We thank the reviewers for the thoughtful comments, addressed below.

Statistical analysis (R2, AC): We have used t-tests because they have been shown to be accurate even for skewed distributions when the data is large (>500) [1]. But we agree that nonparametric tests are more appropriate. We've repeated the analysis using bootstrapped confidence intervals on winsorized data (to cap outliers) and also compared medians with Wilcoxon tests. The results are qualitatively the same: VoiceOver users produce more status updates (bootstrapped win. mean = 3.7 compared to 2.2 for iOS users) and more likes (M=108 vs. 92.8), and receive more status comments (9.7 vs. 7.7) and fewer photo likes (M=14.4 vs. 16.6) (all p < 0.001). We will revise Figs 1 and 2, and use bootstrapped CIs to report effect sizes for all measurements.

Ethical concerns: The analysis was done in aggregate on anonymized data on Facebook's servers, and falls within its TOS. Text analysis was performed automatically and the cited messages were all shared publicly. We will clarify this in the revision. We agree that a deeper analysis of photo content would be interesting but for privacy reasons we did not view images. Instead, we analyzed captions as a proxy for content. Overall, we try to balance the tradeoff between privacy and a better understanding of an underserved community.

We will discuss design implications for automatically identifying visually impaired users (R3). Auto-detection of users’ vision impairment could enable online services to better adapt to their needs. For example, users could be redirected to a specially designed version for visually impaired users, in which visual elements and dynamic content are de-emphasized (e.g. www.amazon.com/access). This approach has been very well-received by the visually impaired community so far.

We will add a table for all activity variables with means, medians, and standard deviations (R1).

Sampling threshold of 3x per month (R3): A sensitivity analysis with different cutoffs for VoiceOver usage (5, 7, 10) shows the results do not differ significantly.

Keyword representativeness in Tables 2 and 3 (R2, R3): Among the 12K VoiceOver users who posted any status updates, 7% of them used at least one of the 10 keywords by the abs. change metric and 5.5% from the prob. change metric (Table 2). Meanwhile, only 0.9% and 0.4% of the iOS sample users did. We used the same algorithm to compute the top 10 distinctive keywords used by iOS sample users. They are (1) with, have, this, that, just, love, time, like, your, what (from status updates) and  (2) this, with, love, from, happy, have, just, time, birthday, good (from captions). We will include these in the revision.

R2 noted that we describe two time periods for data collection, which led to confusion. We didn’t measure any longitudinal changes and will make this clear in the final draft. We simply picked samples in June and analyzed activity in August to ensure that participants in the sample were not brand new users, where novelty might affect the results. We will also revise our description of formulas (1) and (2), in which first and second periods are terminologies used in the trend detection algorithm. We're not measuring changes over time, but rather normalizing counts of terms from the VoiceOver sample by counts from the iOS sample. Same for the network size: we made our observation once and control on Facebook tenure.

We found it difficult to separate app-generated photos from personal photos (R1); people often add personal messages to the default caption in app-generated photos. Overall, our analysis is not intended to provide a complete understanding of how visually impaired users take photos, but share some new insights on the photos they share online. Among all the photos posted, 6.4% of them have ENGLISH captions (5.4% for VoiceOver sample and 6.7% for iOS sample). We will discuss these numbers in the revision.

R2 asked about sample distribution: Users were randomly sampled and sample sizes were large enough to make statistically significant observations.

All reviewers raised concerns about comparing TapTapSee users to iOS and VoiceOver users. Our goal was to obtain a subset of individuals with severe vision impairment. However, this sample is so small that it is potentially biased (R1). We will remove the analysis of TapTapSee users from the network section.

In the revision, we will discuss limitations in greater detail (AC, R1, R3), such as limiting the study to VoiceOver (iPhone) users and their behaviors on FB, and lack of information about the level or history of users' vision impairments.

We will expand the related work section according to the comments (R1, R3). We will also have an external proofreader do a thorough proofreading pass for spelling and grammar.

[1] Lumley T, Diehr P, Emerson S, Chen L., The importance of the normality assumption in large public health data sets. Annu Rev Public Health. 2002;23:151-69.