Point-by-point Responses to the Reviewers' Comments (Rd. 3)

Again, we would like to express our sincere gratitude for the time and effort to review our manuscript.

Reviewer 2

1) abstract

Rewrite "*Given that concrete expressions contain richer information than abstract counterparts,*" as "*Given that concrete positive expressions contain richer information than abstract positive counterparts,*"

I think the reader needs to be alerted that only positive statements will be used. If the test is passed a simple theory could explain it: The more certain the positive information about someone, the more favourable the attitude towards them.I think the study would be more interesting if the statements were neutral, neutral statements would afford a more severe test of uncertainty reduction theory. But that is up to the authors.

Thank you for your advice. We rewrote the relevant sentence as follows. We agree with your idea that using neutral sentences as stimuli is informative for testing URT. Nevertheless, we think that this study can also have a unique contribution to URT literature as it measures and controls the inherent favorability of the sentences while most studies have not conducted such a procedure. Considering that there is no completely "neutral" information, we believe it is certainly an essential procedure when we examine the effect of uncertainty reduction.

"*Given that concrete expressions contain richer information than abstract counterparts, we hypothesize that (1) a person described in more concrete (positive) terms will be perceived as more attractive, and (2) the level of uncertainty will mediate this effect.*"
(lines 32 - 35)

2)

The statements defining uncertainty reduction theory I found in the manuscript included claims such as

"*The Uncertainty Reduction Theory posits that as an individual's uncertainty diminishes, they will be evaluated as more attractive*"

"*URT assumes that when uncertainty is high, a person appears less attractive. Given this prediction and the fact that information-seeking contributes to reducing uncertainty, it is expected that more information will reduce uncertainty and increase attraction.*"

Given the assumptions of the study, the study is in a position to provide evidence against URT. The authors do qualify in places that the theory only applies to the early stage of a relationship; this should be explicitly added to the statement of the theory. The study still tests this theory.

If the authors wish to claim URT could not be falsified by the study, they need to define what URT claims more restrictively. But that would be odd. Why not posit the more general theory - the

version they have actually presented (or the revised version just given) - and then have the study potentially count against it (and indicate this in the Design Table).

Thank you for the comment. The context adopted in this study ("online unilateral communication") was not envisioned at the time of the URT proposal. Thus, we think that even a null result in this study cannot negate the effect of URT in the primary context in which it is supposed to work. However, as you indicated, we realized that it would be unkind to readers. Thus, we decided to clarify that a null result could limit the applicability of URT in the design table.

"*Although results against our hypotheses will not necessarily mean the URT is entirely wrong, they may indicate that the URT cannot apply to the situation of online unilateral communication.*" (Design table)

3)

In the pilot study, equivalent favourability is defined by an equivalence test. Specifically, that means the difference in favourability is significantly less than a threshold value (0.33 Likert units). The complement of this - that is, the implication of what not equivalent means - is that the difference is significantly more than the threshold value. (I previously phrased this in terms of "inference by intervals": The two t-tests the authors perform for TOST is mathematically the same as determining if the 90% CI lies within the equivalence interval, whether one interval lies within another interval. The claim that the abstractness ratings were not equivalent could thus be made if the 90% CI lies outside the equivalence interval. Thus, the authors are indeed using inference by intervals.) Put another way: if the authors claim 0.33 Likert units is only just worth taking note of as a difference, then to show a meaningful difference in abstractness, they need to show abstractness was more than 0.33.

I realise that in the end they will have to select whatever items they have that maximise differences in abstractness, even if some strict statistical rule is not satisfied for all items (as they detail). And this seems OK because it is a priori plausible they can create large differences in abstractness. So the points I just made won't change anything in practice. Still, it would be better to start with an inferentially mutually consistent set of rules, and only soften from there as necessary.

The average difference in abstractness relates to predictions of effect size made later; getting rather more than a minimal scientific relevant amount in abstractness turns out to be important. That increases the case for a strict statistical definition at this stage.

Thank you for the comment. Although our previous plan was that we only use pairs of sentences whose abstractness differed greatly than Cohen's d = 0.3 (corresponding to 0.33 Likert unit), we did not plan to confirm this difference was statistically significant, as you pointed out. Thus, in the new plan, we decided to perform a paired t-test that confirms the difference in the mean value of abstractness is greater than 0.33.

**"*First, a paired one-tailed t-test will be conducted on the means of the rated abstractness of the abstract and concrete sentences for each pair. In this test, we will test whether the rated abstractness of the abstract sentences is higher than that of the concrete sentences by more than 0.33. Thus the null hypothesis will be "μ = 0.33," where μ denotes the mean of the difference in*

*abstractness. This step is to ensure that our manipulation works efficiently enough.***"**
(lines 268 - 272)

"*We set the threshold value throughout those analyses as 0.33 because we set our minimally interested effect size in the preliminary survey as Cohen's d = 0.3, and we assumed the SD of the rating as 1.1 based on our pilot survey.***"**
(lines 282 - 284)

4)

"*Although all pairs will be included in the hypothesis-testing analyses regardless of whether the differences in perceived abstractness of profile turn out significant, we will conduct an additional sensitivity analysis that excludes pairs of profiles whose perceived abstractness do not differ significantly by condition.*"

      This opens the way to inferential flexibility. Just be clear that the main conclusion won't depend on this analysis; or else, how it will depend.

      Thank you for the comment. We added a sentence that makes clear that the main conclusion won't change even if additional sensitivity analyses show different results, as follows.

"*The main conclusion will only depend on the hypothesis-testing analysis, even if the additional sensitivity analysis shows different results.***"**
(lines 371 - 372)

5)

"*The assumed effect size of uncertainty on attraction (β = .51) was determined based on a previous study (Baruh & Cemalcılar, 2018).*"

      What was the effect tested in this study? Make clear it is relevant. Allow anyone to find the effect easily.

      Further, the size of an effect found in a previous study helps fix what might be predicted; but not the smallest size one does not wish to miss out on; and it is the latter type of error that power is meant to control (in the long run). If a previous study found beta=0.51 and that was interesting, that does not mean one would be happy to miss out on beta = 0.4.

      The authors have claimed the smallest size of scientific relevance for favourability and abstractness on a 1-7 Likert scales is 0.33 Likert units. (A scientific justification would be good, though I realize this may be difficult.) It is also highly likely that a given difference in abstractness will produce a smaller difference in attractiveness. The question then becomes, what is the smallest ratio of attractiveness to abstractness difference that would be minimally scientifically relevant? It is hard to say, but a 0.5 reduction still seems big (so if the difference in abstractness was 1 Likert unit, and the difference in attractiveness was 0.5 Likert units). Maybe a reduction to 20%? This is just intuition and it would be good to have a reason for it. Maybe the authors can provide one (e.g. from similar situations in social psychology).

In any case, this way of thinking shows there is a relationship between claims of smallest effect of interest for different analyses. Likewise, it also bears on the size of the indirect effect in the mediation analysis: That can be expressed in the same way (what Likert difference in attractiveness should result from the indirect pathway for a given difference in abstractness?). And the answer for the smallest effect of interest for the indirect pathway should be the same as for the smallest effect of interest for the direct effect of abstractness on attractiveness. (Note the predicted effects would be different. Getting predicted effects is an altogether easier matter. But predicted effects is not what is relevant for power or equivalence testing.)

Thank you for your thoughtful comment. We added an explanation on Baruh & Cemalcilar (2018) that we referred to in the power analysis.

"*The assumed effect size of uncertainty on attraction (β = .51) was determined based on a previous study that reported the relationship between those variables using almost the same question items as ours (Baruh & Cemalcılar, 2018, Study 1).*"
(lines 306 - 308)

We realize that performing power analyses based on the effect size from previous studies is not very informative when we aim to test whether there is an interesting effect or not (Lakens, 2022). However, in this study, there is so much difficulty in determining sample size based on "the smallest ratio of attractiveness to abstractness difference that would be minimally scientifically relevant" for two reasons. First, we will not use abstractness rating as it is in the main analysis, rather set the abstract and concrete conditions and use a dummy variable. Thus, we cannot calculate the necessary sample size for a specific ratio of abstractness to attractiveness, unless we know how much difference in the abstractness ratings between the abstract and concrete conditions. Second, we will not use the sentence alone as stimuli in the main experiment; rather, we will use a profile that contains two positive sentences whose abstractness differs by condition. In this setting, it is difficult to disentangle the effect of the abstractness of two sentences on the attraction of the profile. Due to those realistic constraints, we cannot calculate the sample size in the way you recommended, or even if we can, the calculation is not so informative because of too many uncertain assumptions.

We realize that this is one of the important limitations of our study and will discuss whether the effect size observed in this study is interesting enough or not in the stage 2 manuscript.

6)

The possible theory from pragmatics I was referring to in my previous review is more general than the authors take it to be. The claim is that there is a pragmatically appropriate amount of information to divulge in *any* communication - and that would include early stages of a relationship. The theory in effect states there is an optimum amount of information, and so contrasts with uncertainty reduction theory (the more information the better). Pragmatically, more information will make communication go smoothly up to a point, and then make it worse. But I think this can be dealt with in the discussion of the Stage 2.

Thank you for the valuable comment. It is possible that even in the early stages of a relationship, too much information diminishes attraction, although we are not able to test the theory in this study. We will discuss this point in the Stage 2 manuscript.

Other major corrections

1) Power analysis

We corrected a mistake in the R code for the power analysis of the main experiment and re-calculated the sample size. In the old version, we coded the experimental treatment as a dummy variable 0: abstract and 1: concrete. However, in this setting, the effect of the treatment (β) had been improperly calculated because SD of the treatment was 0.5, while SD of the mediating / outcome variable was 1. To set SD to 1, the coding should be -1: abstract and 1: concrete. We re-performed the analysis and found that 681 participants were enough to achieve 80% power assuming the effect of treatment to uncertainty as $\beta = 0.17$ (small) and the effect of uncertainty to attraction as $\beta = 0.51$.

We replaced the old codes with new ones in the OSF:
- **R: https://osf.io/5evgq?view_only=dc1bb4d7647046ccae4d64ba44448921**
- **html: https://osf.io/ku8dc?view_only=dc1bb4d7647046ccae4d64ba44448921)**

2) Stimuli

We replaced some stimuli with new ones (Appendix: https://osf.io/mwfsb).