**Rians comments on the analyses R code**

**Comment 1**:

It seems to me that we are making statements like ‘ Both deepfake and genuine content influence behaviour’. This appears to be true by some measures (eg, mean intentions, and mean self reported intentions in RQ 1, or RQ7, etc) but not others (particularly the IAT test, RQ 1 or RQ7, etc.). This is also the case in RQ5 (and maybe some others).

In some of the IAT measures, the negative videos don’t seem to have an effect, but the positive videos do (see RQ1 or RQ7, etc). So if we were to use just the IAT measure (which would be a bad idea), then we could only make the statement that the positive videos influenced behaviour, but not the negative videos. Both the mean intentions measures seem to support the original claim that DF and genuine content influence behaviour.

Does this suggest we should drop the IAT measure, or weaken the claim??

**Comment 2:**

In H2C, we make a non-inferior claim about the mean intentions response between the genuine and deepfaked content, for the positive content only. This statement is weaker than for positive and negative content, which holds for the other two measures.

Is this a problem?

**Comment 3:**

H3b.

Here we make the claim that participants incorrectly think genuine content is deepfaked.

‘We expect participants to incorrectly detect’

Should we change this to ‘We expect **some** participants to incorrectly detect’.

Maybe this is just nit-picking? Just thinking this because the threshold is 5% of participants, and I feel ‘we expect participants to ‘ suggests this will be a larger effect. Could just be me though.

**Comment 4:**

This is a general comment about reporting stats. We want to make the claim that

‘We expect participants to be poor at making accurate decisions about whether content is genuine or not (i.e., balanced accuracy not greatly above chance, ≲ .60).’

When we make such a statement in the paper (assuming it’s true), do we just report the actual statistic to quantify it?

**Comment 5:**

The y-axis labels on for the ‘interpret posteriors’ graph need to be fixed.

**Comment 6:**

How are ‘awareness’ and ‘detection’ actually defined? This is probably in the actual code, but might be good to specify in the text to help interpretation.

**Comment 7:**

The titles in RQ5 and RQ6 need fixed, and the title for RQ7 has a typo.

**Comment 8:**

In RQ7, the responses to the mean intentions are negative, even for the positive videos. Is there an error here?

**Comment 9:**

Just a friendly reminder that RQ5 and RQ7 still require edits, around the ###### parts.