The authors have satisfactorily responded to most of my previous comments.

In the comment marked 2.7, the suggestion was to compare PAS ratings from pilot data to the RMS contrast of images. The authors instead compared RMS contrast and other properties across experimental conditions. The suggestions was primarily to test if these qualities were being used as cues for PAS ratings, in which case, the authors could correct the issue (ensure all images have the same RMS contrast) prior to data collection. It is a shame data has already been collected and so the authors cannot correct the methods, where methodological improvement is normally a major benefit of the registered report. I would suggest the authors do compare RMS contrast and PAS ratings in the experimental data, to answer the question "are participants making PAS ratings according to instructions (visibility of contents) or based on some cue such as RMS contrast".

It is a shame the authors cannot implement the suggested changes to the methods. The pilot data suggests the masking paradigm is not strong enough to guarantee the participants had no conscious recognition of the images. I suggest the authors discuss this possibility in their eventual discussion, in relation to the frequency of actually giving a PAS of 1 in the condition that is supposed to be 'unconscious'.

We would like to thank the reviewer for the feedback, and for further clarifying their 2.7 comment. We have now run Bayesian ANOVAs with the low-level property as DVs (RMS contrast, luminance, edge density, and 1/f slope), PAS as categorical predictor, and participants as random effect. This was done on only a subset of 18 template images that were common between Pilots 2 and 3 and the experiments proposed. The results are included in the table below. For Pilot 2, there was moderate to strong evidence for the null for all the parameters, while in Pilot 3 evidence still favoured the null but overall it was inconclusive. Overall, these results suggest no relationship between PAS ratings and low-level image properties. However, we will run the same tests in the main dataset, and report and discuss the results.

Property	**Pilot 2 Free Naming BF (+ error)**	**Pilot 3 (MCQ) BF (+ error)**
RMS contrast	BFnull = 9.46 ± 0.59%	BFnull = 1.42 ± 0.51%
Luminance	BFnull = 11.1 ± 0.6%	BFnull = 1.63 ± 0.52%
Edge density	BFnull = 26.8 ± 0.62%	BFnull = 1.89 ± 0.54%
1/f slope	BFnull = 8.57 ± 0.59%	BFnull = 24.7 ± 0.64%

We have now included the addition below on page 15, along with the table and a short summary of the findings in the Supplementary Materials on page 28.

"To assess if variations in any low-level image properties (RMS contrast, edge density, spatial frequency, and luminance) influenced PAS answers, we will also test whether these properties are predicted by the PAS levels, in separate Bayesian ANOVAs."

Nevertheless, even if we do find that RMS contrast or other low-level properties predict PAS ratings in our main dataset, we do not believe this would be a particular concern. As we described in our previous response, there is no difference between conditions in the low-level properties of the images, so no condition is going to be more accurately rated in relation to the image content than the other. Moreover, it would be valuable for the field to know if such a widely-used subjective measure tracks clarity of low-level properties rather than clarity of content, although we also appreciate that the two types of clarity are not independent.

We will indeed also look at the distribution of PAS answers in the short SOA condition. However, as we mentioned in comment 2.3, we do not expect that all trials will be answered with PAS1, nor do we think this to be a limitation of our design or masking procedure. Indeed, if we take datasets from different papers and look at the distribution of PAS answers in trials where nothing was presented, participants answer with PAS1 in only 82% of trials on average (SD = 8.29%, see figure below, copied from another manuscript in preparation). The pattern of answers we observe in Figure 7 in the Short SOA condition (where something is presented but masked) is not markedly different (66% in Pilot 2 and 60% in Pilot 3).