

- Editing: look to avoid
  - Run-on sentences
  - Fragments
  - Punctuation errors
  - Spelling errors

General Writing Tips

- Editing: look to avoid
  - Run-on sentences
  - Fragments
  - Punctuation errors
  - Spelling errors
  - Subject/verb disagreement

General Writing Tips

- Editing: look to avoid
  - Run-on sentences
  - Fragments
  - Punctuation errors
  - Spelling errors
  - Subject/verb disagreement
- Ask a friend or classmate proofread for you before turning your work in

General Writing Tips

- Editing: look to avoid
  - Run-on sentences
  - Fragments
  - Punctuation errors
  - Spelling errors
  - Subject/verb disagreement
- Ask a friend or classmate proofread for you before turning your work in
- Citations: pick a style and stick with it (apa preferred, but not required)

Formatting Tips

Formatting Tips

- Intro and thesis statement

Formatting Tips

- Intro and thesis statement
  - The intro should provide some general background to your chosen topic in a single paragraph

Formatting Tips

- Intro and thesis statement
  - The intro should provide some general background to your chosen topic in a single paragraph
  - The last sentence should be your thesis statement–a succinct, one-sentence expression of what you will be presenting/arguing

Formatting Tips

- Intro and thesis statement
  - The intro should provide some general background to your chosen topic in a single paragraph
  - The last sentence should be your thesis statement–a succinct, one-sentence expression of what you will be presenting/arguing
  - For the lit review, your thesis might be something like: "With respect to question X, there are two schools of thought, A and B."

Formatting Tips

- Intro and thesis statement
  - The intro should provide some general background to your chosen topic in a single paragraph
  - The last sentence should be your thesis statement–a succinct, one-sentence expression of what you will be presenting/arguing
  - For the lit review, your thesis might be something like: "With respect to question X, there are two schools of thought, A and B."
  - For your final paper, your thesis might be something like: "A is preferable to B for reasons X, Y, and Z."

Formatting Tips

- Body:

Formatting Tips

- Body:
  - This is your main content

Formatting Tips

- Body:
  - This is your main content
  - It should be clear to your reader when you are summarizing someone else's position and when you are presenting your own opinion

Formatting Tips

- Body:
  - This is your main content
  - It should be clear to your reader when you are summarizing someone else's position and when you are presenting your own opinion
  - Use transition words to indicate how each paragraph is connected to the previous one

Formatting Tips

- Body:
  - This is your main content
  - It should be clear to your reader when you are summarizing someone else's position and when you are presenting your own opinion
  - Use transition words to indicate how each paragraph is connected to the previous one
  - Each paragraph should have a purpose in your paper/review, and that purpose should be clear to your reader

Formatting Tips

- Body:
  - This is your main content
  - It should be clear to your reader when you are summarizing someone else's position and when you are presenting your own opinion
  - Use transition words to indicate how each paragraph is connected to the previous one
  - Each paragraph should have a purpose in your paper/review, and that purpose should be clear to your reader
  - Pay attention to the length requirements: you will lose marks if your work is too long/too short

Formatting Tips

- Conclusion:

Formatting Tips

- Conclusion:
  - End with a concise summary of the material you have presented

Formatting Tips

- Conclusion:
  - End with a concise summary of the material you have presented
  - Your conclusion should **not** introduce any new material

Formatting Tips

- Conclusion:
  - End with a concise summary of the material you have presented
  - Your conclusion should **not** introduce any new material
  - Keep the conclusion short

Replication

Replication

What is replication?

- This can refer to different concepts

Replication

What is replication?

- This can refer to different concepts
  - Re-evaluating evidence from previous studies (recall the dioxin study Douglas mentioned)

Replication

What is replication?

- This can refer to different concepts
  - Re-evaluating evidence from previous studies (recall the dioxin study Douglas mentioned)
  - Re-doing experiments

Replication

What is replication?

- This can refer to different concepts
  - Re-evaluating evidence from previous studies (recall the dioxin study Douglas mentioned)
  - Re-doing experiments
- In high school science, most (all?) experiments are replications in the second sense

Replication

What is replication?

- This can refer to different concepts
  - Re-evaluating evidence from previous studies (recall the dioxin study Douglas mentioned)
  - Re-doing experiments
- In high school science, most (all?) experiments are replications in the second sense
- This re-doing of experiments can be further classified:

Replication

What is replication?

- This can refer to different concepts
  - Re-evaluating evidence from previous studies (recall the dioxin study Douglas mentioned)
  - Re-doing experiments
- In high school science, most (all?) experiments are replications in the second sense
- This re-doing of experiments can be further classified:
  - Direct–Closely following the original experiment

Replication

What is replication?

- This can refer to different concepts
  - Re-evaluating evidence from previous studies (recall the dioxin study Douglas mentioned)
  - Re-doing experiments
- In high school science, most (all?) experiments are replications in the second sense
- This re-doing of experiments can be further classified:
  - Direct–Closely following the original experiment
  - Indirect–Varying the methodology to test whether the effect can be generalized (eg. deliberately sampling from a different population than the original study)

Replication: What is it good for?

- There is a consensus that replication is a hallmark of scientific knowledge

Replication: What is it good for?

- There is a consensus that replication is a hallmark of scientific knowledge
- Intuitively, this makes sense

Replication: What is it good for?

- There is a consensus that replication is a hallmark of scientific knowledge
- Intuitively, this makes sense
  - When doing science, we are interested in finding general principles

Replication: What is it good for?

- There is a consensus that replication is a hallmark of scientific knowledge
- Intuitively, this makes sense
  - When doing science, we are interested in finding general principles
  - We aren't interested in one-off, or quirky results

Replication: What is it good for?

- There is a consensus that replication is a hallmark of scientific knowledge
- Intuitively, this makes sense
  - When doing science, we are interested in finding general principles
  - We aren't interested in one-off, or quirky results
  - Results that replicate are more likely to be discovering general principles

Replication: What is it good for?

- There is a consensus that replication is a hallmark of scientific knowledge
- Intuitively, this makes sense
  - When doing science, we are interested in finding general principles
  - We aren't interested in one-off, or quirky results
  - Results that replicate are more likely to be discovering general principles
  - This is especially true with respect to indirect replications

Replication: What is it good for?

A Bayesian account (due to Earp & Tramifow 2015):

- Let T=the result of the original study is true

Replication: What is it good for?

A Bayesian account (due to Earp & Tramifow 2015):

- Let T=the result of the original study is true
- Let F=the replication attempt fails

Replication: What is it good for?

A Bayesian account (due to Earp & Tramifow 2015):

- Let T=the result of the original study is true
- Let F=the replication attempt fails
- Consider the Odds form of Bayes' Theorem:

$$\frac{P(T|F)}{P(\neg T|F)} = \frac{P(T)}{P(\neg T)} \cdot \frac{P(F|T)}{P(F|\neg T)}$$

Replication: What is it good for?

A Bayesian account (due to Earp & Tramifow 2015):

- Let T=the result of the original study is true
- Let F=the replication attempt fails
- Consider the Odds form of Bayes' Theorem:

$$\frac{P(T|F)}{P(\neg T|F)} = \frac{P(T)}{P(\neg T)} \cdot \frac{P(F|T)}{P(F|\neg T)}$$

- Suppose, we are initially very confident that the result is true:

$$\frac{P(T|F)}{P(\neg T|F)} = 50 \cdot \frac{P(F|T)}{P(F|\neg T)}$$

Replication: What is it good for?

- And suppose we are twice as confident that the replication would fail if the original result is false than if it were true:

$$\frac{P(T|F)}{P(\neg T|F)} = 50 \cdot 0.5$$

Replication: What is it good for?

- And suppose we are twice as confident that the replication would fail if the original result is false than if it were true:

$$\frac{P(T|F)}{P(\neg T|F)} = 50 \cdot 0.5$$

- Then, our posterior odds:

$$\frac{P(T|F)}{P(\neg T|F)} = 25$$

Replication: What is it good for?

- And suppose we are twice as confident that the replication would fail if the original result is false than if it were true:

$$\frac{P(T|F)}{P(\neg T|F)} = 50 \cdot 0.5$$

- Then, our posterior odds:

$$\frac{P(T|F)}{P(\neg T|F)} = 25$$

- The key point, is that the likelihood ratio ($\frac{P(F|T)}{P(F|\neg T)}$) will always be less than 1

Replication: What is it good for?

- And suppose we are twice as confident that the replication would fail if the original result is false than if it were true:

$$\frac{P(T|F)}{P(\neg T|F)} = 50 \cdot 0.5$$

- Then, our posterior odds:

$$\frac{P(T|F)}{P(\neg T|F)} = 25$$

- The key point, is that the likelihood ratio ($\frac{P(F|T)}{P(F|\neg T)}$) will always be less than 1
- So, in principle, successive failures to replicate can result in our having more confidence that the original study is false than that it is true (What would the posterior odds be if we reiterated this process 6 times?)

Understanding the Results of Replication Studies

- So, what does a failed replication tell us?

Understanding the Results of Replication Studies

- So, what does a failed replication tell us?
  - That the original result was false?

Understanding the Results of Replication Studies

- So, what does a failed replication tell us?
  - That the original result was false?
  - Not really

Understanding the Results of Replication Studies

- So, what does a failed replication tell us?
  - That the original result was false?
  - Not really
  - But it should undermine our confidence in the result, to a degree

Understanding the Results of Replication Studies

- So, what does a failed replication tell us?
  - That the original result was false?
  - Not really
  - But it should undermine our confidence in the result, to a degree
- And what does a successful replication tell us?

Understanding the Results of Replication Studies

- So, what does a failed replication tell us?
  - That the original result was false?
  - Not really
  - But it should undermine our confidence in the result, to a degree
- And what does a successful replication tell us?
  - That the original result was 100% certainly correct?

Understanding the Results of Replication Studies

- So, what does a failed replication tell us?
  - That the original result was false?
  - Not really
  - But it should undermine our confidence in the result, to a degree
- And what does a successful replication tell us?
  - That the original result was 100% certainly correct?
  - Not really (consider question 2 of the problem set)

Understanding the Results of Replication Studies

- So, what does a failed replication tell us?
  - That the original result was false?
  - Not really
  - But it should undermine our confidence in the result, to a degree
- And what does a successful replication tell us?
  - That the original result was 100% certainly correct?
  - Not really (consider question 2 of the problem set)
  - But it should bolster our confidence in the result, to a degree

Understanding the Results of Replication Studies

- So, what does a failed replication tell us?
  - That the original result was false?
  - Not really
  - But it should undermine our confidence in the result, to a degree
- And what does a successful replication tell us?
  - That the original result was 100% certainly correct?
  - Not really (consider question 2 of the problem set)
  - But it should bolster our confidence in the result, to a degree
- "Humans desire certainty, and science infrequently provides it. As much as we might wish it to be otherwise, a single study almost never provides definitive resolution for or against an effect and its explanation" (OSC p.4716-7)

Studying Replication

- Given that replication is so important, it is surprising that it was largely under-studied (until recently)

Studying Replication

- Given that replication is so important, it is surprising that it was largely under-studied (until recently)
- How do we measure how well we are doing with respect to replicability of results if we are not paying attention to it?

Studying Replication

- Given that replication is so important, it is surprising that it was largely under-studied (until recently)
- How do we measure how well we are doing with respect to replicability of results if we are not paying attention to it?
- This has changed in the past decade or so

Studying Replication

- Given that replication is so important, it is surprising that it was largely under-studied (until recently)
- How do we measure how well we are doing with respect to replicability of results if we are not paying attention to it?
- This has changed in the past decade or so
- There have been large-scale "meta-scientific" studies regarding the replication of results in different areas, including psychology

Studying Replication

- Given that replication is so important, it is surprising that it was largely under-studied (until recently)
- How do we measure how well we are doing with respect to replicability of results if we are not paying attention to it?
- This has changed in the past decade or so
- There have been large-scale "meta-scientific" studies regarding the replication of results in different areas, including psychology
- How are we doing with respect to replication?

Open Science Collaboration

This is a very widely cited study that aimed to estimate replication in psychology:

- Methodology:

Open Science Collaboration

This is a very widely cited study that aimed to estimate replication in psychology:

- Methodology:
  - They selected studies from a single year in three main journals

Open Science Collaboration

This is a very widely cited study that aimed to estimate replication in psychology:

- Methodology:
  - They selected studies from a single year in three main journals
  - Teams of researchers selected a study to attempt to replicate

Open Science Collaboration

This is a very widely cited study that aimed to estimate replication in psychology:

- Methodology:
  - They selected studies from a single year in three main journals
  - Teams of researchers selected a study to attempt to replicate
  - The original researchers were consulted

Open Science Collaboration

This is a very widely cited study that aimed to estimate replication in psychology:

- Methodology:
  - They selected studies from a single year in three main journals
  - Teams of researchers selected a study to attempt to replicate
  - The original researchers were consulted
  - A key statistical result from each study was selected as the focus of replication

Open Science Collaboration

This is a very widely cited study that aimed to estimate replication in psychology:

- Results were measured according to five indicators:

Open Science Collaboration

This is a very widely cited study that aimed to estimate replication in psychology:

- Results were measured according to five indicators:
  1. Significance

Open Science Collaboration

This is a very widely cited study that aimed to estimate replication in psychology:

- Results were measured according to five indicators:
  1. Significance
  2. P-values

Open Science Collaboration

This is a very widely cited study that aimed to estimate replication in psychology:

- Results were measured according to five indicators:
  1. Significance
  2. P-values
  3. Effect size

Open Science Collaboration

This is a very widely cited study that aimed to estimate replication in psychology:

- Results were measured according to five indicators:
  1. Significance
  2. P-values
  3. Effect size
  4. Subjective assessment by the replication team

Open Science Collaboration

This is a very widely cited study that aimed to estimate replication in psychology:

- Results were measured according to five indicators:
  1. Significance
  2. P-values
  3. Effect size
  4. Subjective assessment by the replication team
  5. Meta-analyses of effect sizes

1) Results by significance testing:

- Nearly all of the original studies claimed statistically significant results

1) Results by significance testing:

- Nearly all of the original studies claimed statistically significant results
- The researchers estimated that 89% studies would replicate if the original results were correct

1) Results by significance testing:

- Nearly all of the original studies claimed statistically significant results
- The researchers estimated that 89% studies would replicate if the original results were correct
- The actual number was only 35%

1) Results by significance testing:

- Nearly all of the original studies claimed statistically significant results
- The researchers estimated that 89% studies would replicate if the original results were correct
- The actual number was only 35%
- Weakness: this is a pass/fail standard, and doesn't measure how far from the original findings the replication attempts were

2) Results by difference in P-values:

- Here they measured how many studies came within 95% confidence interval of the attempted replication

2) Results by difference in P-values:

- Here they measured how many studies came within 95% confidence interval of the attempted replication
- This was to avoid the problem mentioned with the first method

2) Results by difference in P-values:

- Here they measured how many studies came within 95% confidence interval of the attempted replication
- This was to avoid the problem mentioned with the first method
- By this metric, 47.4% of studies were successfully replicated

3) Results by comparing effect sizes:

- Effect size–the strength of the relationship between two variables

3) Results by comparing effect sizes:

- Effect size–the strength of the relationship between two variables
- When they compared the effect size they found:

3) Results by comparing effect sizes:

- Effect size–the strength of the relationship between two variables
- When they compared the effect size they found:
  - The original results and replication results were positively correlated (so that's the good news)

3) Results by comparing effect sizes:

- Effect size–the strength of the relationship between two variables
- When they compared the effect size they found:
  - The original results and replication results were positively correlated (so that's the good news)
  - However, the original results tended to report larger effect sizes than the replication attempts

3) Results by comparing effect sizes:

- Effect size–the strength of the relationship between two variables
- When they compared the effect size they found:
  - The original results and replication results were positively correlated (so that's the good news)
  - However, the original results tended to report larger effect sizes than the replication attempts
- This suggests that the original findings were in the right direction, but amplified

4) Results by subjective assessment:

- The replicators were asked if the study had successfully replicated

4) Results by subjective assessment:

- The replicators were asked if the study had successfully replicated
- These results were nearly identical to the first method (significance results)

4) Results by subjective assessment:

- The replicators were asked if the study had successfully replicated
- These results were nearly identical to the first method (significance results)
- This would seem to indicate that significance tests are highly regarded as a "gold standard" for research (maybe too highly regarded...we'll look at that next week)

5) Results by meta-analysis:

- Rather than counting individual studies according to whether they met some standard, this method combines the results of both studies (original and replication) and measures the confidence interval of the result

5) Results by meta-analysis:

- Rather than counting individual studies according to whether they met some standard, this method combines the results of both studies (original and replication) and measures the confidence interval of the result
- By this metric, 68% of the studies were successful

5) Results by meta-analysis:

- Rather than counting individual studies according to whether they met some standard, this method combines the results of both studies (original and replication) and measures the confidence interval of the result
- By this metric, 68% of the studies were successful
- One problem with this method of testing is the potential for publication bias to have suppressed data that would be relevant for a complete meta-analysis

Summing up

- So...how are we doing with respect to replicating scientific results?

Summing up

- So...how are we doing with respect to replicating scientific results?
- 

Sources of the problem

We can identify two broad categories of proposed sources of this problem:

- Systemic problems

Sources of the problem

We can identify two broad categories of proposed sources of this problem:

- Systemic problems
  - **Publication bias**

Sources of the problem

We can identify two broad categories of proposed sources of this problem:

- Systemic problems
  - **Publication bias**
  - Over-reliance on statistical significance tests

Sources of the problem

We can identify two broad categories of proposed sources of this problem:

- Systemic problems
  - **Publication bias**
  - Over-reliance on statistical significance tests
  - **The incentive structure to publish innovative results**

Sources of the problem

We can identify two broad categories of proposed sources of this problem:

- Systemic problems
  - **Publication bias**
  - Over-reliance on statistical significance tests
  - **The incentive structure to publish innovative results**
- Individual problems (Questionable Research Practices)

Sources of the problem

We can identify two broad categories of proposed sources of this problem:

- Systemic problems
  - **Publication bias**
  - Over-reliance on statistical significance tests
  - **The incentive structure to publish innovative results**
- Individual problems (Questionable Research Practices)
  - P-hacking

Sources of the problem

We can identify two broad categories of proposed sources of this problem:

- Systemic problems
  - **Publication bias**
  - Over-reliance on statistical significance tests
  - **The incentive structure to publish innovative results**
- Individual problems (Questionable Research Practices)
  - P-hacking
  - **HARKing**

Sources of the problem

We can identify two broad categories of proposed sources of this problem:

- Systemic problems
  - **Publication bias**
  - Over-reliance on statistical significance tests
  - **The incentive structure to publish innovative results**
- Individual problems (Questionable Research Practices)
  - P-hacking
  - **HARKing**
  - Outright fraud

Systemic Problems

Publication Bias

- Some journals explicitly do not publish replication studies

Publication Bias

- Some journals explicitly do not publish replication studies
- What's the motivation to try a replication study if it has 0 chance of getting published?

Publication Bias

- Some journals explicitly do not publish replication studies
- What's the motivation to try a replication study if it has 0 chance of getting published?
- "Journal reviewers and editors may dismiss a new test of a published idea as unoriginal. The claim that 'we already know this' belies the uncertainty of scientific evidence..."

Publication Bias

- Some journals explicitly do not publish replication studies
- What's the motivation to try a replication study if it has 0 chance of getting published?
- "Journal reviewers and editors may dismiss a new test of a published idea as unoriginal. The claim that 'we already know this' belies the uncertainty of scientific evidence..."
- "...Innovation points out paths that are possible; replication points out paths that are likely. Progress relies on both." (OSC, p.943)

Publication Bias

- Some journals explicitly do not publish replication studies
- What's the motivation to try a replication study if it has 0 chance of getting published?
- "Journal reviewers and editors may dismiss a new test of a published idea as unoriginal. The claim that 'we already know this' belies the uncertainty of scientific evidence..."
- "...Innovation points out paths that are possible; replication points out paths that are likely. Progress relies on both." (OSC, p.943)
- This reminds me of a distinction from our old friend Reichenbach...

Publication Bias

- Some journals explicitly do not publish replication studies
- What's the motivation to try a replication study if it has 0 chance of getting published?
- "Journal reviewers and editors may dismiss a new test of a published idea as unoriginal. The claim that 'we already know this' belies the uncertainty of scientific evidence..."
- "...Innovation points out paths that are possible; replication points out paths that are likely. Progress relies on both." (OSC, p.943)
- This reminds me of a distinction from our old friend Reichenbach...
- Context of discovery vs context of justification!

Significance Testing

- Some suggest that $P < 0.05$ is too relaxed of a threshold, and that the benchmark should be $P < 0.01$

Significance Testing

- Some suggest that $P < 0.05$ is too relaxed of a threshold, and that the benchmark should be $P < 0.01$
- Others suggest we need to pay more attention to other statistical factors

Significance Testing

- Some suggest that $P < 0.05$ is too relaxed of a threshold, and that the benchmark should be $P < 0.01$
- Others suggest we need to pay more attention to other statistical factors
  - Effect size

Significance Testing

- Some suggest that $P < 0.05$ is too relaxed of a threshold, and that the benchmark should be $P < 0.01$
- Others suggest we need to pay more attention to other statistical factors
  - Effect size
  - Statistical power

Significance Testing

- Some suggest that $P < 0.05$ is too relaxed of a threshold, and that the benchmark should be $P < 0.01$
- Others suggest we need to pay more attention to other statistical factors
  - Effect size
  - Statistical power
- Others suggest that NHST needs to be replaced by Baysian statistical inference

Significance Testing

- Some suggest that $P < 0.05$ is too relaxed of a threshold, and that the benchmark should be $P < 0.01$
- Others suggest we need to pay more attention to other statistical factors
  - Effect size
  - Statistical power
- Others suggest that NHST needs to be replaced by Baysian statistical inference
- We will be looking at this issue more closely next week

Questionable Research Practices

HARKing

Hypothesising After Results are Known

- Recall the HD model of theory confirmation:

HARKing

Hypothesising After Results are Known

- Recall the HD model of theory confirmation:
  - Hypothesis predicts something will be observed

HARKing

Hypothesising After Results are Known

- Recall the HD model of theory confirmation:
  - Hypothesis predicts something will be observed
  - The prediction comes true

HARKing

Hypothesising After Results are Known

- Recall the HD model of theory confirmation:
  - Hypothesis predicts something will be observed
  - The prediction comes true
  - Hypothesis is confirmed

HARKing

Hypothesising After Results are Known

- Recall the HD model of theory confirmation:
  - Hypothesis predicts something will be observed
  - The prediction comes true
  - Hypothesis is confirmed
- We noted that this is not a valid inference

- Recall Popper's adaptation:

- Recall Popper's adaptation:
  - Hypothesis predicts something will be observed

- Recall Popper's adaptation:
  - Hypothesis predicts something will be observed
  - Prediction fails

- Recall Popper's adaptation:
  - Hypothesis predicts something will be observed
  - Prediction fails
  - Hypothesis is disconfirmed

- Recall Popper's adaptation:
  - Hypothesis predicts something will be observed
  - Prediction fails
  - Hypothesis is disconfirmed
- We noted that there were still some issues here

- Recall Popper's adaptation:
  - Hypothesis predicts something will be observed
  - Prediction fails
  - Hypothesis is disconfirmed
- We noted that there were still some issues here
- Underdetermination: we don't know whether it is the hypothesis or some background assumption that has been disconfirmed

- Recall Popper's adaptation:
  - Hypothesis predicts something will be observed
  - Prediction fails
  - Hypothesis is disconfirmed
- We noted that there were still some issues here
- Underdetermination: we don't know whether it is the hypothesis or some background assumption that has been disconfirmed
- Recall the example of the discovery of Neptune

The discovery of Neptune is actually an example of HARKing:

- Newton's laws of motion (together with some assumptions) predicted the orbit of Uranus

The discovery of Neptune is actually an example of HARKing:

- Newton's laws of motion (together with some assumptions) predicted the orbit of Uranus
- There were observed anomalies between the predictions and observations

The discovery of Neptune is actually an example of HARKing:

- Newton's laws of motion (together with some assumptions) predicted the orbit of Uranus
- There were observed anomalies between the predictions and observations
- Airy HARKed that Newton's inverse square law was breaking down over long distances

The discovery of Neptune is actually an example of HARKing:

- Newton's laws of motion (together with some assumptions) predicted the orbit of Uranus
- There were observed anomalies between the predictions and observations
- Airy HARKed that Newton's inverse square law was breaking down over long distances
- Le Verrier HARKed that there was an 8th planet causing these anomalies

The discovery of Neptune is actually an example of HARKing:

- Newton's laws of motion (together with some assumptions) predicted the orbit of Uranus
- There were observed anomalies between the predictions and observations
- Airy HARKed that Newton's inverse square law was breaking down over long distances
- Le Verrier HARKed that there was an 8th planet causing these anomalies
- Galle verified that Le Verrier's hypothesis was indeed correct

The discovery of Neptune is actually an example of HARKing:

- Newton's laws of motion (together with some assumptions) predicted the orbit of Uranus
- There were observed anomalies between the predictions and observations
- Airy HARKed that Newton's inverse square law was breaking down over long distances
- Le Verrier HARKed that there was an 8th planet causing these anomalies
- Galle verified that Le Verrier's hypothesis was indeed correct
- So what's the problem with HARKing?

Prediction vs Accommodation

There is actually a long-standing debate in philosophy of science about the value of prediction over accommodation

- The debate dates back to the 18 and 19th centuries

Prediction vs Accommodation

There is actually a long-standing debate in philosophy of science about the value of prediction over accommodation

- The debate dates back to the 18 and 19th centuries
- Herschel and Whewell argued that theories that made novel predictions that came true could not have done so by chance

Prediction vs Accommodation

There is actually a long-standing debate in philosophy of science about the value of prediction over accommodation

- The debate dates back to the 18 and 19th centuries
- Herschel and Whewell argued that theories that made novel predictions that came true could not have done so by chance
- Mill, and later Keynes, opposed this view, thinking that "hypothesizing" was mere speculation

Prediction vs Accommodation

There is actually a long-standing debate in philosophy of science about the value of prediction over accommodation

- The debate dates back to the 18 and 19th centuries
- Herschel and Whewell argued that theories that made novel predictions that came true could not have done so by chance
- Mill, and later Keynes, opposed this view, thinking that "hypothesizing" was mere speculation
- On their view, so long as the theory accommodated the observations, the temporal order did not matter

Prediction vs Accommodation

There is actually a long-standing debate in philosophy of science about the value of prediction over accommodation

- The debate dates back to the 18 and 19th centuries
- Herschel and Whewell argued that theories that made novel predictions that came true could not have done so by chance
- Mill, and later Keynes, opposed this view, thinking that "hypothesizing" was mere speculation
- On their view, so long as the theory accommodated the observations, the temporal order did not matter
- We know which side Popper took in this debate

Prediction vs Accommodation

There is actually a long-standing debate in philosophy of science about the value of prediction over accommodation

- The debate dates back to the 18 and 19th centuries
- Herschel and Whewell argued that theories that made novel predictions that came true could not have done so by chance
- Mill, and later Keynes, opposed this view, thinking that "hypothesizing" was mere speculation
- On their view, so long as the theory accommodated the observations, the temporal order did not matter
- We know which side Popper took in this debate
- Kerr sides pretty strongly with Popper

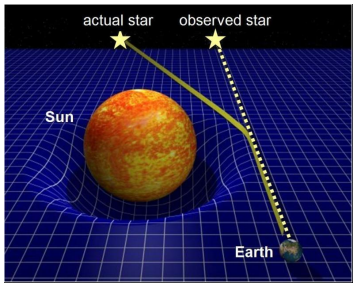
Prediction vs Accommodation

There is actually a long-standing debate in philosophy of science about the value of prediction over accommodation

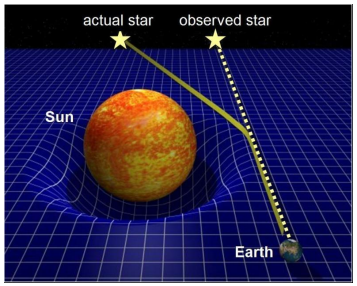
- The debate dates back to the 18 and 19th centuries
- Herschel and Whewell argued that theories that made novel predictions that came true could not have done so by chance
- Mill, and later Keynes, opposed this view, thinking that "hypothesizing" was mere speculation
- On their view, so long as the theory accommodated the observations, the temporal order did not matter
- We know which side Popper took in this debate
- Kerr sides pretty strongly with Popper
- Contemporary Bayesians typically side with Mill and Keynes

Recall an Example from Week 1:

- General relativity **predicts** that gravity bends light

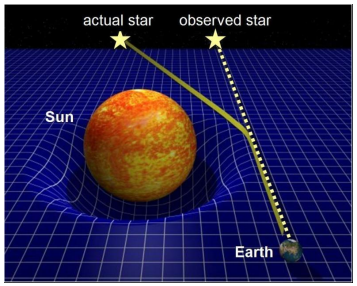


Recall an Example from Week 1:



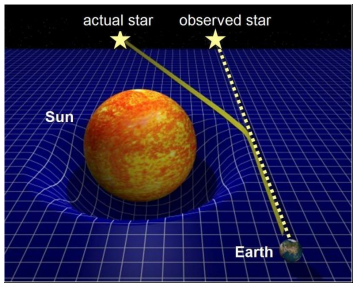
- General relativity **predicts** that gravity bends light
- This effect has been observed, initially by Eddington in 1919

Recall an Example from Week 1:



- General relativity **predicts** that gravity bends light
- This effect has been observed, initially by Eddington in 1919
- General relativity also **accommodates** all the previously observed data with respect to the position of the planets

Recall an Example from Week 1:



- General relativity **predicts** that gravity bends light
- This effect has been observed, initially by Eddington in 1919
- General relativity also **accommodates** all the previously observed data with respect to the position of the planets
- Does this not count in favour of the theory?

So, maybe HARKing isn't bad in itself...

- What is bad is when it is not acknowledged for what it is

So, maybe HARKing isn't bad in itself...

- What is bad is when it is not acknowledged for what it is
- This would be like if Le Verrier had presented himself having predicted the anomalies, and then took them as definitive proof for the existence of the 8th planet

So, maybe HARKing isn't bad in itself...

- What is bad is when it is not acknowledged for what it is
- This would be like if Le Verrier had presented himself having predicted the anomalies, and then took them as definitive proof for the existence of the 8th planet
- Rather, Le Verrier presented his hypothesis as needing more testing

So, maybe HARKing isn't bad in itself...

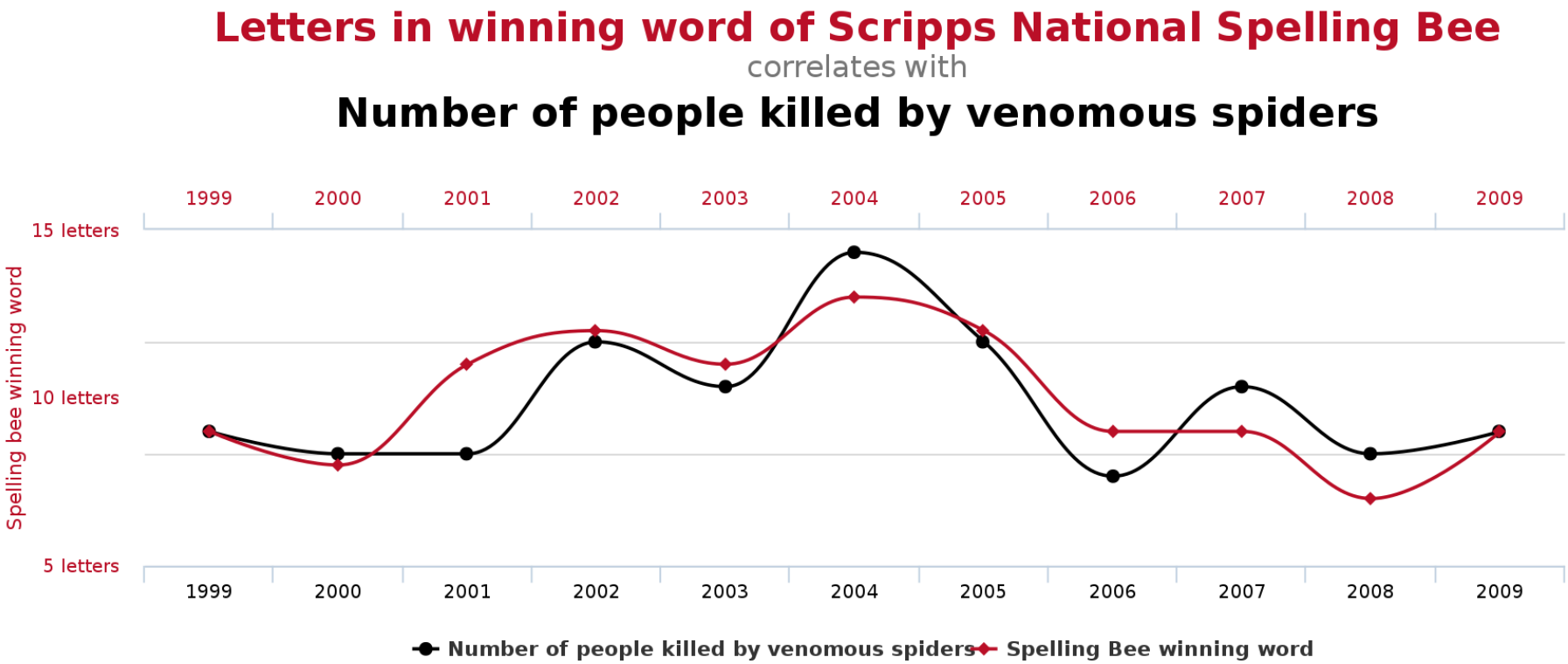
- What is bad is when it is not acknowledged for what it is
- This would be like if Le Verrier had presented himself having predicted the anomalies, and then took them as definitive proof for the existence of the 8th planet
- Rather, Le Verrier presented his hypothesis as needing more testing
- This is a subtle, but significant difference

So, maybe HARKing isn't bad in itself...

- What is bad is when it is not acknowledged for what it is
- This would be like if Le Verrier had presented himself having predicted the anomalies, and then took them as definitive proof for the existence of the 8th planet
- Rather, Le Verrier presented his hypothesis as needing more testing
- This is a subtle, but significant difference
- Kerr: HARKing belongs to the context of discovery more than the context of justification

So, maybe HARKing isn't bad in itself...

- What is bad is when it is not acknowledged for what it is
- This would be like if Le Verrier had presented himself having predicted the anomalies, and then took them as definitive proof for the existence of the 8th planet
- Rather, Le Verrier presented his hypothesis as needing more testing
- This is a subtle, but significant difference
- Kerr: HARKing belongs to the context of discovery more than the context of justification
- Why? Since the studies were not designed to test the HARKed hypothesis, the correlations noted could just be spurious



Source: "Spurious Correlations" by Tyler Vigen

The problem isn't so much that HARKing is *post hoc* (translation: after this), but *ad hoc* (translation: for this)

- Ad hoc hypotheses are not derived from a theory

The problem isn't so much that HARKing is *post hoc* (translation: after this), but *ad hoc* (translation: for this)

- Ad hoc hypotheses are not derived from a theory
- Rather, they are added to a theory to save it from being falsified

The problem isn't so much that HARKing is *post hoc* (translation: after this), but *ad hoc* (translation: for this)

- Ad hoc hypotheses are not derived from a theory
- Rather, they are added to a theory to save it from being falsified
- This doesn't automatically mean an ad hoc hypothesis is false

The problem isn't so much that HARKing is *post hoc* (translation: after this), but *ad hoc* (translation: for this)

- Ad hoc hypotheses are not derived from a theory
- Rather, they are added to a theory to save it from being falsified
- This doesn't automatically mean an ad hoc hypothesis is false
- But, without further testing, these are merely speculative

The problem isn't so much that HARKing is *post hoc* (translation: after this), but *ad hoc* (translation: for this)

- Ad hoc hypotheses are not derived from a theory
- Rather, they are added to a theory to save it from being falsified
- This doesn't automatically mean an ad hoc hypothesis is false
- But, without further testing, these are merely speculative
- They are proposing an explanation for why the theory's prediction failed

The problem isn't so much that HARKing is *post hoc* (translation: after this), but *ad hoc* (translation: for this)

- Ad hoc hypotheses are not derived from a theory
- Rather, they are added to a theory to save it from being falsified
- This doesn't automatically mean an ad hoc hypothesis is false
- But, without further testing, these are merely speculative
- They are proposing an explanation for why the theory's prediction failed
- As Kerr notes, they are often the first explanation that occurs to the researcher

The problem isn't so much that HARKing is *post hoc* (translation: after this), but *ad hoc* (translation: for this)

- Ad hoc hypotheses are not derived from a theory
- Rather, they are added to a theory to save it from being falsified
- This doesn't automatically mean an ad hoc hypothesis is false
- But, without further testing, these are merely speculative
- They are proposing an explanation for why the theory's prediction failed
- As Kerr notes, they are often the first explanation that occurs to the researcher
- What are the chances that the first explanation that occurs to the researcher is correct?

The upshot

HARKing is bad practice:

- Maybe not because prediction is better than accommodation

The upshot

HARKing is bad practice:

- Maybe not because prediction is better than accommodation
- But because it presents a hypothesis as having been tested when it hasn't

The upshot

HARKing is bad practice:

- Maybe not because prediction is better than accommodation
- But because it presents a hypothesis as having been tested when it hasn't
- The experiment that suggested the HARKed hypothesis was not designed to test it (for example, you might need to choose your sample by a different method to test the HARKed hypothesis than you would for the original hypothesis).

Reading for next time:

- Travers et al. "Null Hypothesis Significance Testing and p Values"
- Howson & Urbach "Bayesian Reasoning in Science"
- Romero & Sprenger "Scientific self-correction: the Bayesian way" (it's ok to skim this one)