Now it's time to flex your critical evaluation skills. Read the following descriptions of an experiment and its analysis, identify the flaws in each, and describe what you would do to correct them.

1. The Sith Lords are concerned that their recruiting slogan, "Give In to Your Anger," isn't very effective. Darth Vader develops an alternative slogan, "Together We Can Rule the Galaxy." They compare the slogans on two groups of 50 captured droids each. In one group, Emperor Palpatine delivers the "Anger" slogan. In the other, Darth Vader presents the "Together" slogan. 20 droids convert to the Dark Side after hearing Palpatine's slogan, while only 5 droids convert after hearing Vader's. The Sith's data scientist concludes that "Anger" is a more effective slogan and should continue to be used.

Emperor Palpatine could just be a more charismatic and effective public speaker than Darth Vader and therefore, regardless of which slogan he delivers will be more persuasive. Either both speeches should be delivered by the same person – or the group should be split into 4 equally balanced groups – half will be addressed by emperor palpatine and half by darth vader, but then each of their respective groups will be split in two again and half will get the Anger slogan and the other half will get the Rule the Galaxy slogan. The data can still be analyzed as two groups ‘Anger’ vs ‘Rule’ – the bias presented by any extra charisma or personality trait effecting the choices will be balanced within

1. In the past, the Jedi have had difficulty with public relations. They send two envoys, Jar Jar Binks and Mace Windu, to four friendly and four unfriendly planets respectively, with the goal of promoting favorable feelings toward the Jedi. Upon their return, the envoys learn that Jar Jar was much more effective than Windu: Over 75% of the people surveyed said their attitudes had become more favorable after speaking with Jar Jar, while only 65% said their attitudes had become more favorable after speaking with Windu. This makes Windu angry, because he is sure that he had a better success rate than Jar Jar on every planet. The Jedi choose Jar Jar to be their representative in the future.

This would have been better balanced by splitting the 8 planets so that 2 friendly and 2 unfriendly planets were given to each representative. It is much easier to convince friendly faces than unfriendly faces and Jar Jar definitely had an unfair advantage.

1. A company with work sites in five different countries has sent you data on employee satisfaction rates for workers in Human Resources and workers in Information Technology. Most HR workers are concentrated in three of the countries, while IT workers are equally distributed across worksites. The company requests a report on satisfaction for each job type. You calculate average job satisfaction for HR and for IT and present the report.

Analysis must be done carefully to account for the imbalance between the two groups. Job satisfaction could depend heavily on country culture, so this must be highly scrutinized when assessing differences between groups.

1. When people install the Happy Days Fitness Tracker app, they are asked to "opt in" to a data collection scheme where their level of physical activity data is automatically sent to the company for product research purposes. During your interview with the company, they tell you that the app is very effective because after installing the app, the data show that people's activity levels rise steadily.

People confident in their workout routines are more likely to opt in to a tracker than those who know they are more likely to flake or otherwise not maintain a fitness routine. Since the tracker is optional it more than likely represents a highly biased population leaning towards people motivated to work out.

1. To prevent cheating, a teacher writes three versions of a test. She stacks the three versions together, first all copies of Version A, then all copies of Version B, then all copies of Version C. As students arrive for the exam, each student takes a test. When grading the test, the teacher finds that students who took Version B scored higher than students who took either Version A or Version C. She concludes from this that Version B is easier, and discards it.

There could be some bias present in the order that the students came in. For some reason the more well-studied students arrived in the middle of the group. OR there was more cheating present in the B group and since the tests were not randomly ordered in the pile it was possible for students with the same test to sit next to each other and “borrow” each others’ answers.

Discuss your answer to each of these questions with your mentor. For a deep and irreverent dive into the critical evaluation of experiments (among other sources of information), check out [Calling Bullshit in the Age of Big Data](http://callingbullshit.org/).