

The Long-Term Effects of FB App Integrity Interventions

Joe [REDACTED] · Last edited Today at 6:14 AM · 29 minute read

Save ...



The Long-Term Effects of FB App Integrity Interventions

TL;DR

- A common external narrative about Facebook is that integrity and engagement are opposing forces, and that Facebook must choose between products that are 'good for the world' or 'good for business'. Short-term tests of small sets of products tend to show that integrity products have a neutral or (occasionally) negative impact on engagement. Using the Minimal Integrity Holdout experiment, I tested whether this integrity-engagement tradeoff holds over the course of multiple years and a large number of Facebook App Integrity's products (i.e., 'FI products').
- Results showed that FI products significantly improved Facebook's most important engagement metrics -- DAP/MAP, as well as several MSI engagement metrics. Though it took months-to-years for these effects to become significant, they have remained significant for the past 6 months. These DAP/MAP effects were largely replicated in the Personalized Integrity Holdout experiment (i.e., a more circumscribed version of the Minimal Integrity Holdout experiment specific to a single FI team focused on reducing bad experiences that users perceive, also a long-term holdout). There is

<https://fb.workplace.com/notes/337086114890650>

HIGHLY CONFIDENTIAL (COMPETITOR)
HIGHLY CONFIDENTIAL (ATTORNEY'S EYES ONLY)

Defendant's Exhibit

DX0154

Case Number:
D-101-CV-2023-02838

1/35

META3047MDL-005-00000261
METANMAG-002-00225551

no guarantee that these effects would extend to non-FI integrity products, future FI products, or even to these same FI products in the future; but these findings do suggest that such generalization is plausible.

- In contrast, results indicated that FI products tended to significantly reduce a major engagement goal metric commonly employed as a guardrail for integrity products: core sessions. However, FI products significantly increased core sessions among newer users, marginal users, and teens, as well as core sessions with 'positive active signals' (e.g., likes/loves, original broadcast posts, friend requests). FI products also significantly increased a discounted version of core sessions (i.e., $\log[\text{core sessions}]$).
- Overall, findings suggest that FI products tend to have positive effects on both integrity and (eventually) engagement. Recommendation: To engender an ideal Facebook ecosystem -- i.e., a highly engaging ecosystem with high integrity -- we should invest in more (and more intense) integrity products.
- These findings highlight that integrity and engagement can be positively correlated -- under certain circumstances, we may not need to choose between the two. Recommendation: We should incentivize more teams to simultaneously improve both integrity and engagement (e.g., by making engagement metrics/products more integrity-aware, as we've done with MSI).
- These findings also show that, in the context of FI products, core sessions is not an accurate long-term indicator of DAP/MAP. Recommendation: Move away from using metrics like core sessions as guardrails for integrity products; explore sessions metrics related to quality sessions (e.g., $\log[\text{sessions}]$) that better predict future DAP/MAP as potential guardrails.
- These findings also highlight how product effects on the Facebook ecosystem may change in important ways over the course of months and years. Recommendation: Prioritize the monitoring of long-term product effects across the company, expand the coverage of integrity-related long-term holdouts, and explore medium-term tests of integrity product effects on DAP/MAP.

Introduction

A common external narrative is that integrity and engagement are fundamentally opposing forces, and Facebook tends to err on the side of engagement. A more cynical version of this narrative is that integrity products are 'good for the world' but 'bad for business', and Facebook errs on the side of engagement products that are 'good for business' but 'bad for the world'. In contrast to this external narrative, Facebook's internal strategy has been framed in terms of 'user value' rather than an integrity-engagement dichotomy. Like all latent constructs, user value cannot be directly observed or directly measured. We must infer user value from indirect

measures that we create and continually refine, and different people may have different views on what user value is and how to best measure it.

Facebook has taken a multifaceted approach to user value that includes goals and metrics related to both integrity and engagement: Integrity aims to improve user value by reducing harms/bad experiences and positively influencing key individual/societal outcomes (e.g., [see here](#)); engagement related to meaningful social interactions (MSI) aims to improve user value by [improving user wellbeing](#); and engagement related to sessions/DAP aims to improve user value by [fulfilling user needs and providing a good user experience](#). Metrics in each of these domains continually evolve to better approximate user value. For example, in 2019 Facebook began goaling on sessions for conceptual and technical reasons, and because sessions predicted DAP and was correlated with positive user sentiment (see [here](#) and [here](#)). But because research showed that sessions shorter than 3 seconds and more than 50 sessions in a day were not reflective of greater user value, Facebook began goaling on 'core sessions' (i.e., sessions >3 seconds, capped at 50 sessions per day) in 2021 (see [here](#) and [here](#)). Ongoing work continues to refine this metric to make it more reflective of user value (e.g., the [quality sessions](#) work stream). For example, [very recent work](#) suggests that 'Discounted Sessions' (i.e., the 1st session of the day is weighted more than the 2nd, the 99th is weighted more than the 100th, etc.) predicts future DAP better than core sessions.

Despite their common goal, different approaches to improving user value sometimes conflict with one another. For example, studies have shown that integrity products can cause reductions in bad experiences (e.g., FI end of half holdout studies from [H2 2019](#), [H1 2020](#), and [H2 2020](#)), generate more satisfaction with News Feed ([██████████ 2020](#)), and produce greater perceived legitimacy of Facebook ([██████████ 2021](#)). But work has also shown that integrity products tend to have a neutral (or sometimes negative) effect on engagement metrics (e.g., [Engagement section of ██████████ 2019](#)). Likewise, engagement products can generate more core sessions and DAP while simultaneously increasing the prevalence of integrity harms like misinformation (e.g., [H1 2021 FAST Holdout](#) experiment). To manage such conflicting effects, engagement metrics often act as [guardrails for integrity products](#) (i.e., teams agree that integrity products cannot regress engagement metrics [such as core sessions] beyond a certain point), and [vice versa](#). This situation has led some internally to view integrity and engagement as functionally (if not conceptually) in opposition, and may help fuel the external narrative that Facebook has a choice between integrity and engagement, and often chooses engagement.

But the full relationship between integrity and engagement remains unclear, in part because most of our data on this relationship involves short-term evaluations of a single product or small set of products. It's possible that the long-term effects of integrity products differ from their short-term effects. Notably, outside of Integrity, there is precedence for differences in short-term vs. long-term product effects on engagement ([see here](#)). Short-term evaluations are limited because these timescales only represent a snapshot of activity in the [complex adaptive system](#) that is Facebook. Such systems evolve over time in response to feedback -- the relationship between integrity and engagement may involve an indeterminate number of feedback loops, each interacting with many other feedback loops, and each evolving over time. Although a complete

understanding of these loops may be impossible, many who work in Integrity have hypothesized that a general pattern should emerge from this complexity: **With enough integrity products and enough time, Integrity's effect on engagement will reach a tipping point, such that Integrity begins to increase engagement.** In other words, many who work in Integrity believe that the choice between integrity and engagement may be a false dichotomy -- under the right circumstances, it may be possible to have both.

Methods

The primary goal of this study was to test this hypothesis by utilizing the Minimal Integrity Holdout. This viewer-side holdout experiment was created on March 20, 2019 (~2.5 years ago) and includes two groups: (a) The Production Group, which receives most FI products as they're launched (i.e., this group is exposed to an ever-increasing number of integrity products over time); and (b) The Holdout Group, which only receives FI products deemed absolutely necessary (e.g., enforcements like debunking by Third Party Fact Checkers, specific demotions for countries at risk for offline violence; [see here for more information](#)).

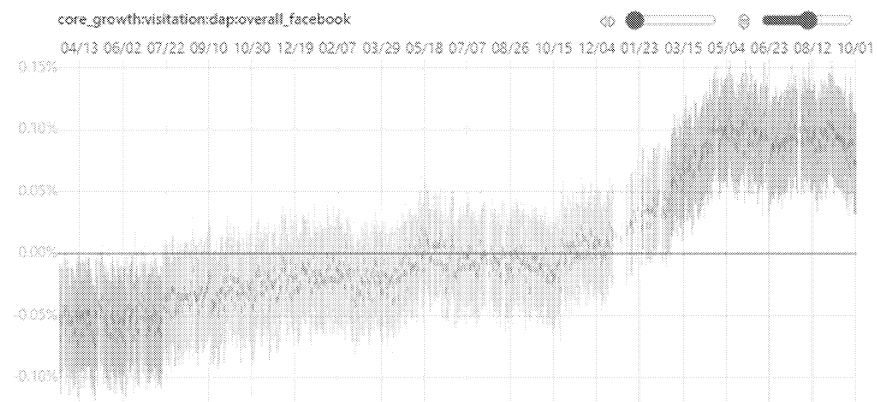
A common point of confusion is the nature of the products in this holdout experiment. So to clarify, this experiment is probably best thought of as the Minimal **Feed** Integrity Holdout experiment because it currently only includes Feed-based integrity products -- it does not include integrity products related to Groups, Watch, and so forth. It also only includes products launched to the Integrity Production Config, which means that it does not include integrity-related products launched to the Relevance Production Config, such as most civic- or health-related integrity products (e.g., products related to downstream MSI replacement, civic ranking: engagement-based worth your time, among others). Moreover, this viewer-side holdout doesn't capture the effects of producer-side interventions like informative reshare friction well, since such changes don't directly impact what the viewers in the holdout see, only what they share. The present results may not generalize to products beyond those included in the Minimal Integrity Holdout experiment.

Comparisons between the Production and Holdout Groups can allow us to observe the causal effect of FI products on engagement and integrity over time. Notably, however, these causal effects will likely underestimate the true effects of FI products because of second-order effects. This is because all users -- including those in the Holdout Group -- receive the indirect effects of FI products (i.e., a higher-integrity Facebook ecosystem), which lessens the experimental impact of FI products.

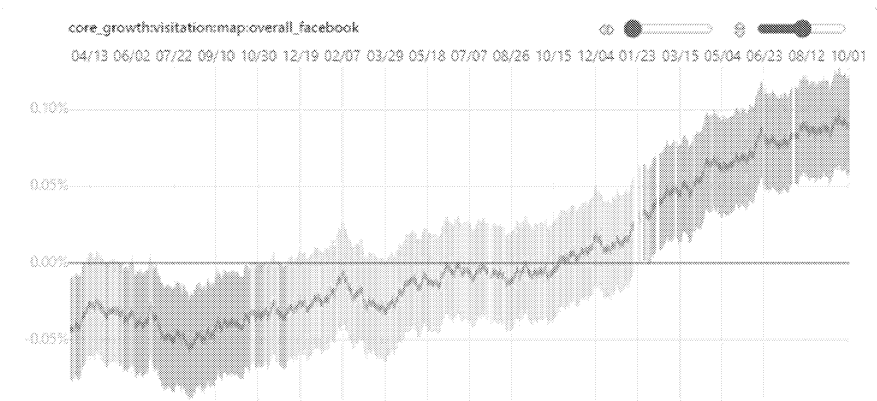
Below, I compare the Production/Holdout Groups on major metrics of interest. Each figure below is in terms of percent delta ($[(\text{Production} - \text{Holdout}) / \text{Holdout}]$), with positive numbers indicating higher values in the Production Group (i.e., the group that received all integrity launches). All confidence intervals are 95% CIs; green bars/dots = significant effect in what is usually considered a 'good direction' for the metric; red bars/dots = significant effect in what is usually considered a 'bad direction' for the metric.

Results

FI Products Improved Daily Active People (DAP) and Monthly Active People (MAP) Across 2.5 years (<https://fburl.com/deltoid3/ur4mlsq2>)



FI products gradually increased DAP across ~2.5 years (March 20, 2019 - October 1, 2021)

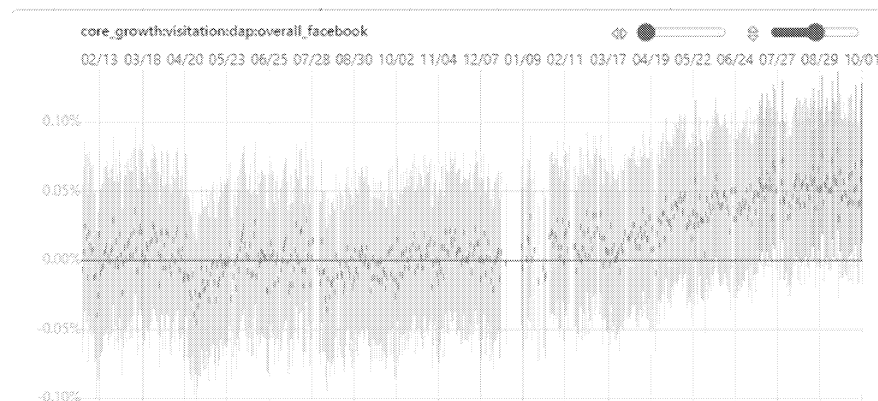


FI products gradually increased MAP across ~2.5 years (March 20, 2019 - October 1, 2021)

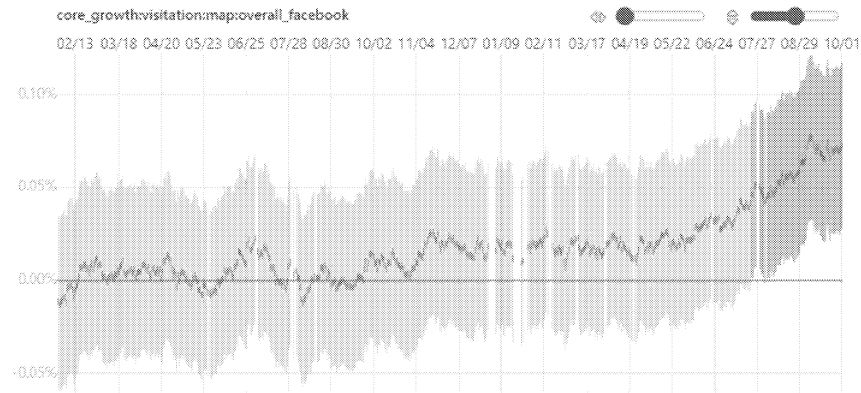
Summary. FI products have a significant and sustained positive effect on DAP and MAP. These metrics have been on an upward trajectory for nearly two years, with this trajectory becoming steeper approximately one year ago and reaching statistical significance several months ago. Of note, this finding is directionally consistent with the observation that Integrity has improved other non-Integrity specific metrics in the longer term, like News Feed Sentiment. The relative deltas for DAP/MAP are each approaching **0.1%**, which translates into roughly 2.78 million additional daily active users and 3.53 million additional monthly active users. These gains are meaningful but modest; they pale in comparison to those produced by some of the products that more directly target DAP/MAP. For example, the H1 2021 FAST Holdout experiment shows that engagement-focused products increased DAP by **0.27%**, and long-term holdouts for 'People You May Know (PYMK)' show that PYMK increased DAP by **3.5%** and MAP by **2.4%**.

Although the goal of the current project is to focus on the broad relationship between integrity and engagement, some notable moderators (aka 'cuts') of these effects include evidence that these DAP/MAP effects tended to be especially strong in emerging markets (see [here for regional analyses](#)), marginal users (i.e., L28 < 10 & sessions/day <= 15, or L28 >= 10 & sessions/day <= 10), and users aged 15-23 (which is consistent with evidence that younger users report more negative experiences on the app); DAP/MAP effects applied similarly regardless of Facebook tenure.

Of note, there's a similar gradual DAP/MAP increase in the long-term holdout experiment of a specific FI team, the Personalized Integrity team (Production = receives launches from this team; Holdout = does not; **see Figures below** top = DAP, bottom = MAP; January 2020 - Oct 2021). Compared to the Minimal Integrity Holdout experiment, this experiment has been running for less time (1.5 years; since January 2020) and includes a smaller number of integrity products (those specific to one team). Nonetheless, the Production Group shows an upward trajectory for DAP/MAP, reaching statistical significance in late July/early August 2021. These findings add to the present results showing that, in the long-term, FI products increase DAP/MAP.



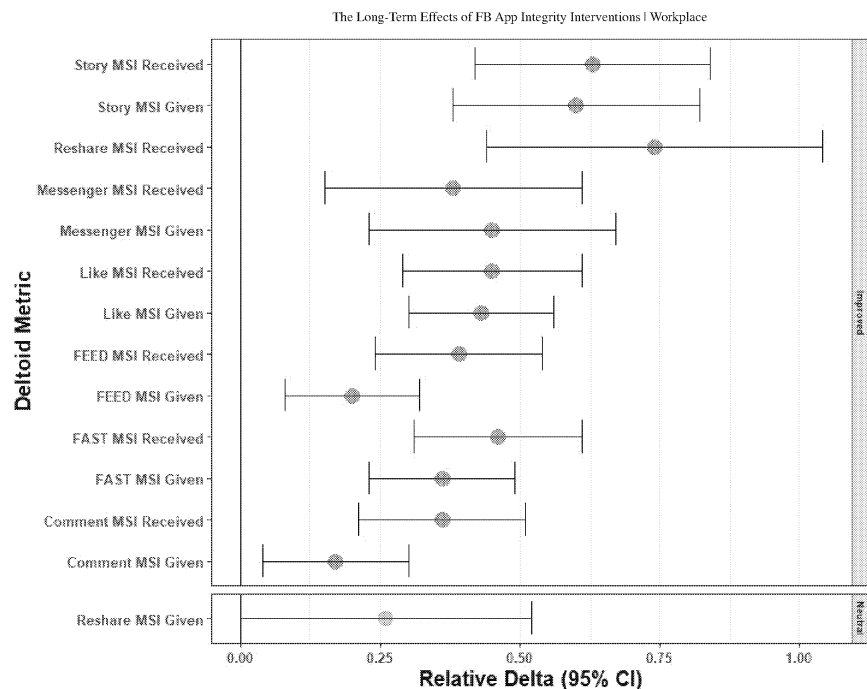
*Personalized Integrity Products have gradually improved **DAP** over the course of 1.5 years in the Personalized Integrity Holdout (i.e., the long-term holdout for a specific FI team)*



*Personalized Integrity Products have gradually improved **MAP** over the course of 1.5 years in the Personalized Integrity Holdout (i.e., the long-term holdout for a specific FI team)*

FI Products Improved Several MSI Metrics (<https://fburl.com/deltoid3/88i2bz69>)

August 19, 2021 - October 19, 2021 Results Shown Below

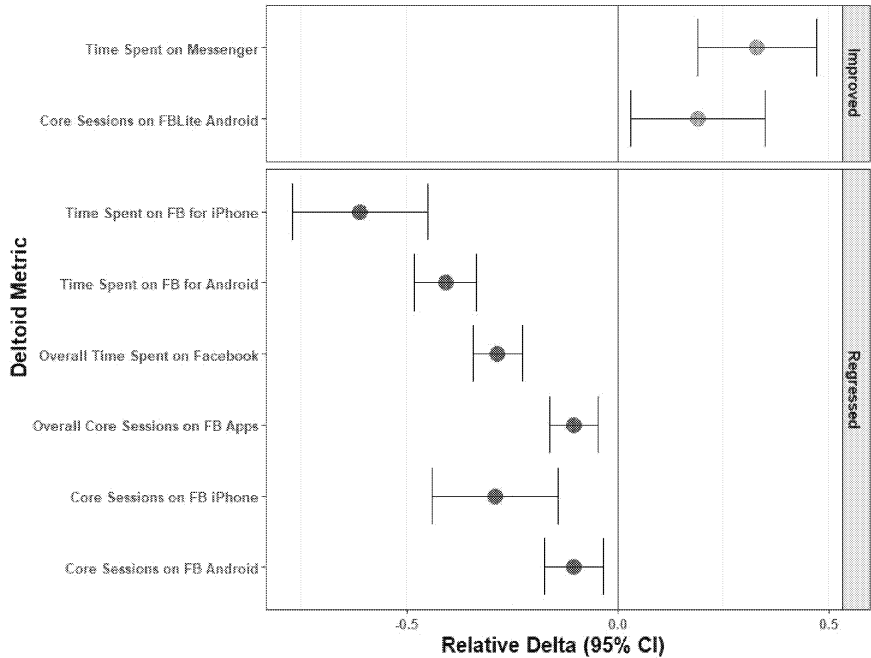


FI Products improved a wide range of MSI metrics

Summary. The Production Group showed significant improvements in most major meaningful social interaction (MSI) metrics, including MSI related to Stories, Messenger, FEED/FAST, comments, likes, and reshares (though Reshare MSI Given was not quite significant). In general, improvements tended to be stronger for MSI Received metrics than MSI Given metrics -- the lone exception to this pattern was Messenger MSI. Although data on MSI is only available in the Minimal Integrity Holdout from the past 1.5 years, it appears that MSI improvements generally preceded DAP/MAP improvements by several months. Not depicted in the above figure, but also worth noting, FI products also tended to improve metrics related to MSI ([see here](#)), such as Messenger DAP, Stories DAP, Marketplace Interactions, and original broadcast posts. This latter effect is particularly valuable because original broadcast posts have been on the decline for years. More original posts means more inventory for the platform, which means a healthier social network. But it is worth noting that the most well-known integrity launches that have negatively impacted MSI in pre-tests (e.g., comment/share term removal for Civic content) are not part of the Minimal Integrity Holdout.

FI Products had Mostly Negative Effects on Core Sessions & Time Spent
(<https://fburl.com/deltoid3/cz9g8wza>)

Oct 6, 2021 - Oct 19, 2021 Results Shown Below



FI products regressed most broad core sessions and time spent metrics

Summary. FI products had a positive impact on Messenger (both Time Spent and number of messages sent), but generally regressed time spent and core sessions on Facebook (with the exception of core sessions among FBLite Android users). Although the focus of this study was on broad effects, 16 moderators (aka ‘cuts’) of the effect of FI products on core sessions were evaluated (see [this document](#) for a summary and links to relevant Deltoid results). **The TL;DR of these cuts is that FI products tend to significantly increase higher-quality sessions (and sessions among higher-priority groups like marginal users) and significantly decrease lower-quality sessions.** For example, FI products increased sessions among newer users (i.e., 2 years or less), marginal users (i.e., L28 < 10 & sessions/day <= 15 or L28 >= 10 & sessions/day <= 10), across most non-News Feed product segments (e.g., ads, dating, gaming, stories, watch, etc.), and among teens. FI products also tended to generate an increase in sessions with positive active signals (e.g., like/love reacts, sending a friend request, original broadcast posts), shorter sessions (i.e., <1min), and sessions that included posts with high ‘Worth Your Time’ scores. In contrast, FI products significantly regressed sessions among more tenured users (i.e., 2+ years),

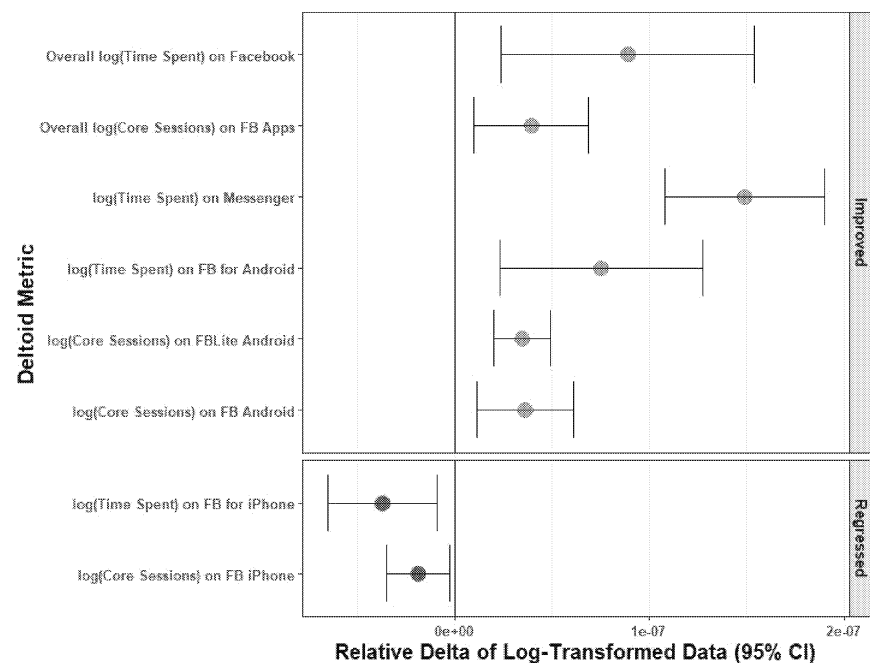
non-marginal users, and across News Feed product segments. FI products also produced significantly fewer sessions with a negative activity signal (e.g., anger react, leave group, snooze, unfollow, etc.) and longer sessions (i.e., sessions >5min). FI products' overall tendency to regress core sessions seems at odds with its positive effects on DAP/MAP.

It's also worth noting that the Personalized Integrity Holdout shows a similar but muted version of the Minimal Integrity Holdout DAP/MAP-Sessions pattern: whereas DAP/MAP have gradually and significantly increased over time (see above), core sessions has gradually and significantly decreased over time.

FI Products had Mostly Positive Effects on log(Core Sessions) & log(Time Spent)

(<https://fburl.com/deltoid3/rm5u1f1f>)

Oct 6, 2021 - Oct 19, 2021 Results Shown Below



FI Products mostly improved broad log(core sessions) and log(time spent) metrics

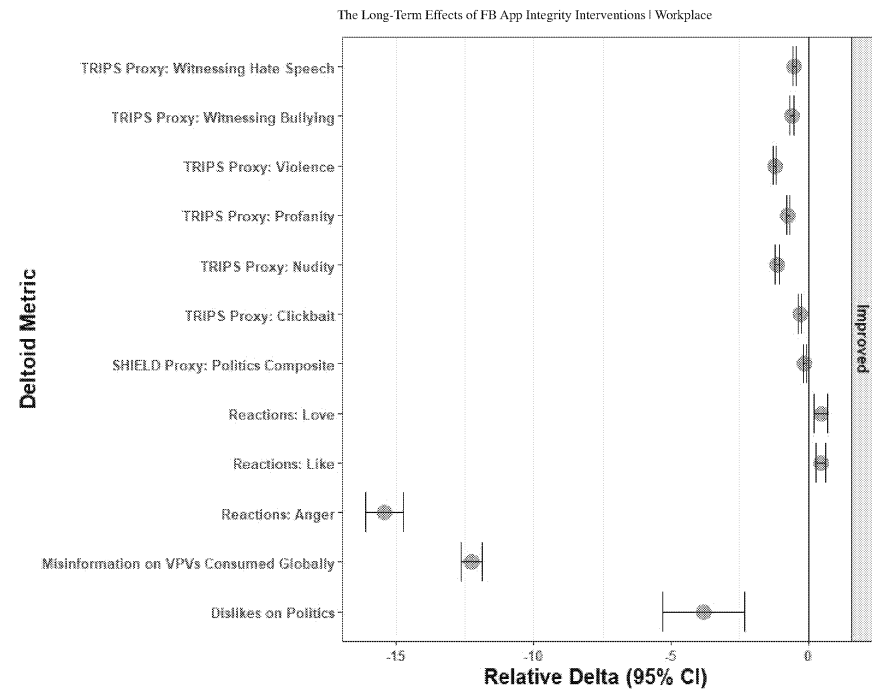
Summary. Recent work by Scott [REDACTED], Anand [REDACTED], Franz [REDACTED], Evan [REDACTED] and many others has continued to refine the sessions metric to increase its relevance to user value. One strategy involves 'Discounted Sessions', which is where subsequent sessions count less than initial sessions. One relatively simple way to discount sessions is to log-transform sessions data. Recent work by Anand [REDACTED] suggests that $\log(\text{sessions})$ predicts future DAP better than core sessions across over 800 experiments (e.g., core sessions predicting DAP two weeks later, $r = 0.29$; $\log[\text{sessions}]$ predicting DAP two weeks later, $r = 0.37$). The analyses depicted in the figure above show that FI products increase most metrics related to $\log(\text{core sessions})$ and $\log(\text{time spent})$ -- the exceptions being $\log(\text{core sessions})$ and $\log(\text{time spent})$ for iPhone. $\log(\text{sessions})$ works by more heavily weighting users' initial sessions, which might lead one to reason that we would obtain similar findings by capping sessions at 20, 10, or 5. However, FI products significantly regress sessions when they're capped at 20, and have no significant effect on sessions capped at 10 or 5 ([see here](#)). The Personalized Integrity Holdout experiment produced similar (but again more muted) effects: $\log(\text{sessions})$ were non-significantly increased (most $ps < 0.10$) in the Production Group relative to the Holdout Group.

Taken together, these findings indicate that, compared to core sessions, $\log(\text{sessions})$ is a better measure of user value both in terms of its association with DAP and its association with a higher-integrity ecosystem. Although other aspects require further study (e.g., moveability of $\log[\text{sessions}]$), these findings suggest that $\log(\text{sessions})$ may be a better overall user value metric and integrity guardrail metric than core sessions.

FI Products had Very Positive Effects on Integrity Metrics

(<https://fburl.com/deltoid3/hq881mlz> for TRIPS Proxies, Reactions, and Misinformation VPV Prevalence; <https://fburl.com/deltoid3/3mdxyh4n> for Politics-Related metrics)

Results for TRIPS Proxies, Reacts, and Misinformation VPV Prevalence are July 19, 2021 - October 19, 2021; Results for Politics-Related Metrics are October 9, 2021 - October 11, 2021 (due to availability limitations)



FI Products broadly improved integrity metrics

Summary. Across numerous integrity metrics (only 12 of which are included above for illustration purposes), FI products had a very strong and sustained impact. These effects ranged from reducing the prevalence of harmful VPVs (e.g., prevalence of misinformation VPVs globally); increasing desired behaviors (e.g., Like and Love Reacts); decreasing undesired behaviors (e.g., Anger Reacts); and reducing the perception of integrity harms (e.g., TRIPS Proxies) and dissatisfaction with politics content (e.g., SHIELD Proxy) on Facebook. These findings demonstrate that FI products' effects on integrity metrics are both broad and enduring, which may be particularly notable given that these effects are likely substantially understated due to this viewer-side holdout being unable to capture second-order effects. For example, in prior End of Half studies (e.g., see Table at below from [REDACTED] 2019), the relative reduction in integrity problems between Production and Holdout groups (e.g., 6% relative reduction in Engagement Bait) has tended to be many times smaller than the absolute reduction in the prevalence of integrity problems (e.g., a 40% absolute reduction in Engagement Bait).

Metric	Current	Product Impact	Time Series Impact
Exaggeration	2%	-10%	-60%
Withholding	2%	7%	-60%
Weak Ad Farm	2.50%	-5%	-60%
Strong Ad Farm	0.25%	Too noisy	-85%
Engagement Bait	2.50%	-6%	-40%
Useless MSI	6%	NA	-15%

Table from Turner (2019) illustrating that relative reductions in integrity harms (which holdout experiments capture; i.e., 'Product Impact') tend to be far smaller than absolute reductions in integrity harms (which represent broad improvements in the Facebook ecosystem that holdout experiments cannot capture well; i.e., 'Time Series Impact')

Are These Effects Real?

These effects -- especially the DAP/MAP effects -- may have major implications for Facebook, so it was important to evaluate the possibility that they were caused by some kind of technical artifact (vs. integrity products). To do this, I drew on Joseph [REDACTED] post on [Why DAP Moves](#), and worked with Ian [REDACTED], Joseph [REDACTED], Evan [REDACTED], Brian [REDACTED], Anand [REDACTED], Jiayi [REDACTED] and others to interrogate these effects. See [this document](#) for an overview of some of the alternatives considered (i.e., broken experiment; random noise; change in account status; feature limits, strike volumes, and penalties; and some other as-yet unidentified issue with the Minimal Integrity Holdout experiment) and the results of exploring these alternatives. **The TL;DR is that it appears that FI products genuinely caused the aforementioned effects** -- we could find no issues with the Minimal Integrity Holdout experiment, and replication of patterns in the Personalized Integrity Holdout experiment (see above) further suggest that the Minimal Integrity Holdout results are valid. Nevertheless, it is always possible (though decreasingly plausible) that some other alternative explanation accounts for these effects.

Explaining These Effects

The results indicated that, over the course of 2.5 years and with an ever-increasing number of FI products, the Production Group (relative to the Holdout Group) showed significant improvements in DAP/MAP, MSI-related engagement, and various integrity metrics. Although this study directly demonstrates these effects, it cannot directly show *why* these effects occurred (i.e., what mediates these effects). Based on indirect evidence, I hypothesize that these effects are explained by a large number of FI products acting in complex ways over a long period of time. Evidence for this hypothesis includes:

- The changes in DAP/MAP were gradual (see above).

- We might expect a more obvious inflection point if these effects were caused by a single product or small set of products, and/or if this were a short-term/immediate product effect.
- DAP/MAP improved significantly in both the Minimal Integrity Holdout experiment and the Personalized Integrity Holdout experiment (and the former includes the products of the latter), but these improvements occurred **at different times** (see above).
 - It's hard to square this asynchrony with a 'single product' or 'small set of products' explanation.
- There were no significant DAP/MAP improvements in FI End of Half AXIS/SHIELD studies (2019H2; 2020H1; 2020H2; 2021H1) that tested the effects of specific sets of integrity products over a short period of time.
 - This is inconsistent with these effects being caused by a single product/small set of products and short-term product effects.
- FI products had a broad effect on a wide range of metrics that tend to improve DAP (see the last two sections of this document).
 - This is consistent with the idea that a large set of FI products had a small impact on several feedback loops within the Facebook ecosystem, gradually generating improved engagement.
- These effects apply across the full range of 'Facebook Ages' (i.e., users who recently activated accounts and users who activated their account years ago).
 - This finding is most consistent with the interpretation that FI products have gradually transformed the Facebook ecosystem into a more engaging place. More recent users have immediately benefited from this improved ecosystem whereas more tenured users took months or years to benefit from this (i.e., until the weight of our cumulative interventions were strong enough to have an effect).

These findings converge to support the aforementioned hypothesis, but work is ongoing to more directly evaluate the possibility that a product or small set of products primarily caused the present results.

Insights & Recommendations

Several insights and recommendations can be gleaned from these findings. Below, I briefly discuss what I think are the broadest and most important takeaways:

Insight 1: FI products have positive effects on both integrity and (eventually) engagement. FI products had a powerful and sustained impact on a broad range of integrity metrics. They also

had a positive sustained impact on log(sessions), MSI, and DAP/MAP-related engagement. In short, FI products add substantial long-term user value by cleaning up the Facebook ecosystem and making it more engaging. The catch is that it seems to take months or years for FI products to have this dual transformational impact on the Facebook ecosystem. It is, of course, unclear if the present findings would generalize to products **not** included in the Minimal Integrity Holdout, but these results suggest that it is reasonable to hypothesize that other integrity products may have similar effects.

Recommendation 1: Invest in more integrity products, especially for marginal users and youth.

When we build integrity into the fabric of our products (vs. using integrity simply as a guardrail), we can produce positive effects on both integrity and engagement. To maximize user value and engender an ideal ecosystem -- that is, an ecosystem that is both highly engaging and has high integrity -- we would benefit from investing in more (and more intense) integrity products and from recognizing that their positive effects on engagement may take months or years to become apparent.

Integrity products directed at two groups of great interest to Facebook -- youth/young adults and marginal users -- may be particularly fruitful. Youth/young adults are vital to Facebook's continued success, but have driven engagement softness over the past several months ([see here](#)). Marginal users are by definition less engaged users, and improving their engagement is highly valuable for Facebook. A 2020 analysis estimated that improving sessions among marginal US users is nearly 10x more valuable than improving sessions among non-marginal US users: moving a US user from 1 to 10 sessions is worth \$0.37 and moving a US user from 50 to 60 sessions is worth \$0.04. Notably, engagement-focused products tend to most powerfully increase sessions among non-marginal users ([see here](#)). FI products tended to improve both integrity metrics and engagement metrics (including core sessions) among marginal users and youth/young adults, indicating that FI products improved overall user value in these groups. Youth/young adults and marginal users appear to be particularly responsive to FI products -- investing in more (and/or more intense) integrity products directed at these two groups may be an especially worthwhile direction.

Insight 2: It's possible for products to simultaneously improve integrity, MSI-related engagement, and sessions/DAP-related engagement.

Some popular narratives about Facebook are rooted in the idea that we must choose between integrity and engagement. But as the present findings make clear, for a wide swath of Integrity on FB App, this is a false dichotomy: we don't need to choose between integrity and engagement, and we don't need to develop products that primarily target just one of these domains. To maximize user value, we may be best off designing products to simultaneously target integrity, MSI-related engagement, and sessions-related engagement.

User value is a latent construct, meaning that it is not directly observable and may not have any universally agreed upon boundaries or necessary/sufficient features. In other words, user value

means different things to different people (and possibly different things to the same person at different times). At Facebook, integrity, MSI-related engagement, and sessions-related engagement have each been associated with user value. One way to think about *this version* of user value is in terms of an emergent model (cf. [application of emergent models to emotions](#)). In an emergent model, a phenomenon (e.g., user value) emerges from/is caused by other phenomena that can be correlated, but are not necessarily correlated (e.g., integrity, MSI-related engagement, sessions-related engagement). An emergent model of this version of user value is depicted in the figure below.



An emergent model of user value. According to this model, user value emerges from Integrity, Sessions/DAP-related engagement, and MSI-related engagement. User value is maximized when all three domains are activated; to the extent that a product improves one domain (e.g., sessions/DAP) but regresses another (e.g., integrity), its effect on user value is diminished.
(Figure Credit: Susy [REDACTED])

Based on this model, we can make the following inferences:

- Integrity, MSI-related engagement, and sessions-related engagement can each generate user value
- User value is maximized when all three of these domains are activated; ideally, products would simultaneously maximize all three domains
- Integrity (at least in terms of FI products) improves MSI- and sessions-related engagement, greatly improving user value in the process
- To the extent that a product focused on one of these domains (e.g., sessions-related engagement) regresses another of these domains (e.g., integrity), that product's

positive effect on user value is diminished

Recommendation 2: Incentivize products that improve overall user value via multiple routes (i.e., not just via integrity or MSI or Sessions/DAP). A few existing products and work streams have moved beyond an integrity-engagement dichotomy to directly address multiple domains. For example, MSI metrics have been continually modified to filter out low integrity interactions (e.g., see here); efforts to improve the integrity of Top-N content (see here and here) aim to ensure the integrity of Facebook's most engaging content; Integrity has joined with News Feed to reduce non-recommendable content (e.g., Engagement Bait) among the most widely viewed content on the platform; and work on products like p(Inspire) might also be considered attempts to present users with engaging content that has integrity. But it remains uncommon for products and work streams to explicitly attempt to improve both integrity and engagement. Doing so may greatly facilitate efforts to improve overall user value. For example, a product that causes small-to-moderate improvements in both integrity and engagement may generate greater user value than a product that causes very large improvements in just one of these domains. As described in greater detail here, one way to help accomplish this might be to intentionally develop products that maximize pleasant tension.

Insight 3: In the context of FI products, core sessions is not an accurate indicator of long-term effects on DAP/MAP. Core sessions is often used as a guardrail for integrity products, in part because it is thought to be an accurate indicator of how a product will affect more important engagement metrics like DAP/MAP (see e.g., general association between core sessions and DAP), which typically take longer to move significantly. But there are documented instances where sessions and DAP are inversely correlated (e.g., here and here), and the present results add to this list. For FI products, core sessions may be misleading -- it might suggest that a product lowers engagement broadly when it significantly improves other engagement metrics (e.g., MSI, DAP/MAP).

Recommendation 3: Move away from core sessions as a guardrail for integrity products; consider metrics related to quality sessions instead (e.g., log[core sessions]). Research is ongoing to understand quality sessions, and metrics related to quality sessions may ultimately prove to be more relevant to both integrity and engagement than metrics like core sessions. For example, recent work suggests that, compared to core sessions, 'Discounted Sessions' (i.e., each daily session is weighted less than the prior session) may be a better predictor of future DAP. As shown above, FI products significantly improved a specific form of Discounted Sessions called log(core sessions), and these improvements were correlated with improvements in higher-quality core sessions (and reductions in lower-quality core sessions). Other recent work suggests that core sessions does not incentivize growth for marginal users -- one of the most important and valuable types of users to grow. Because log(sessions) focuses primarily on a user's first several sessions, it may more effectively incentivize growth for marginal users. The relationship between sessions and monetary value seems to follow a logarithmic function (see here); the present

results suggest that the relationship between sessions and user value may also follow a logarithmic function. In any case, by virtue of its stronger associations with both DAP and integrity, $\log(\text{core sessions})$ seems to be a better reflection of user value than core sessions and, correspondingly, may be a better product evaluation metric. Continued work on quality of sessions may generate a sessions metric that approximates user value even better than $\log(\text{core sessions})$.

Insight 4: Short-term product evaluations may miss crucial product effects. Most Facebook products are designed to be an indefinite part of the Facebook ecosystem, with many products sticking around for decades. But most of these same products are evaluated based on how they affect the Facebook ecosystem over the course of a few weeks. This mismatch can be problematic when a product's effect on the ecosystem changes over time (and/or when the ecosystem itself changes over time). The evolution of FI products' effects on the Facebook ecosystem is likely just one example of how product effects can change over time ([see here for a couple of additional examples](#)) -- we currently have little insight into how common such evolutions are. However, it's not feasible to make product launches contingent on long-term evaluations. This would greatly limit Facebook's agility, and it's always possible that product effects will change after the long-term evaluation period.

Recommendation 4: Prioritize monitoring of long-term product effects on the Facebook ecosystem. This long-term information should be weighed heavily when making decisions about changing past products, launching similar products, and developing new products. Ideally this would occur for all major products, but at the very least it would be helpful for this to occur with collections of similar products, as with the Minimal Integrity Holdout. In the absence of such prioritized monitoring, Facebook is forced to rely on the tenuous assumption that short-term effects generalize to long-term effects.

Insight 5: Medium-term tests may allow us to detect integrity product effects on DAP/MAP. The effect of FI products on DAP/MAP was significant for users whose accounts were only 3-12 months old. One possibility is that, despite the speed of these effects on users, the effects of new integrity products on users would still take a year or longer to detect. This is because, in this scenario: new integrity product → FB ecosystem effects (several months to manifest) → effects on DAP/MAP (a few months to manifest). Another possibility is that the speed of these effects on users indicates that it may be possible to detect effects on DAP/MAP in a shorter time frame (e.g., 3-6 months) -- this may be especially likely for collections of products (vs. a single product). In this scenario: new integrity product(s) → FB ecosystem effects (a few months or less) → effects on DAP/MAP (a few months). It is currently unclear which of these scenarios would be most accurate, and the scenario may vary depending on the nature of the integrity product(s) and the

state of the FB ecosystem at the time. But whatever the case, these findings indicate that it may be possible to observe significant effects of integrity products on DAP/MAP in medium-term tests. Should this approach bear fruit, it may allow integrity products to more directly estimate broader engagement effects rather than estimating these effects with sessions.

Recommendation 5: Explore the possibility of medium-term tests for integrity products. Such tests may be most viable for large collections of products, such as those typically included in FI-wide end of half studies.

Acknowledgements

- The findings above reflect the hard work, ingenuity, and dedication of hundreds of people working in FB App Integrity and related groups over the past several years. This note simply describes an outcome of your efforts.
- Special #thanks Evan [REDACTED] for setting up the Minimal Integrity Holdout experiment and providing assistance throughout this project.
- #thanks Ian [REDACTED] and Samantha [REDACTED] for invaluable advice, feedback, and guidance on this project.
- #thanks Brian [REDACTED] Joseph [REDACTED] Jiayi [REDACTED] and Anand [REDACTED] for helpful feedback and technical assistance on this project.
- #thanks Apala [REDACTED] Anna [REDACTED] Panagiotis [REDACTED] and David [REDACTED] for feedback on and support for this project.

 259

90 comments

 Like

 Comment

 Share

Most relevant ▾



Sarah [REDACTED]

Camille [REDACTED] Komal [REDACTED] Jan [REDACTED] Anthony [REDACTED] Swarup [REDACTED] Ksenia [REDACTED] Pamela [REDACTED] Swati [REDACTED] (PM), Nir [REDACTED] Robert [REDACTED] Albert [REDACTED] Rochelle [REDACTED] Josiah [REDACTED]

"These findings highlight that integrity and engagement can be positively correlated -- under certain circumstances, we may not need to choose between the two. Recommendation: We should incentivize more teams to simultaneously improve both integrity and engagement (e.g., by making engagement metrics/products more integrity-aware, as we've done with MSI)."

2y Like Reply

3 

Panagiotis

The analysis is quite insightful and the results are really encouraging about the long-term impact of our integrity efforts on user retention.

This analysis is quite relevant to recent discussions we have with various partner teams on trade-offs between short-term and long-term user engagement.

CC **Andre, Joey, Hong, Amit**

2y Like Reply

10  **Sam**

This note is nothing short of revolutionary. It puts to bed the canard that we need to choose between Integrity and Growth. Rather we should focus on long term value and retention, and work to improve the proxy metrics we use to measure our incremental progress. The work we're doing to improve the Core Sessions metric (cc **Scott**) seems more relevant than ever. We should also unblock Integrity ranking work that might regress Core Sessions in favor of other indicators of long term value, such as reduction in bad experiences and misinformation, and improved sentiment (cc **Nitin** for WYT and FSQ work).

2y Like Reply

23  **Sam**cc **Kaushik Ryan Morgan Guy** for Social Impact and Integrity

2y Like Reply

2 **Matt** 

These findings should be game-changing. I'll now be citing this in most of my future work. Great job!

2y Like Reply

8  **Matt** 

Kristen, **Joo**, **Jason**, **Billy**, **Joke**, **David**, **Alex**, **Tom Wynter** -- new longitudinal research shows that integrity promotes long-term engagement/growth.

2y Like Reply

6  **Alex** 

We've been talking about this kind of analysis for literally YEARS; I'm so so happy you've produced this work – and also that the results look so promising!

I think this is a very interesting debate we'll continue to have over the "value" of time spent versus all the other metrics you've called out here. Thank you in particular for digging into heterogeneous effects, because we've always struggled to look beyond means and understand the true, nuanced impact of our work.

2y Like Reply Edited

21  

- Ravi** 
Great analysis that I predict will be cited widely in the AI org as a way that our integrity and relevance work is complementary and essential for driving longer term product outcomes. cc **Ram Patrick Ali Vishal Nipun Yue Cristian Chunlei Dan Harry Nate**
- 2y Like Reply Edited 12  
- Jacky** 
Olivia  **Morgan**  **Mark**  -- long run session impacts
- 2y Like Reply 
- Morgan** 
FYI, this is consistent with earlier results that showed that Friending Integrity was also growth-positive <https://docs.google.com/.../1l6Eg-DsiHv8uU8M.../edit...>
-  Connect Google Drive account to see this preview 
- 2y Like Reply 10  
- Wilson** 
Nick  **Nicole**  **Yiwei** 
- 2y Like Reply 4  
- Nicole** 
Wilson  thanks for the tag this is the perfect evidence for the value of promote the good and we can expand this to think ecosystem wide between businesses and people too.
- Jeffrey**  tagging for consumer value this is a good long-term vision for how we could apply meaningful conversion metric down the line that combines ACE and BAE and other signals.
- Indira**  great piece we can draw upon as a way to think about our collaborations across Growth & Integrity 🤝 But also something long-term to discuss about how legitimacy and trust interact.
- 2y Like Reply
-  Write a reply...

Jan [REDACTED]

Thank you **Joe** and everyone involved for compiling these thoroughly vetted findings!

Assuming that predicting long-term retention is a difficult task in general (for which we might lack adequate quantitative metrics), have we considered running something like a prediction-market-based holdout in which we only track those integrity experiments that FB integrity employees generally believe might improve long-term retention?

(Along the lines of what has been done in the field of psychology
<https://www.science.org/doi/10.1126/science.aac4716> and at Google <https://anaandjelic.typepad.com/.../googlepredictionmarke...>)

2y Like Reply

8

Alex [REDACTED]

Jan [REDACTED] that's honestly a cool idea

2y Like Reply

Jan [REDACTED]

To elaborate, maybe we'll find that employees' aggregated intuition is a better predictor of long-term outcomes than the short-term two-week-experiment effects we often consider in experiment review. If so, it might help us evolve decision-making at Facebook from the narrow "data wins arguments" mantra, which might have served the company well to make growth-oriented decisions, but no longer does.

2y Like Reply Edited

3

Kelsey [REDACTED]

Jan [REDACTED] This is a really cool idea and reminds me also of Eva Vivalt and Stefano DellaVigna's work and their new forecasting platform <http://evavivalt.com/.../announcing-the-launch-of-the-...>

EVAVIVALT.COM

Announcing the Launch of the Social Science Prediction Platform!

2y Like Reply

Write a reply...

Glenn [REDACTED]

Two words: redefines expectations. A few more words: redefines how we understand and build successful Meta products

2y Like Reply

8

Ian [REDACTED]

CC Ashish [REDACTED] Nino [REDACTED] Gayatree [REDACTED] John Hegeman Kayla [REDACTED] Dan [REDACTED] Chengcheng [REDACTED] Gabriela [REDACTED] FYI

2y Like Reply Edited

2

- Nicole** [redacted]
Yinka [redacted] this may be useful for the trust work and collaboration between Integrity & Growth!
- 2y Like Reply 2 🍌
- Jake** [redacted]
Mei [redacted] **Charlie** [redacted] **Peng** [redacted] **Mohammad** [redacted] on long-term effects of integrity interventions. Interesting that most long-term effects are positive (except session)
- 2y Like Reply 🍌
- Tatiana** [redacted]
 cc: **Maria, Pany, Nic, Ardee, Emily, Falis**
- 2y Like Reply 🍌
- Sam** [redacted]
 cc **Rebecca** [redacted]
- 2y Like Reply 2 🍌
- Michael** [redacted] 🔒
Yasho [redacted] on the quantifying the value of integrity using holdouts
- 2y Like Reply 🍌
- Apala** [redacted] 🔒
 Thank you **Joe** [redacted] for doing this analysis and vetting this analysis endlessly before publishing it and thank you **Samantha** [redacted] for providing him guidance all along. I could not be prouder of both of you ❤️
- And thank you to all who have done the important work that's reflected and validated in this analysis.
- 2y Like Reply 4 🍌
- Sarah** [redacted] 🔒
 CC: **Courtney** [redacted] and **Katie** [redacted]
- 2y Like Reply 2 🍌❤️
- Apala** [redacted] 🔒
 Regarding the question , 'if the effects of DAP/MAP growth are results of strong interventions in late 2020/H1 2021 as opposed to cumulative interventions over a period of time' -
- There were no significant DAP/MAP improvements in FI End of Half AXIS/SHIELD studies (2019H2; 2020H1; 2020H2; 2021H1) that tested the effects of specific sets of integrity products over a short period of time.
 This is inconsistent with the effects mentioned in the note being caused by a single product/small set of products and short-term product effects.
- 2y Like Reply 2 🍌🍌

-  **John** 
cc **Prathyusha**
2y [Like](#) [Reply](#) 2 
-  **Shirui** 
cc: **Jiaqi** , **Sai** , **Aniket** 
2y [Like](#) [Reply](#) 4  
-  **Erin** 
cc: **Billy** , **Prerna** , **James** , **Justin** 
2y [Like](#) [Reply](#) Edited 2  
-  **Billy**  
Erin  Wow, what a wonderful and timely piece of research for our roadmap discussions!

How easy is it for us to extend the study to Watch? This would be so so so powerful!!

cc **Rui, Albert**
2y [Like](#) [Reply](#) 
-  Write a reply...
-  **Albert** 
cc: **Robert** , **Eugene** , **Praveen** , **Prerna** 
 **Mohammad** , **Srivatsan** 
2y [Like](#) [Reply](#) 3 
-  **Srivatsan** 
super cool. i am curious to trends for key integrity metrics in the 2 long term holdouts so that we can better tie cause and effect. also curious to see sentiment trends for this to see if sentiment moved in line with DAP.
2y [Like](#) [Reply](#) 3 
-  **Devra** 
Super neat research!!!! I'm curious if you account for people that have left the platform entirely. If there is differential attrition, then the results would overstate (if more dissatisfied people left from the treatment group) or understate (if more dissatisfied people left from the holdout group) the true effect. Given the long time horizon, the results accounting for differential attrition could be quite different than those that don't.
2y [Like](#) [Reply](#) Edited 3 

Morgan  **Devra**  this is difficult to study directly because of anonymization. I looked into that question a bit, but I don't think it's possible to conclusively answer over long periods because, for users with deactivated/deleted accounts, we do not retain a log of whether the user themselves chose to leave or FB deleted their account.

There is a small but stat sig decrease in deleted accounts (-0.01%). But it's hard to say what that means or root cause it further.

2y Like Reply 

Devra  **Morgan**. -0.01 is pretty small, but so were the substantive differences between groups. And hard to know which way any differential attrition would go if due only to real users deleting accounts. From reading, the treatment seems to affect just ranking and wouldn't include things like FB deleting fake accounts. Is that correct?

2y Like Reply Edited

 Write a reply...

Matt   I love the finding "FI products also tended to improve metrics related to MSI (see here), such as Messenger DAP, Stories DAP, Marketplace Interactions, and original broadcast posts. This latter effect is particularly valuable because original broadcast posts have been on the decline for years. More original posts means more inventory for the platform, which means a healthier social network." This echoes repeated other findings (@ me for deets) that creators curb their posting and commenting behaviors due to prior bullying and harassment experiences.

2y Like Reply

8  

Emily  **Damon** for a good integrity <> growth win.

Vincent to our convo about need for more long term measurement

2y Like Reply Edited

3 

Damon  **Emily**  thanks! this has been making the rounds in DS land 😊. In growth, we have some similar, if less crisply causal stories too so it's really nice to see this in engagement land <https://fb.workplace.com/groups/366391746742780/posts/6966332773415278>

2y Like Reply

3  

 Write a reply...

Apala  
Andreas 

- 2y Like Reply 2
- Apala**
Jimmy
- 2y Like Reply
- Talal**
Jason
- 2y Like Reply 2
- Talal**
David
- 2y Like Reply
- David Gillis**
Wow – This is super interesting and heartening work. I especially like the emergent model of user value, and think we are seeing a pattern of systemic > silver bullet solutions leading to improved experiences for people on our platforms (e.g. also thinking of multi-factor approach to meaningfully reducing hate and hostile speech prevalence).
- 2y Like Reply 7
- Scott**
This is great!
- 2y Like Reply 2
- David**
cc **Ed** - "These findings highlight that integrity and engagement can be positively correlated -- under certain circumstances, we may not need to choose between the two. Recommendation: We should incentivize more teams to simultaneously improve both integrity and engagement"
- 2y Like Reply
- Corinne**
Chris - consider this approach for showing fairness investments improve product
- 2y Like Reply
- Andreas**
This is such interesting work -- thank you for sharing this note. Much to think about!
- 2y Like Reply
- Caner**
cc: **Stefan, Cristian, Ivan**
- 2y Like Reply
- Stefan**
CC **Ali** **Ram** **Cornelia**
- 2y Like Reply

-  **Ram** 
cc **Nipun** 
2y Like Reply 
-  **Patrick** 
cc **Amit, Joey, Jason, Jerome, Srinivas, Maria**
2y Like Reply 3  
-  **Jerome** 
Patrick Pantel this looks like an amazing note that I think we should socialize a lot more. But let me read it in full first.
2y Like Reply 2  
-  **Jason Gaedtke**
cc **Guy Rosen, Veronika Belokhvostova, Michael Salem**
2y Like Reply 
-  **Glenn** 
Jerome  wait, you are reading something before resharing it? Do you know how to use social media? 
2y Like Reply 2  
-  **Vini** 
cc: **Michael Tamar Gemma Maxime David** FYI for a fantastic articulation of the net positive effects of the integrity interventions we enable on growth
2y Like Reply 
-  Write a reply...
-  **Andrew** 
I would love for some of these results to be shared in our public blog or newsroom.
2y Like Reply 4  
-  **Sarah** 
Andrew +100
2y Like Reply 
-  Write a reply...



David [REDACTED]

Joe [REDACTED] I want to say off the bat thank you for such a thoughtful and well sourced note.

I do want to raise the following concern. You note that

We might expect a more obvious inflection point if these effects were caused by a single product or small set of products, and/or if this were a short-term/immediate product effect.

There were no significant DAP/MAP improvements in FI End of Half AXIS/SHIELD studies (2019H2; 2020H1; 2020H2; 2021H1) that tested the effects of specific sets of integrity products over a short period of time.

FI products had a broad effect on a wide range of metrics that tend to improve DAP (see the last two sections of this document).

I highlight these conclusions because there is a distinct risk that the lack of clear signal actually suggests something else is going on. That is, from an internal validity POV the absence of a clear inflection point in an interrupted time series analysis would suggest the long term trend could also be due to other concomitant historical events.

Edit: to add some clarity when we look at product launches we're looking at patterns in concomitant, interrupted time series design. The lack of clear intervention effects make me concerned that something else is going on here. Namely some kind of *Treatment-related attrition*. Attrition refers to any loss of participants from posttest measurement who were initially randomized whether because of refusal to answer questionnaire items, failure to show up for the posttest measurement session, moving away from the study site, or the like. Particular concerns with respect to internal validity arise when attrition is related to treatment assignment. For example, students with low math ability may drop out of the innovative new program, whereas students with high math ability may drop out of the standard program.

Or in this case people who stop using FB but keep using messenger.

Here's what I mean:

Going to back to Joe's **note** that you cite:

One really important point is that Overall Facebook visitors includes both Facebook App (Blue) users and Messenger users. So, if a user just uses Messenger, they're included in Overall Facebook DAP.

Given that there aren't any movements due to clear product changes or other integrity launches, what if the shift is due to people stopping using Blue, while still using Messenger.

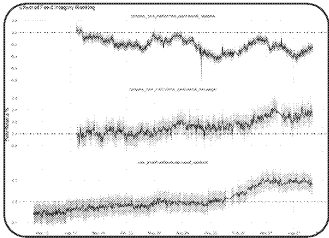
Based on the Deltoid links you provided above (again thank you for the links), we have some evidence that speaks to this in the note and in the deltoid links: Time Spent Overall on Facebook is down in production relative to holdout, while Time Spent Overall on Messenger is up. Importantly, the period of increase in visitation:dap:overall_facebook coincides with improvements in Time Spent on Messenger, see the attached figure.

Now time spent does not equal DAP but given that DAP in the figure above is made up of people using Blue and Messenger it would be good to decompose the effects to make

sure that FI helps improve DAP by keeping people coming back to BLUE and not because of movements on Messenger only.

We do have metrics for dap break down by Blue and Messenger and believe those will better answer the question about DAP movements for Blue app by itself.

- core_growth:visitation:dap:overall_blue
- core_growth:visitation:dap:messenger_overall



2y Like Reply Edited

2

Joe [REDACTED]

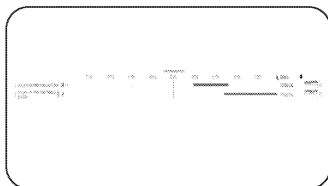
David [REDACTED] Thank you for raising this issue and thoughtfully explaining your reasoning. There's been a major SEV affecting Deltoid, so I've only just been able to dig into a few analyses. I'll try to succinctly give my thoughts below:

(1) It's important to note that many (if not most) people have disagreed with my view that there's no obvious inflection point in the DAP time series. I presented a daily time series, which is bound to look fairly smooth and make it harder to see major discontinuities. But even within this graph, several have pointed out to me that there appears to be an inflection point around mid-October or possibly mid-November of 2020. I refer to this in passing in the 'summary' section under the DAP/MAP graphs as a 'steeper trajectory' over the past year. But I'm afraid that my language in the the section that you noted was less precise and did not fully incorporate the observation of a steeper trajectory starting in late 2020. So the premise of there being no inflection point (and more reason to consider a potential explanation like differential attrition) may not be supported.

(2) However, it's still possible that differential attrition may play a role, at least to some degree. On this point I would defer to **Joseph [REDACTED]** and **Morgan [REDACTED]** both of whom have done some terrific work on addressing this kind of question in general (and also worked through some of this issues in the Min Integrity Holdout). I believe that Morgan noted some of his findings/thoughts in a comment above, and results on Joe [REDACTED]'s 'change in account status' metric can be found here: <https://fburl.com/deltoid3/2fap33hv> (i.e., null effect on this metric). At this point, it doesn't look like differential attrition plays a role here (or at least a big enough role to fully explain the DAP findings), though it may be difficult to fully rule out.

(3) In terms of the overall Facebook DAP effect being driven by Messenger DAP, this does not appear to be the case (see attached for screen shot, and Deltoid query here: <https://fburl.com/deltoid3/w74j7m50>). I apologize for omitting these analyses from the note. I believe that they were in a prior draft, along with a greater focus on FI products increasing Story DAP, Marketplace Interactions, etc -- but the note was getting far too long!

I referenced this briefly in the note, but we are very much continuing analyses to try to identify what specific caused these effects on DAP/MAP, MSI, log(sessions), integrity metrics, etc. My default hypothesis has been that it's a complex combination of many things acting over a long period of time, but we may end up finding that the cause is more specific (hopefully specific products and not technical artifacts). Thank you again!



Rachel [REDACTED]

Wow. This is so fascinating! A couple questions **Joe**:

1. A lot of the scrutiny is not just about engagement but also about FB *profiting* from x (hate speech, misinfo, etc). Engagement may be a proxy but do we have **#s** for that? (totally get why we might not publish that and happy to offline this)

2. Do we have an understanding of what some of the biggest categories of product drivers are for these metric moves? Very interesting to see bigger effects with youth / marginal users but what are the biggest things that worked that we should be doing more of?

2y Like Reply

3

Ian [REDACTED]

Hi **Rachel** [REDACTED] nice seeing you here! Re: #2, here is a double-click we did into Young Adults (Integrity perception, prevalence, and the impact of our work to date), including some thoughts on future next steps:
<https://fb.quip.com/B3KHAeo8UdV0>

I don't think we have numbers on hand that say "doing intervention X causally increased engagement metric Y for Young Adults" but do think working to provide them with more/stronger protections is very worthwhile (very positive for Integrity, and likely at worst neutral on positive value as estimated through engagement metrics). CC **John** [REDACTED] **Meng** [REDACTED]



QUIP

Prevalent Integrity Problems for Young Adults in US/CA

2y Like Reply Edited

3

Rachel [REDACTED]

Ian hi!!!! So good to see you (albeit in a post!)

2y Like Reply



Write a reply...

Vishwanath [REDACTED]

Many have said already in better ways how amazing this note is. Great job to everyone sourcing the diverse analysis on this to present a very comprehensive picture.

I know this is scope to FI products, but I thought I'd add here one more data point, the note of which I can't find anymore - during our initial exploratory analysis back in 2015 for some of the work on women in emerging markets, we ran similar long term cohort analysis which showed correlation between people who had more bad experiences also had less long term retention and churned more. 😊

2y Like Reply

8

Vishwanath [redacted]
Ahh, **Shankar** [redacted] and **Emily** [redacted] who were folks who worked on it. 😊

2y Like Reply



Shankar [redacted]
this was my analysis (<https://fb.workplace.com/notes/405140920879464/>) and this is Emily's (intriguingly called Part 1 but I can't find Part 2): <https://fb.workplace.com/notes/2717020245226888/>

Long-term retention and gender gap in India: What do we know? Do we know things? Let's find out! (LTRAGGIII! WDWK? DWKT? LFO!)

2y Like Reply



Write a reply...

Will [redacted]
cc **Christian, Roxie, Michelle, Ellery, Amanda**

2y Like Reply



Eugene [redacted] ✓
For years we've guessed this might be the case for improving ecosystem quality on different products, but this is the first time I've seen the correlation clearly measured. Thank you for doing it!

2y Like Reply



Becky [redacted] ✓
Gaurav Jonathan Elie Fritz Santhosh useful food for thought as we devise success metrics for Account Status
There could be some nuances to pay attention to when using engagement metrics as guardrails, especially Core Sessions.

2y Like Reply Edited



Jake [redacted]
Really insightful overview, thanks **Joe** for your work on this and spelling out your findings/recs here so clearly!
These findings highlight that integrity and engagement can be positively correlated -- under certain circumstances, we may not need to choose between the two. Recommendation: We should incentivize more teams to simultaneously improve both integrity and engagement (e.g., by making engagement metrics/products more integrity-aware, as we've done with MSI).

cc; **Priscilla** on engagement/integrity trade-offs in FBAI

2y Like Reply



Miranda [redacted]
Excellent, really thoughtful work. I really really appreciate it

2y Like Reply

2   Mia [REDACTED]

Really compelling work and super helpful insights!!

2y Like Reply

 Mia [REDACTED]

cc Liz [REDACTED]

2y Like Reply

 Melissa [REDACTED] 

Joe [REDACTED] Is it possible to make this post more widely viewable? There is a lot of knowledge and insight here that would be valuable for folks outside of FB Integrity.

2y Like Reply

2  Parisa [REDACTED] 

Jacqueline [REDACTED] Emily [REDACTED] [REDACTED] Tangent but related study to the earlier discussion of long term business effect

2y Like Reply

 Matt [REDACTED] 

Stephen [REDACTED] this is the fantastic paper I mentioned about long term effects of FB integrity interventions.

1y Like Reply

 Marie [REDACTED]

cc Jeffrey [REDACTED]

1y Like Reply  Rachel [REDACTED]

cc Max [REDACTED]

1y Like Reply

 Max [REDACTED]

cc Gavin [REDACTED]

1y Like Reply

 Yaniv [REDACTED]

Shai [REDACTED]

1y Like Reply

 Karin [REDACTED]

cc: Tsatska (Taz) [REDACTED] Rachel [REDACTED] - for Project IETI

1y Like Reply

 [REDACTED]

cc Yihuan [REDACTED] Don [REDACTED]

1y Like Reply

Ian [redacted]
For anyone interested in our latest findings and plans on this workstream, you can check out our recent notes here...

Latest go-forward plans
<https://fb.workplace.com/.../1641.../posts/1605980229872290/>

What more we learned in 2022:
<https://fb.workplace.com/.../1641.../posts/1605861946550785/>

 OLD Facebook Integrity: FYI OLD

1y Like Reply

Matan [redacted]
cc **Arnon** [redacted] FYI

50w Like Reply

Matan [redacted]
Tali [redacted]

50w Like Reply

Write a reply...

Sergio [redacted]
Sean [redacted]

43w Like Reply

Sam [redacted]
Monica [redacted]

32w Like Reply

Monica [redacted]
Robert [redacted] **Eric** [redacted] **Katie** [redacted] **James** [redacted]

32w Like Reply

Eric [redacted]
Oh wow... This is an extremely important note/finding! Great presentation of the results and clearly written. Fantastic work with this! Thanks for sharing **Monica** and **Sam**!

32w Like Reply

Jessica [redacted]
cc **Jack** [redacted]!

1w Like Reply