Design Check

Anonymised for Peer Review

Table of contents

In this supplementary material, we present two follow up analyses that test the efficacy of our experimental and analytic designs. First, we present an analysis investigating whether people were equally as persuaded by a consensus arguing in favor of the claim compared to against it. Secondly, we investigate the possibility of demand effects induced by our within subjects design, by seeing whether our results differed between the first and last trials fo the experiment.

Are people equally persuaded by claims arguing for (pro) and against (con)?

Previous studies using our paradigm found that people were more or less equally convinced when posts argued for versus against the claim. For this reason, to preserve power in the individual level analyses, we transformed the direction of belief changes on trials arguing against the claim rather than adding argument direction as a predictor. But was this a safe assumption? It's possible that the claim type X consensus interaction might only hold for tweets arguing a certain direction.
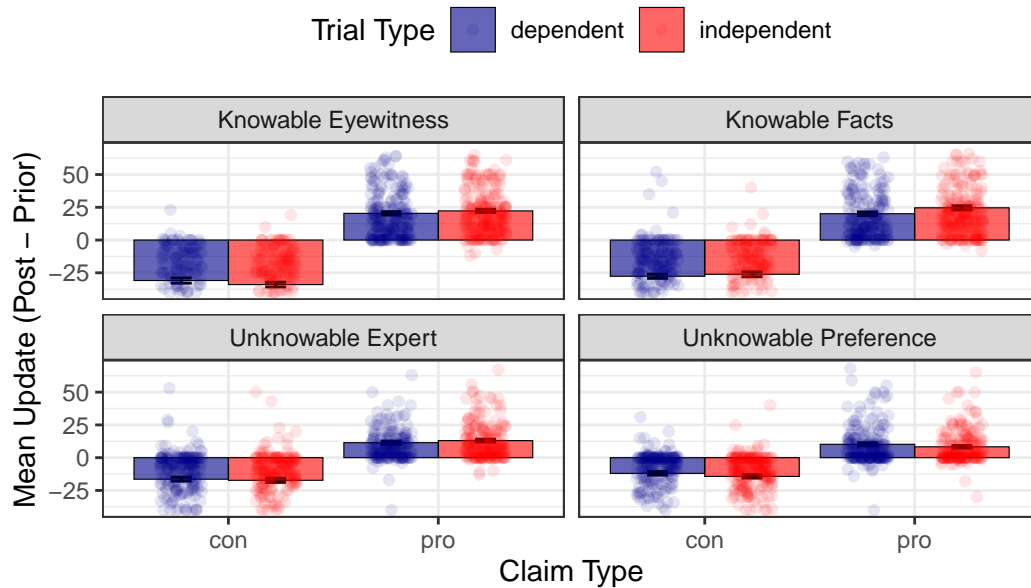
Although people seemed to be generally more convinced by posts arguing against the claim, the trend whereby a concensus was more convincing when the claim was more knowable appears to hold. Additionally, the difference between independent and dependent conditions appears to remain similar for pro and con.

Do the result remain when we only look at the first trial?

Due to the nature of within subjects design, it is possible that the results we observed were due to demand effects. One way to explore this possibility is to see whether our results hold on the first trial, before thy have seen any trials from other conditions. It will also be informative to look at the last trial, to see whether there appears to be any systematic changes from beginning to end.

The main limitation of this approach is that it substantially limits statistical power. Consequently, we did not perform any statistical analyses, as any null results would likely be due to a lack of power. However, the same qualitative trends are visible in both the first and last trials as observed in the entire sample. The plot below compares the first and last trials of session 1 and 2, with the numbers on the plot indicating the number of observations in each condition (showing there are sufficient n in each group to make it unlikely that any extreme responses would have affected the results too much).
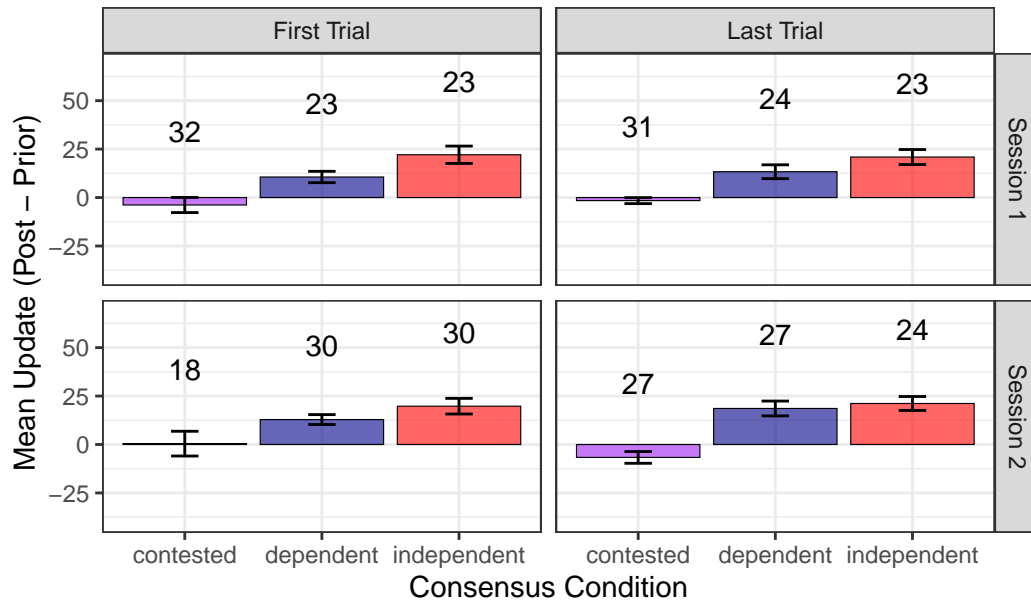
Though it is less likely that these results would be effected by demand effects, it is also worth examining whether the claim type finding holds.
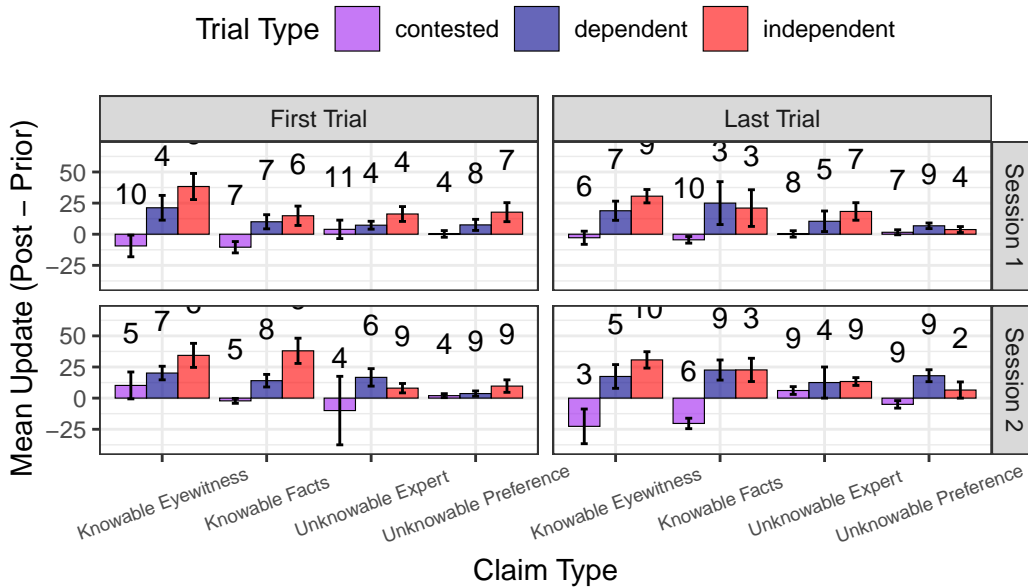
The plot above shows more variability in the differences between consensus conditions, but also that the number of responses in each cell is much smaller and often quite imbalanced, making it difficult to interpret trends with any certainty. However it does appear that in general, there is a larger difference between an independence consensus and no consensus in the more "knowable" conditions, supporting the conclusion from our full analyses.