Tennessee v. Meta
Case No. 23-1364-IV

DX0340

PSC Self Review H2 2021

Kyle Andrews

Highlights

- The success of v1 BEEF led to data science recommending that BEEF be used for top-level goals for well-being in H1 2022. BEEF provided the means to operationalize user perceptions as goals at the org level, a departure from the traditional reliance on rater prevalence, and a first for IG.
- BEEF data was also used to create the Creator Well-Being Operational Scorecard, which Sam[redacted] led at the request of Adam; validate and improve Relevance Integrity's p(objectionable) and p(report) models, in addition to measuring their progress vs a holdout; serve as a goaling metric for 2022 for the Equity team (audience limitation question); provide the Youth team issue cuts by early and late teens for the first time, which they used the data to make the case to senior management for supervised experiences for younger users, as well as additional safeguards around messaging; underlie Arturo's recommendations to senior leadership, including a briefing with Adam where he reviewed the BEEF deck and a note summarizing his recommendations for the well-being function; and improve the age consistency modeling work of Central Integrity, among other uses.
- Continued as a survey reviewer, completing 39 SRT review assignments this half. Also served as a reviewer for the IG Safety and Community Health RFP, completing 17 reviews. Conducted 13 interviews (2x team average). Continued to formally and informally mentor other researchers (see below).

Results/Impact Details

- **Bad Experiences and Encounters Framework (BEEF) Survey, v1:** Analyzed, reported, and socialized the v1 BEEF Survey, a new area of research with broad impact across multiple areas of the company. BEEF v1 provided an overarching understanding of the bad experiences people face on IG, beyond policy-violating issues; a test of whether we've made progress vs our half holdouts in Well-being and Relevance Integrity; a ranked priority of issues for IG to prioritize in the future; a deeper understanding of the contextual factors surrounding issues, beyond prevalence; and insight into the most impacted subgroups (teens, creators, etc) per issue. Executing BEEF successfully required working broadly across Instagram, in addition to working with FB's SHIELD team and becoming the IG TRIPS point of contact for Central Integrity. The study influenced how teams goal and focus their work:
    - The success of v1 BEEF led to data science recommending that BEEF be used for top-level goals for well-being in H1 2022. BEEF provided the means to operationalize user perceptions as goals at the org level, a departure from the traditional reliance on rater prevalence.
    - BEEF data on creators was incorporated into the Creator Well-Being Operational Scorecard, which Sam[redacted] led at the request of Adam, and helped prioritize which issues are most important for creators, a company priority.
    - The data also was used by Relevance Integrity to validate and improve their p(objectionable) and p(report) models, in addition to measuring their progress vs a holdout.
    - The Equity team used the BEEF audience limitation question to investigate whether perceptions of audience limitation correlate with metrics of reach, which it does. As a result, the Equity team is using that question as part of their goaling for 2022.

- The Youth team used the data to make the case to senior management for supervised experiences for younger users, as well as additional safeguards around messaging. BEEF provided data on harms broken out by early teens vs late teens, which the team hadn't had before; this was used to inform policies for early teens.
- Arturo used the BEEF results as the basis for his recommendations to senior leadership, including a briefing with Adam where he reviewed the BEEF deck, a note summarizing his recommendations for the well-being function, and direct recommendations (via email) to Mark, Sheryl, Adam, and Chris Cox.
- Worked with Carlos [redacted] Central Integrity UXR, to share analyses to the Age Verification Working Group. The stated age vs imputed age of BEEF respondents was valuable data for them as they work to improve age consistency.
- The BEEF data helped the IG Support team understand better which issues users feel supported vs not supported; it also helped evolve the way in which they ask about support (reflected in the v2 BEEF survey).
- The #1 theme for the Experiences Platform team in H1 2022 is to proactively reduce bad experiences, which BEEF question development and survey results will help measure.
- The H1 roadmap for Discovery Safety proposes using BEEF/pTRIPS to track and reduce problem areas.

- **Graviton notebook development for survey weighting.** One of the obstacles faced in the BEEF v1 analysis was weighting the survey data to population. Unlike on FB, where inverse propensity score weights (IPSW) are automatically emailed to researchers after they run a survey, IG researchers have to do it manually or not at all (mostly the latter, although some people calculate rake weights in R. Rake weighting isn't as good as IPSW, though). I wanted to use BEEF as an opportunity to uplevel IG's quant work by developing a Python notebook that could implement Graviton to calculate IPSW, and Jikuo was kind enough to help. I worked with Jikuo to identify the appropriate covariates, interfaced between him and the survey infra/Graviton team, and tested the different iterations he produced. Once we had a working notebook, I created an annotated template notebook for others to use. I then hosted a workshop for IG quant researchers, and taught them how to use it. Since then I've helped other researchers weight data with the notebook, and answer their questions.

- **BEEF v2:** Developed the v2 BEEF survey, which will launch in H1. I spent a considerable amount of time this half (with Kathryn's help) building consensus on the scope and goals for BEEF v2 and v3. This involved working across teams and disciplines to understand the various measurement needs within Well-being, and disentangle them so I can propose the best methodology for each need. For the v2 survey itself, I worked across Well-being and Relevance Integrity to understand the different experiments that have been launched to each holdout this half so I can ask the right issue questions to capture team impact appropriately. For the follow-up BEEF questions, I collaborated with researchers in Well-being and Relevance Integrity to understand the knowledge gaps they have around particular issues, as well as worked with Anya Drabkin to ensure the questions asked don't violate any guidelines re: emotion measurement.

- **Privacy Awareness and Findability Study:** Worked with the privacy team to design and launch a survey intended to gather baseline data on the awareness and perceived findability of eight different privacy controls. This data will serve as a reference point for future surveys, to track the team's progress driving down privacy people problems. When Kathie [redacted] the team's new researcher, joined I on-boarded her to the project. She then analyzed the data and created the report.

- **Rapid Feedback Taskforce:** Arturo and others created a 'taskforce' to explore how to append surveys to negative user actions, in order to gather more data on specific, traceable experiences. I worked with them to understand the limitations of the survey pool, the research and design rules in place for soliciting user feedback, and the need to involve ICs in the process.
    - One outcome of this work was the negative user action (NUA) upsell experiments, implemented by the Experiences Platform team. This data will help the team gather data for experience personalization work, understand context around NUAs, and see what percentage of NUAs occur on violating content.
    - A second outcome is the Community to Actor Feedback project, which will alert people if something they posted is reported by others, and offer them a chance to take it down. The hope is that this will slowly build empathy and new norms of what is acceptable to post, and provide IG with an additional avenue to reduce bad experiences beyond punitive measures.

- **Idiomatic:** Continued to manage the Idiomatic contract, and edit/post monthly reports internally.
    - This half I also worked closely with Dabeet [redacted], who wanted to use Idiomatic to code the open-ends from a large sentiment survey he ran to help explain sentiment drops in CAU and GFW. Understanding the drop was a request from Adam, and Vijay [redacted] (head of Marketing Decision Science) was leading the effort. With Kristin's permission, I interfaced between him and Idiomatic, and did the paperwork to have his project included in our contract so it could start quickly.
    - Due to new data sharing requirements, I completed a renewed Third Party Assessment for Idiomatic, as well as worked with FB IT to implement single sign-on (SSO) for Idiomatic instead of passwords. Both processes were not quick, and involved a lot of coordination and fact-finding.
    - At Kristin's request, I put together a document outlining my perspective on the usefulness of open-end coding, and what I think are the best next steps as we deprecate the IG sentiment tracker.
    - Odi [redacted] (head of Customer Experience) and Sandy Dondici (VP of Product & Service Operations) asked me to document my experience with Idiomatic because they're considering hiring Idiomatic for their teams.

- **Survey review:** Continued as a survey reviewer, completing 39 SRT review assignments this half. Also served as a reviewer for the IG Safety and Community Health RFP, completing 17 reviews.

- **Interviewing:** Conducted 13 interviews this half (2x team average). Conducted interviews for all variations of skill interviews (screen, qual, quant, mixed method), plus training interviews.

- **Mentoring:** Continued formal and informal mentoring of new researchers in and outside of IG. I'm the culture buddy/mentor for Chad [redacted] (Messenger), and meet with him biweekly. I'm also sought out for guidance/expertise often on an informal basis. Examples this half include working with Kate [redacted] re: incorporating likes into Stories; Becky [redacted] re: account status surveys; Kathleen [redacted] re: the youth team's off-platform parent monitoring survey; Kate [redacted] on feed sentiment and crossposting; Caroline [redacted] re: IG sentiment measurement; Eric [redacted], re: onboarding to well-being; Rishi [redacted] re: onboarding to feed; Kenzie re: logistics of building a session-level survey; Lauren [redacted] on creator sentiment measurement; Jonathan [redacted] re: applications of his root cause analysis tool and connecting to potential IG users; Ryan [redacted] re: shopping survey troubleshooting; and Noa [redacted] On, re: IG-wide rapid feedback survey management.

- **Company response to leaks:** After IG well-being research surfaced in the September leak, I was asked to gather and document all of the work I had done over the past 3 years on Project Daisy. This included all survey configs, datasets, research plans, and reports. In addition, I worked with leads and DSS to document all the decisions I made during the Project Daisy research process, and answer their questions about the work that had been done by myself and other researchers.

Context Elements

Tenure: I joined Meta in H1 of 2017 (March) and joined my current team in H1 of 2021 (February).

Role: My role is a horizontal quant role on the Well-being Foundations team, partnering with the different subteams on IG well-being, relevance integrity, Central Integrity, and FB App Integrity. The functions I've primarily partnered with are UXR (to develop and disseminate findings), DS (to align work streams and collect data that can build and refine ML models), and leadership (to develop our IG well-being measurement strategy, and to disseminate findings).

Environment: Notable events or organizational changes that impacted individual's scope/work:

- The company's response to the data leaks this half led to a significant tightening of the rules regarding how and where we can share reports, who is allowed to see them, and how we go about socializing our findings. While necessary, they increase the amount of time it takes to develop, launch, and socialize findings. The leaks themselves also caused a good deal of thrash, in that I had to take time to document the details around a bunch of older studies for leads, in case they were leaked too.
- As the Foundations team is new and different people/functions have different opinions of the role research should play, much more time is spent building consensus compared to my previous teams.