Reviewer Comment	Author reply
Dear Christian Thurn and colleagues, Thank you for all careful revisions and detailed responses to previous feedback. Two reviewers were able to return to carry out another feedback round and they both were very satisfied with this improved version. They only had a few minor suggestions. I'll let you consider that feedback in a final revision, but I won't invite the reviewers anymore for a third round. I agree this version is good and near-ready for IPA	Dear Veli-Matti, Thanks for your additional constructive comments. We considered the reviewers' feedback and you find our changes and replies to the reviewers below.
I have only one comment of my own. This concerns H1.	H1 indeed posed problems to us.
- Because you're testing a confirmatory hypothesis, it would be good to explicitly justify why you expect a certain outcome in the Study 3 section before H1. Currently you write, "we are interested in how the number of characters in a play relates to the complexity... our goal is to understand the relation of the number of characters to the complexity of networks in theatre plays," which is an exploratory description. But your H1 is a confirmatory test ("We test the hypothesis that the number of characters positively predicts complexity (H1)") so it would be important to briefly recap in Study 3 section why do you predict a positive result.	We completely changed this section, as we found it really difficult to derive clear hypotheses given a lot of untested assumptions in the current state of research. Instead, we decided to investigate the relation between the number of characters and complexity exploratorily.
- It feels to me that your assumptions ("we assume that plays are more likely to be well-received and popular if they make it possible for recipients to follow the narrative") predict the null insted of a positive correlation. I might be mistaken, but it would be worth clarifying what you expect and why.	We are now more clear in our writing as this project does not provide answers to the question whether the complexity of character networks is an adequate proxy for the demand that a play puts on recipients' cognitive systems. This is a possibility for future research. Instead, we clarify that we test the often implied link between the

	number of characters of a play and complexity in sort of a convergent validity approach.
- Because this is a hypothesis test, and PCI RR is very strict about justifying effect sizes in hypothesis tests, it would be good to have some kind of justification for r < .30s as small effects. I know it's difficult to think about justification in this context. I also find it challenging to help with this—especially as I lack a comprehensive understanding of Kolmogorov complexity in the present context—but here are some ideas:	We agree that justifying ES is difficult and arbitrary in this context, wherefore we dropped the hypothesis and decided to interpret the correlation between the number of nodes and complexity with benchmarks used for test-retest reliability.
● One option would be to seek existing character network data in fiction and see what r=.3 looks like. E.g., in fiction (of any media), what is r=.3 in terms of complexity? Can you effectively separate actual works of fiction by complexity? This could help both you and readers grasp the raw effect size and justify it.	
● Another option that comes to mind would be to simulate data with r=.3 and see how the effect size appears in these simulated instances. Being able to pinpoint reasonable raw differences even in simulated form would be better than nothing.	
● A third option could be to select one actual play by Shakespeare, provide a description of its character network, and demonstrate a hypothetical raw change of .3 in practice.	
● A fourth (meta)option would be to take a Bayesian approach and rely on (non-informative) priors. Some justification would still be nice to have but the basis of the rationale would be less problematic for inference.	
● (The same concerns RQ4 but since it's exploratory it doesn't matter. Btw, also noticed you don't mention alpha	We do not mention alpha but always will use 95% CIs

anywhere in the paper -- is it 5% throughout?)	
Some of the above ideas may be unfeasible, so please read them primarily as food for thought; I hope they guide you to the best solution from your own topic-expert position. As a standard note, I refer you to PCI RR evidence thersholds and these two papers by Zoltan Dienes on specifying theoretically relevant effect sizes for statistical hypothesis testing: https://doi.org/10.1525/collabra.28202 https://doi.org/10.1037/cns0000258 I sadly don't have any good examples from theatre or literature, but if you wish to have practical examples of effect size justification from other fields, contact me and I will seek some from PCI RR archive. If you are unsure about anything else or wish to discuss, you can email (as usual) before submitting. Best wishes, Veli-Matti Karhulahti	
Reviewer 1	
This is a far more coherent proposed research plan with a clearer rationale as to why do the study. The authors responses to the feedback and new plan, have from my perspective, addressed the key issues with the initial draft, and addressed methodological concerns. As before the analytical approach appears robust. I have made a few comments below, not all in relation to this stage of the process but as things to consider when interpreting findings and justifying the approach in their final write up.	Dear James, Thank you for reviewing our manuscript again. We are happy about your positive evaluation and the helpful comments.
• There has been a clarification of the "slices" and what these are, that makes the use of the analytical approach more comprehensible and justifiable.	Thanks, your previous comments helped to clarify this aspect.

• The response to my questioning of the justification of the scene as opposed to entrances and exits was very clear, and I think the response mentioning that using scenes is more feasible/ reproducible is actually a strong argument and could be used to justify this study further. As using entrances and exits will not necessarily produces the same results as scenes. Within that argument the strength of the scene is that it is not subject to individual interpretation of entrance and exits. However, I think their phrasing of "Please note that slicing a play based on exits and/or entrances automatically slices a play by scenes too" will ideally need unpacking in their final research as there would be potentially network implications.	We are glad that you found our response clear. That entrances and exits may not necessarily produce the same result as scenes is one reason for us to conduct study 4, in which we compare different operationalizations of the network. We further unpacked our reasoning in line 249f.
• In relation to the use of speaking characters, there will at some point need to be a justification and consideration in the interpretation of the results in terms of what constitutes a speaking character. Some characters such as Livinia in Titus Andronicus will spend part of the play interacting with other characters (communicating) but not speaking or in some plays there could be scenes with significant onstage action and minimal dialogue.	Thank you for this input. We will address this point in the limitations section of the Stage 2 report.
• They could perhaps make more of the applicability to literary analysis and the application of the validation approach, however, this can be done in the final stages/ write up.	We agree that the discussion would greatly profit from formulating our insights in a way that could prove helpful for other researchers interested in (interdisciplinary) analyses of cultural products. We will consider this in Stage 2.
Reviewer 2	
Dear Authors, I was very pleased to read the revised Stage 1 protocol and your responses to the initial feedback. I'd like to acknowledge the substantial improvements made to the paper.	Dear Matúš, We thank you for reviewing our manuscript again and for your positive and helpful comments.

The introduction, including the study's rationale, are now well-described. The methodology for the current four studies is sound and highly rigorous. Before suggesting in-principle acceptance, I'd like to request a few more edits and clarifications.	
In several instances, the authors assume that the number of characters in theatre plays is mainly determined by recipients' cognitive capacity. However, this upper limit for theatre plays might be influenced by other factors (e.g., historical and technical/pragmatic considerations). It might be worthwhile to briefly mention these alternative explanations.	We assume that the demand that is put onto a system is positively and monotonously related to the complexity. It is the goal of Study 3 to analyse how closely the number of nodes tracks complexity. Importantly, we do not make any assumptions about which factors determine the number of characters in plays. Instead, we merely rely on the assumption that plays are more likely to be popular if character networks are simple enough for the recipients to sufficiently follow the narrative. Networks with more nodes have the potential to be more complex than networks with fewer nodes. However, it is in principle possible for larger networks to result in lower complexity than smaller networks. When it comes to the number of characters in a play, we agree that there are many factors that might explain differences, including contextual information. Since such questions are not the focus of the present project, we will not expand on these points in the theoretical background for the Stage 1 report. However, we will include a brief mention in the discussion of the Stage 2 report in case that this would help contextualize some of the findings. In our planned analyses, we include contextual information in the context of Study 3. Based on your comment, we rephrased the relevant sentence to make it more clear that contextual information will be included (see lines 226 to 228).
A more accessible description of some of the concepts and methodology would benefit a broader readership. For instance, examples of how to apply Kolmogorov complexity (as a focal construct of the present study) in the context of theatre plays would be helpful. Specific	On lines 71 to 77, we made our way of applying Kolmogorov complexity to character networks more transparent and suggest a tutorial for further reading. In line 80-85 we added an example applied to theater plays. In the discussion we will

examples applied to the context of theatre plays, along with brief insights into their potential implications, would also be very beneficial for Study 3. (Btw, I really liked your second operationalization of complexity described in Study 3.)	evaluate which aspects of complexity won't be captured by our approach. We tried to keep these additions brief, but readers interested in the detail could look at the R scripts that we will share publicly.
In Table 2, the authors state "If many plays lie above 3*IQR, they represent outliers of high complexity. If no or very few plays lie outside 3*IQR then Shakespeare's plays are not particular with regard to complexity." Please specify the exact number of plays that you consider as "very few."	Good spot. We now replaced the interpretation in Table 2 with a formulation that is more consistent with the text.
After reading your responses, I would appreciate a summary of the perceived/expected limitations and challenges mentioned in the paper. I'm not insisting on this, as it's a bit unusual request at Stage 1, but reflecting on this at this stage could greatly improve transparency and increase confidence in the methodological choices and results. All the best, Matúš Adamkovič	We agree that being clear about limitations at this stage is very helpful. We are keeping a list of limitations that we are constantly updating. Yet, not all limitations will be eventually relevant (e.g., if they concern cases that will not appear in the data). The most important considerations in this regard are reported in the main text of the Stage 1 manuscript. In response to other reviewer comments we also added a limitation in the footnote (ll. 86-88) to avoid wrong expectations. We decided to elaborate meaningful limitations in the discussion section of Stage 2 and are open to further reviewer input at Stage 2.