p-value: If you find that in only a few of the new groups the treatment cookies are liked more (e.g., 3.4% of the time), you think it's unusual and reject the idea that the treatment cookies are not liked more. But if in most new groups the treatment cookies are liked more, you think it's not unusual and don’t reject that idea

this is what not makes sense to me. If you find that only a few groups like the treatment cookies more...then why on earth would you reject the null and say people like the treatment more when that only happened 3% of the time?

**ChatGPT said:**

ChatGPT

Let's clear up the confusion. It seems there’s a mix-up in interpreting the p-value and the conclusion about rejecting the null hypothesis. Here’s a simplified explanation:

**Understanding p-value in Hypothesis Testing**

1. **Null Hypothesis (H₀)**: There is no effect; the treatment is not better. In this case, it means the average purchase amount for the treatment group is not greater than the control group.
2. **Alternative Hypothesis (Hₐ)**: The treatment is better. In this case, it means the average purchase amount for the treatment group is greater than the control group.

**How p-value Works**

* **Bootstrap Sampling**: You create many simulated samples from your data and calculate the difference in means for each one.
* **Compare Differences**: You check how often the differences in these simulated samples are as large or larger than the difference you observed in your real data.
* **p-value Calculation**: This percentage is the p-value. It tells you how common or rare it is to see a difference as extreme as the one you observed if the null hypothesis were true.

**What the p-value Means**

* **Small p-value (e.g., 0.034)**:
  + **Interpretation**: If only 3.4% of your bootstrap samples show a difference as large as or larger than what you observed, it means that such a difference is rare if the null hypothesis were true.
  + **Conclusion**: Because such a large difference is rare under the null hypothesis, it suggests that the observed difference is unlikely to be due to random chance. Therefore, you reject the null hypothesis and conclude that the treatment likely has a greater average purchase amount.
* **Large p-value (e.g., 0.900)**:
  + **Interpretation**: If 90% of your bootstrap samples show a difference as large as or larger than what you observed, it means that such a difference is common if the null hypothesis were true.
  + **Conclusion**: Since observing such a difference is not unusual, you don’t have enough evidence to reject the null hypothesis. It suggests that the observed difference could easily happen by random chance, so you don’t conclude that the treatment is significantly better.

**Why Small p-value Means Rejection**

* A small p-value indicates that the observed difference is rare and unlikely to have happened by chance under the null hypothesis. This rarity suggests that the null hypothesis (no effect) is unlikely to be true, so you reject it in favor of the alternative hypothesis (that the treatment is better).

**Visual Example**

1. **Imagine you’re rolling a die.** If you roll a 6, and only 3% of the rolls (in your simulated rolls) show a 6 or higher, this is rare. If you see this 6, it’s unlikely due to chance, so you might believe there’s a reason for this result and adjust your belief accordingly.
2. **If 90% of rolls show a 6,** this is not unusual, so seeing a 6 doesn’t change your belief much; it could just be due to chance.

**Summary**

* **Small p-value**: The observed effect is rare if the null hypothesis is true, so you reject the null hypothesis (supporting that the treatment is likely better).
* **Large p-value**: The observed effect is common if the null hypothesis is true, so you fail to reject the null hypothesis (supporting that there might not be a significant difference).

In your case, finding that only a small percentage of the bootstrap samples have a difference as large or larger than the observed difference means that such a difference is rare and not likely due to random chance. Hence, you reject the null hypothesis.