

Instructor guide: Confirmation Bias unit

This instructor guide is designed as a flexible framework to help you deliver the Confirmation Bias unit effectively.

- Each lesson is summarized with key takeaways and includes cues for additional, step-by-step directions for activities.
- Use the video and supplementary slide-by-slide annotations with speaker notes provided to familiarize yourself with the material, streamline your lesson preparation, and enhance classroom discussions.
- Feel free to adapt and customize the content to fit your teaching style and your students' needs. Get access to the slides [here](#), then navigate to File → Make a Copy to get started.

Overview & introduction

Summary:

This unit on **confirmation bias** explores the subtle yet powerful ways that our predisposition to privilege our prior beliefs can influence every stage of scientific research. By understanding how confirmation bias can distort experimental design, data collection, analysis, and interpretation, students will learn essential strategies to design more rigorous, transparent, and reproducible studies. This unit is ideal for early-career researchers and advanced students who wish to strengthen their critical thinking skills and safeguard against biased reasoning.

Why use this unit:

- This unit on confirmation bias equips students with a fundamental understanding of cognitive biases, especially confirmation bias, which is often the root of many research errors.
- Each lesson blends theoretical insights with practical activities, ensuring that learners not only recognize bias but can also implement strategies to minimize its impact in their work.
- Real-world examples and interactive activities encourage students to reflect on their own decision-making processes and develop more robust research practices.

Lesson overviews

Lesson 1: Our biased brains (note that the lesson begins before the title slide)

Lesson summary:

This lesson introduces confirmation bias as a fundamental error in cognition. It details how initial beliefs influence the gathering and interpretation of information. The classic Wason task (1960) illustrates the pervasive nature of this bias, and highlights the need for critical evaluation in research.

Goal:

Build an intuition around identifying cognitive bias and taking steps to mitigate it.

Activity overview: (~3-5 minutes)

Participants engage in a modified version of the Wason 2-4-6 task) to experience firsthand the important distinction between trying to falsify vs trying to support an initial hypothesis.

Link to activity:

<https://monolith-test-1.vercel.app/activities/hms-cbi-01/>

Step-by-Step activity instructions:

1. **[Instructor] Prompt students to start typing numbers into the 3 input fields.**
 - a. **Remind them that they will either see “TRUE” or “FALSE” when they hit the “TEST” button.**
 - i. **This indicates if their number sequences match the “secret rule” for this session.**
 - b. **Note: Each user will have a random “secret rule,” so if you’d like pairs or teams to work together on the same one, have a representative be the only one who pulls up the activity/have them share a single device for that group.**
2. **Type numbers into the 3 input boxes, then hit the “TEST” button to see if your guess is “true” or “false” in regards to the “secret rule.”**

3. Repeat until you're ready to draft a hypothesis.
4. Click the "GUESS THE NUMBER RULE" button to reveal an input box.
5. Type your hypothesis for what the "secret rule" is into the input box.
6. Click "SUBMIT" to discover what the "secret rule" is and if you guessed correctly.
7. Click "CONTINUE" to review your guesses in relation to the guesses of others.
8. **[Instructor] Spend time with students reviewing their guesses. Make sure that by the end of the discussion, there is a conversation about falsifying your hypothesis.**
 - a. **Use the discussion questions in the unit to help facilitate discussion.**

Activity takeaway:

A hypothesis is only as strong as the effort spent to try and refute it.

Lesson 2: Formulating rigorous hypotheses

Lesson summary:

This lesson focuses on the pitfalls of designing experiments that fail to rigorously test hypotheses. Instructors will discuss the importance of developing falsifiable and specific hypotheses, in order to minimize the opportunities for confirmation bias to unduly influence hypothesis testing.

Goal:

Learn to design specific, falsifiable, and contextual hypotheses so that experiments provide evidence that helps support well-justified conclusions.

Activity overview: (~10 minutes)

Students will work solo or in small groups to improve a hypothesis and study design. The activity guides students to make choices to assess the conditions under which a hypothesized effect will be true.

Link to activity:

<https://monolith-test-1.vercel.app/activities/hms-cbi-02a/>

Step-by-Step activity instructions:

1. **[Instructor], Students will see the research study they are going to improve.**
 - a. **Encourage students to read the displayed Hypothesis and Study Design carefully before proceeding..**
2. The screen displays the original Hypothesis and Study Design.
 - a. Click NEXT to advance to Screen 2.
3. You will see that you are revising the Hypothesis of the research study.
 - a. Multiple revised hypothesis options are provided.
 - b. Guide students to select the hypothesis that best demonstrates the principles of **specificity** and **falsifiability**.
4. Click Submit to confirm the selection.
 - a. Feedback will appear explaining the strengths or weaknesses of the selected hypothesis.
 - b. If you selected the best choice, the selection is disabled and the NEXT button is enabled
 - c. If you selected a different choice, you can revise your choice and then re-submit
5. Once the best hypothesis is selected, click NEXT to advance to Screen 3.
6. You will now revise the Study Design of the research study.
 - a. The original hypothesis will be replaced with the improved hypothesis selected on Screen 2, with an indicator that the hypothesis was improved.
7. Select an improved Study Design
8. Click Submit to confirm the selection.
 - a. Feedback about your choice will appear.
 - b. If you selected the best choice, the selection is disabled and the NEXT button is enabled.
 - c. If you selected a different choice, you can revise your choice and then re-submit.
9. Click NEXT to advance to the Congrats screen.
10. **[Instructor] Lead discussion while on the results screen.**
 - a. **Discuss how changes to the Study Design influenced alignment with the hypothesis**
 - b. **Use the discussion questions in the unit to help facilitate discussion.**

Activity takeaway:

Rigorous research requires a specific, falsifiable hypothesis and a study design that tests it.

Lesson 3: Researcher degrees of freedom

Lesson summary:

This lesson unpacks researcher degrees of freedom. These are the myriad design and analysis choices that quietly steer findings toward a favored conclusion. Using an fMRI case study, it shows how post-hoc region selection (“double dipping”), ill-matched controls, and participant pre-screening each bias results toward supporting a localized neural-correlate hypothesis (H_1). The discussion then broadens to general researcher choices (e.g., when to stop collecting data, which outcomes to highlight, how to handle anomalies) and argues that, without forethought, each becomes a conduit for confirmation bias. Countermeasures include formulating specific, falsifiable hypotheses, pre-registering protocols, masking and randomizing, and transparently reporting all results.

Goal:

Recognize and constrain researcher degrees of freedom so experimental choices illuminate truth instead of amplifying confirmation bias.

Activity overview: (~10 minutes)

In this session, learners are given a case study illustrating various choices that can lead to bias. They will then **select the choice that produces the largest bias in favor of the provided hypothesis, H_1** . If you have a small group, consider having a single person show their screen to all participants so that the group can debate these “would you rather” prompts as they go. If you have a larger group, consider dividing them into small groups.

Heads up!

This activity uses some technical language that may be unfamiliar to those without experience in fMRI research. The point of using this specific example isn't to intimidate you or your students; it's to drive home the point that we can still find rigor

issues even if we aren't deeply familiar with a subject area. Here is a glossary of terms and how they affect hypothesis testing:

- **Multiple comparison correction** does not favor H_1 , but may not be an ideal in obtaining sufficient power for the study if many voxels will be examined.
- **Analyzing the most correlated voxel** strongly favors H_1 because it double-dips, using the same data to determine the parameters of the analysis and to perform the analysis.
- **Stopping data collection early** favors H_1 , because it conditions on the outcome; the amount of data collected will depend on the interim results.
- **A simple control** seems like a good idea, but adding in another kind of picture (landscapes), could be an unnecessary source of noise.
- **Analysis Parameter: Smoothing** favors H_1 because it loosens the parameters of the analysis to find a promising result.
- **Personality screen** strongly favors H_1 because it selects a biased sample that seems aligned with that hypothesis.

Link to activity:

<https://monolith-test-1.vercel.app/activities/hms-cbi-03>

Step-by-Step activity instructions:

1. **[Instructor], you will have the option of sharing the activity with your students via QR Code or a direct link.**
2. Read the H_1 for the fMRI case study: "Emotional pictures produce stronger BOLD responses than neutral pictures in the amygdala."
3. Then read the 2 choices provided and click on the option you believe would most strongly bias the results towards supporting H_1 .
4. Then type out your reasoning for why this action would bias the experiment in favor of H_1 in the input box and click the "ROUND 2" button.
5. Repeat, then click the "ROUND 3" button.
6. Repeat, then click the "SUBMIT" button.
7. Compare your results with those of your peers on the final page. You can toggle through everyone's answers and discuss your thought process.
8. **[Instructor] Lead discussion while on the results screen. Be sure to toggle between the different responses provided in each round.**

- a. **Hover over the visualization to see a pop-up with the text description of the choices.**
- b. **Scroll up and down to reveal student responses.**
- c. **Use the discussion questions in the unit to help facilitate discussion.**

Activity takeaway:

Some well-intended research practices introduce bias into our experiments.

Lesson 4: Mitigating bias through masking

Lesson summary:

This lesson explains masking (also called blinding) as a crucial method to reduce observer bias in experiments. Practical examples and step-by-step protocols demonstrate how rigorous masking improves accuracy and reliability of research findings. Students are invited to explore the problematic impact bias has on distorting the effect size of experiment results.

Goal:

Explore the design and implementation of rigorous masking protocols to ensure data collection and analysis remain free from observer bias, a specific type of confirmation bias.

Activity overview: (~1-2 minutes)

Students will **explore the impact of bias on spurious results** in scientific research.

Heads up!

This activity uses some technical terms. Here are some quick definitions for each of these terms:

- **Sample Size:** The number of observations or data points included in a study or experiment. Note that sample size should be reported for all levels: typically, the numbers of subjects, assays, and datapoints. Larger sample sizes generally increase the reliability and statistical power of a study.
- **Bias Amount:** A measure of systematic error introduced into a study, which can distort the results away from the true effect.

- **True Effect Size (d):** A quantitative measure of the magnitude of the difference or relationship being studied. A larger effect size means a stronger relationship or difference between groups.
- **Probability of detecting a significant result:** The likelihood that an observed result is statistically significant, meaning it is unlikely to have occurred by random chance.

Link to activity:

<https://monolith-test-1.vercel.app/activities/hms-cbi-04/>

Step-by-Step activity instructions:

1. **[Instructor] It might make the most sense to treat this activity as an extended discussion.**
 - a. **Students can still work solo, but the exploration works best with prompting about how different configurations influence effect size while segueing into experiences students have had in their labs or when reading published papers.**
2. You are provided with two sliding bars. One represents "Sample size" and the other represents "Bias" in a given experiment.
3. Click on each dot to move it along its respective sliders.
4. Try different pairings to see how these factors distort the relationship between true effect size and the probability of statistical significance.
5. The relationships you explore will be visualized in the graph below the 2 sliders.
6. **[Instructor] If you haven't been a part of the discussion thus far, make sure you lead a brief discussion, and connect this exploration with the next segment of the unit, showcasing the impact of not masking and not randomizing on effect sizes.**

Activity takeaway:

A failure to mask can introduce substantial risk of spurious results.

Lesson 5: How good is your mask?

Lesson summary:

This lesson explores the challenges associated with accidental unmasking and the assessment of masking effectiveness in experimental settings. Strategies for identifying subtle cues and preventing unmasking are discussed to ensure that research outcomes remain as unbiased as possible. Students are given the opportunity to develop a take-home tool for use in their own research.

Goal:

Develop an awareness of potential unmasking cues and how to safeguard experimental outcomes through effective masking.

Activity overview: (~12-15 minutes)

Participants will review an experimental scenario and identify practices with the potential to lead to accidental unmasking. They will then propose solutions to improve masking procedures. Make sure you give all participants ample time to read the experiment, and encourage them to use the hint function to proceed if they get stuck.

Link to activity:

<https://monolith-test-1.vercel.app/activities/hms-cbi-05/>

Step-by-Step activity instructions:

1. **[Instructor] Students will have the option to complete this activity as an individual or group.**
 - a. **If you want to run it for a class to only see their peers' results, click the "AS A GROUP" button to receive a session link that each member of the class will need to use to have their results coordinate on the results page.**
 - b. **If students select the "AS AN INDIVIDUAL" button, the results page at the end will show results from "global users" outside of your class.**
2. Review the details of the behavioral study on sleep patterns and, solo or with your group, think through all the ways this study might be accidentally unmasked.
3. Type your observations into the input box.

4. **[Instructor] Prompt students to use the hint feature and chat about options with each other to continue thinking about risk points to unmask.**
5. Click "Get Hint" whenever you feel stuck to help you think of different ways the mask might fail.
6. When you are done, click the "NEXT" button.
7. Next, along the left side of the screen, you will see the responses of other users along with your own.
8. **[Instructor] Use this results page as a brief opportunity for discussion where students share about their choices and notes. How and why did they make those decisions?**
9. On the right side of the screen, you can supply advice about different facets of the study, the Study Team, the Participants, the Data Accumulation, or the Environmental Conditions. You can also skip giving advice and click "Open Checklist" directly.
10. **[Instructor] Remind students that offering advice is optional, but be sure to give students more time if they choose to do this.**
11. If you do provide advice, you will be taken to a screen to review advice you and other users have given. You can then click the "Open Checklist" button.
12. **[Instructor] For the checklist, prompt students to take a moment to consider what they will really need to remember/self-check in the future for their own experiments.**
13. On the checklist page, feel free to add custom checklist items to help you evaluate your own research using the text entry and "Add" button.
14. When you are ready, click the "Download PDF" button to save this list as a PDF.
15. **[Instructor] This activity is fairly discussion-heavy, but feel free to wrap up the activity with any final thoughts or observations before segueing back into the lesson that picks up on the question of formally assessing if an experiment was properly masked.**

Activity takeaway:

Masking issues can arise at multiple places throughout studies, but careful critique and creative solutions can overcome these issues.

Lesson 6: Analytical practices to mitigate bias

Lesson summary:

This lesson shifts the focus from data collection to analysis, and reveals how confirmation bias lurks in every post-collection decision. Using the StudentLife dataset as a sandbox, learners experience how easy it is to unearth a “significant” correlation and then face the subtle distinction between a descriptive pattern and a causal claim. The lesson contrasts exploratory and confirmatory research, highlights questionable practices (like HARKing, p-hacking, and the garden of forking paths), and offers antidotes like clear labeling of exploratory findings, pre-registered analysis plans, and transparent reporting.

Goal:

Master analytic discipline (distinguishing exploratory curiosity from confirmatory testing) to prevent your data-driven insights from drifting into bias-driven illusions.

Activity overview: (~5 min)

Students will work with a sample dataset and one of two preset hypotheses. They will evaluate the dataset to see if it supports their given hypothesis. They will ultimately **discuss the implications** of making analytic choices with a pre-determined hypothesis.

Heads up!

All participants will be randomly assigned to one of two groups. Half will be asked to demonstrate that daily activities *do* influence student outcomes, while the other will demonstrate that daily activities *do not* influence student outcomes.

Link to activity:

<https://monolith-test-1.vercel.app/activities/hms-cbi-06/>

Step-by-Step activity instructions:

1. **[Instructor] Students will be given 1 of 2 possible hypotheses to inform their thinking. If possible, let students stay in the dark about the existence of a counter-hypothesis their peers might have. Part of the “twist” of the**

experience is to see how far students will go from exploratory thinking to confirmatory thinking.

2. You'll see a hypothesis at the top of the screen in relation to the broader research question: "Do daily activities affect student outcomes?"
3. Looking at the visualization of the data provided, use the radio buttons in the 2 categories (Student outcomes and daily activities) to see what types of relationships you notice between a given activity and an outcome.
4. **[Instructor] While students are working, gently model/prompt the students to explore different choices and notice how selecting various combinations reveals different relationships.**
5. When you've noticed a relationship that connects to your given hypothesis, leave the radio dials on those 2 items, and then click the "SUBMIT" button.
6. Next, you will be able to compare your input with your peers.
7. **[Instructor] When you lead a discussion, point out that different students had conflicting hypotheses. Discuss what connections they saw.**
 - a. **If they selected choices that deliberately "demonstrated" that their hypothesis was "true," discuss how they decided to bypass selections that revealed data that countered their hypothesis.**
 - b. **Use the discussion questions in the unit to help facilitate discussion.**

Activity takeaway:

It is easy to find a pattern in data when one goes looking for it. This practice is very important for exploratory analysis, but becomes problematic when conclusions are drawn.

Lesson 7: Data masking in machine learning models

Lesson summary:

This lesson shows that delegating analysis to algorithms does not free us from confirmation bias. In fact, machine-learning models can easily exploit any hint of data leakage. Using a Parkinson's-disease detector as a case study, the lesson shows how evaluating a model on the same patients used for training produces inflated accuracy. It is effectively a machine-learning analogue of unmasked treatment groups. The counter is rigorous data partitioning (i.e., training, validation,

and test sets with zero overlap and representative variability) in addition to vigilance against subtler leaks (like iterative hyper-parameter tuning on the test set, duplicated samples, temporal dependence, or information-rich data augmentations). Without such safeguards, models seem to confirm our hopes in the lab yet falter in reality.

Goal:

Learn to detect and block data leakage so your machine-learning models deliver honest, generalizable performance rather than echoing built-in bias.

Activity overview: (~3 min)

Students will interact with a basic simulation of a machine learning experiment for detecting a condition like Parkinson's Disease, to understand the impacts of leakage on model results.

Link to activity:

<https://monolith-test-1.vercel.app/activities/hms-cbi-07>

Step-by-Step activity instructions:

1. **[Instructor] Students can work together or solo.**
 - a. **Explain that each subset (A, B, C, D) represents data from 10 patients.**
 - b. **Remind students that toggling a button assigns that subset to use in the Training or Testing set.**
 - c. **Emphasize that their selections directly determine how the model is built and evaluated.**
2. Under Model Setup, take note of the Training Set and Test Set sections.
 - a. Each section contains four toggle buttons labeled A, B, C, D.
 - b. Toggling a button on includes that subset of data in the selected set.
 - c. Toggling it off removes it from that set.
3. In the first phase, your goal is to maximize the model's performance.
 - a. Select all four subsets (A, B, C, D) under the Training Set.
 - b. Select at least one subset (A, B, C, or D) under the Test Set.
 - c. When a valid selection is made, the Train Model button will become enabled.

4. Click the Train Model button to build your model.
 - a. The model's performance results will appear in the Results table.
5. You will receive a prompt to click NEXT to proceed.
 - a. If desired, click Reset to clear your selections.
 - b. You may make alternative Model Setup selections and train another model.
 - c. Each additional trained model will add a new entry to the Results table.
6. In the second phase, your goal is to perform a proper train/test split.
 - a. Select at least one subset for the Training Set.
 - b. Select at least one subset for the Test Set.
 - c. Ensure there is no overlap between Training and Testing subsets.
 - d. When a valid, non-overlapping selection is made, the Train Model button will become enabled.
7. Click the Train Model button.
 - a. The new model's performance will be added to the Results table.
8. Review the models listed in the Results table.
 - a. Each model includes its Test Accuracy.
9. You will be prompted to answer a multiple-choice question:
 - a. Select which model's Test Accuracy you believe is most indicative of its true accuracy on new data.
 - b. The answer choices correspond to the models displayed in the Results table.
10. Click Submit to lock in your selection.
 - a. Your chosen model will be highlighted in the Results table.
 - b. The Reveal Out-of-Sample Accuracy button will become enabled.
11. Click Reveal Out-of-Sample Accuracy.
 - a. An Out-of-Sample (OOS) Accuracy column will be added to the Results table.
 - b. A Change column will also appear, indicating how performance differs from the Test Accuracy.
12. You will receive a feedback prompt encouraging you to reflect on why some models' test accuracies differed from their out-of-sample performance.
13. Click COMPLETE ACTIVITY to finish the activity.

14. [Instructor] At this stage, lead discussion and have students share about their experience.

- a. Ask students what patterns they observed between Test Accuracy and Out-of-Sample Accuracy.**
- b. Use the discussion questions in the unit to help facilitate discussion.**

Activity takeaway:

Reserving a share of the data for testing is a necessary component of building a reliable model.

Lesson 8: Beyond confirmation bias: additional disruptions to rigor

Lesson summary:

This lesson broadens the discussion to include additional biases that can distort science. It emphasizes the importance of adopting systematic safeguards to minimize the cumulative impact of these biases on research integrity.

Goal:

Build a comprehensive understanding of how various cognitive biases and rigor issues connect to each other and negatively impact research.

Activity overview: (~15 min)

Participants will map out a “bias ecosystem” by connecting different biases and discussing real-life examples where these interactions have led to research pitfalls.

Heads up!

This activity introduces a lot of terms for a variety of biases. If students have questions about any particular bias, you can learn more by searching the [catalogue of biases](#).

Link to activity:

<https://monolith-test-1.vercel.app/activities/hms-cbi-08>

Step-by-Step activity instructions:

1. **[Instructor] Note that there is an option on the landing page to have students work individually and as a group.**
 - a. **If you want to run it for a class to only see their peers' results, click the "AS A GROUP" button to receive a session link that each member of the class will need to use to have their results coordinate on the results page.**
 - b. **If students select the "AS AN INDIVIDUAL" button, the results page at the end will show results from "global users" outside of your class.**
2. You should see boxes in the playable field with text describing different types of bias.
3. You can click on any box (bias) to drag it to a different position within the field.
4. You can click the center of any box to start drawing a line that can be connected to another box (bias).
5. Once you connect 2 biases, click "Click to edit" to type in a description of how these biases are connected.
6. You can connect more than one bias to another.
7. Once you've completed your "bias map", you can click "EXPORT AS PNG" or "EXPORT AS PDF" to save.
8. Click "SUBMIT" to advance to the next page and compare your bias map with other users.
9. **[Instructor] At this stage, lead a discussion and have students share about their experience and the connections they've made.**
 - a. **Segue to the discussion questions in the unit to connect students to the next portion of the lesson.**

Activity takeaway:

Different kinds of rigor issues and biases can replicate, cause, inform, or otherwise relate to one another.

Note for Instructors:

Each unit is estimated to comprise approximately 3 hours of instructional time (approx. 15-20 minutes per lesson), but variances in discussion length, student

needs, experiences with the interactive activities, or instructor customization may yield different unit and lesson durations.

Observations & final notes

Concepts likely to challenge students:

- Students may have emotional reactions in places where the instructional content differs from their prior experience. This is okay! You're helping them to learn how to do more rigorous work.
 - We advise letting disgruntled students express their points of disagreement, then gently encouraging them to consider why the materials might disagree with what they've been taught previously.
 - **Remember:** There's nothing wrong with asking a student to hold on to their grievance to let you conclude the unit.
 - **Think we got it wrong?** We want to improve! Email us at c4r@seas.upenn.edu.
- Differentiating between exploratory and confirmatory analysis and the pitfalls of "double dipping" in data.
- Recognizing subtle instances of unmasking in experiments, especially when biases operate at multiple levels.
- Designing research protocols that adequately control for the myriad ways confirmation bias can infiltrate decision-making.
- Navigating the technical details of machine learning modeling, particularly in parsing out different types of cross-validation.

Final Reminder for Instructors:

This guide is intended as a flexible framework for presenting the Confirmation Bias unit. Instructors are encouraged to adapt the content to best suit their teaching style and the needs of their students. Feel free to expand on any section, incorporate additional examples, or integrate further interactive elements. Remember, this is your presentation: use it as a starting point and customize it to best serve your teaching.