Tennessee v. Meta
Case No. 23-1364-IV
PX-00030

"Addiction"/Problematic Use Design Audit Plan [WIP]

POCs: Elena Davis & Jennifer Guadagno
[+ Crystine [redacted] depending on degree to which we want to combine with dark patterns work]

Design Partners: TBD (ideally someone from Jason [redacted]'s Z1 team, possibly someone from FB Design System on Product Foundations?)

Outstanding questions

- Is "addiction"/ problematic use audit the same as dark-pattern work? Should they be combined or kept separate, with addiction/problematic use audit informing work happening on dark patterns and vice versa? - Elena/Crystine/Jennifer to figure out
- How closely to work with Margaret's team on this? Where are they at, are goals aligned?
- Include Instagram as well?
- Who should we partner with externally? Tristan or others? (see pros/cons below)

Context

There is growing concern in the public that technology companies are deliberately manipulating users by designing addictive products that can harm well-being and subvert self-control. The prevalence of problematic (general) internet use ranges between 1.5-8.2% in the US and Europe (link), and academics are debating whether 1. FB is truly "addictive", and 2. what standardized definitions we should use to identify FB addiction or problematic FB use. Regardless of the clinical definitions here - given that currently there is NO medically defined FB addiction disorder - what are the real people problems in this space? The well-being team has reframed the addiction narrative to focus on the ways that FB may be contributing to use patterns that people find difficult to control despite negative impacts in their lives, and to identify and correct those contributing factors. Specifically, what are the ways that FB can seem rewarding or reinforcing without delivering real value, and what are the types of habitual FB behaviors that people find unwanted and difficult to change?

Background for the audit

- Research on reward learning and habit formation processes point to certain behavior patterns that promote repeated, automatic behaviors that may begin to operate outside of an individual's control.
- We hypothesize that unpredictable, low value rewards and cues that promote frequent, automatic, undesired behaviors are contributing to reward learning and habit formation in a way that is not aligned with users' intentions or preferences.
- These may lead to feelings of being manipulated, feeling lack of control related to certain behaviors, and feelings of dependence on checking or going on Facebook that could be related to lower well-being - and could be fueling the subjective, colloquial experiences of feeling "addicted" to Facebook.
- See more about "addiction" narrative here
- In keeping with the engagement model of the well-being team, we have defined guidelines for how teams throughout the company can be incorporating well-being when designing and building products. Some of these guidelines are particularly relevant for identifying and reconsidering aspects of the platform that are currently promoting "addictive"-like behaviors. (See guidelines in Related work section below).

Research Plan

ALIGN WITH INTERNAL TEAMS

Work with Margaret Stewart and other design ethics leaders to understand what design changes would be in line with the goals of reducing low value rewards and improving user control. Work with Crystine [redacted] and team to understand how this work fits into and can benefit dark pattern work. See if this should be expanded to Instagram as well.

CREATE CLEAR DEFINITIONS/CRITERIA

Partner with design to refine a definition of what aspects of features promote "addictive"-like/problematic behaviors to target in the audit and how they're different from/related to dark patterns. Specifically, features are problematic if they promote:

- Unpredictable, low-value rewards (e.g., certain notifications)
- Low-control or automatic behaviors (e.g., autoplay)
- Frequent behaviors that are pressured or unwanted (e.g., read receipts)
- Define other criteria

DESIGN AUDIT

Perform a design audit of the platform with key external behavioral experts and with internal designers and researchers, to identify features that may lead to habitual and unwanted / low-value behaviors

How we'll work with external expert:

- Share and iterate on our definition of characteristics of features that promote "addictive"-like/problematic behaviors. Establish how that's different from/related to dark patterns and specify whether we're after just former or both (TBD from our internal discussions).
- Engage with a third party to conduct an independent audit and / or review of the features we have identified
- Form an alliance where the third party can vouch for the thoroughness and relevance of our approach for targeting the "addiction" claims

Possible external experts to partner with:

- Tristan Harris (and others in his camp: eg, Justin Rosenstein)
    - Lots of pros and cons here - all included at very end of doc, but summary is biggest risk is that he points out things that we don't agree with or are unfounded and we don't make changes he's suggesting
- IDEO
- MIT Media Lab
- Academics specializing in addiction and technology (eg, UK researchers: Daria Kuss and Mark Griffiths - informational meeting set up in mid-September). Others?
- ~~Joe Edelman - value based design (part of Tristan's camp)~~

EXPLORE/TEST ISSUES RAISED IN AUDIT

Work with relevant product teams to follow up audit with user research, tests, lit reviews/gathering scientific support on certain issues raised in audit.

- Eg., test potential drivers of habitual, repeated behaviors by creating variants on the platform and examining associated behavior changes (if badge color being red is raised as an issue test effects of changing notification badge color)

GOALS

- Prioritized, validated list of concerning product experiences that should be fixed or improved
- Take prioritized list into a design sprint to mock up fixes or alternative features to address main problems
- Consider expanding or adding to well-being principles to incorporate new "anti-addiction"/ problematic use principles
    - E.g., "Behaviors that are frequent should also be valuable."
- Inform and help move along work on dark patterns (Crystine [redacted] team and Margaret Stewart's team)

Related work within the company

- Well-being engagement model and design principles (link, selected principles copied below): Devina [redacted] and Jennifer Guadagno
    - Give people the most value in the time they have
        - Do: recognize that people have different intentions when coming to Facebook
        - Don't: design to increase growth metrics without also benefiting the individual
    - Provide the transparency needed for people to make decisions with intentionality and clarity
        - Do: make it clear what will happen as a result of their actions
        - Do: provide defaults that are good for people's well-being, while empowering them to control their Facebook experience
        - Don't: impose behaviors on people that are against their best interests
    - Create an environment that has a positive impact on people, and that is easy to participate in
        - Don't: create features designed for good that can come off as self-serving
- Dark Pattern spotting (link): Crystine [redacted]
- Margaret's Ethical Design work

Additional Thoughts & Info

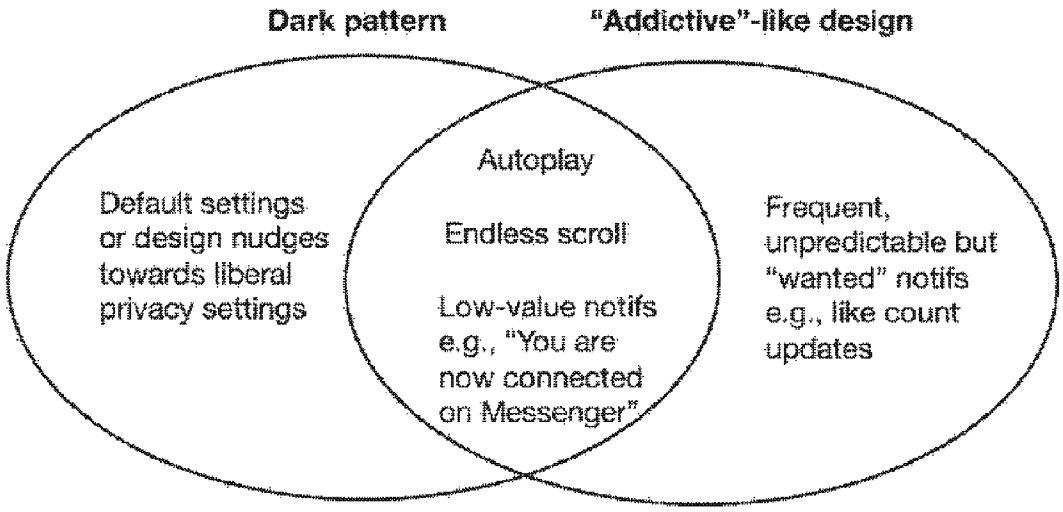
Q: How is an "addictive"-like design element similar vs. different from a "dark pattern"?

A: "Dark pattern" is a UX term referring to interactions that are deceptive, or that trick you into doing something you didn't want to do.

These include misleading or manipulative features or settings that are designed to encourage a particular behavior that is not necessarily in the user's best interests. These can include single-instance behaviors.
"Addictive"-like design elements are those that promote repeated or poorly regulated behaviors that may have a negative impact on the individual over time as they continue behaviors that aren't in their best interests.

A dark pattern deviates from user expectations; an "addictive"-like design results in a behavior pattern that is not wanted or valuable for the user over time (the function of the design is to promote repeated maladaptive or low-control behaviors).

Examples of cases that may be both "dark patterns" and "addictive"-like: All Roads Lead to Watch (since it seems like it could promote automatic or unintentional behaviors when coupled with autoplay), and low value or "fake" notifications like alerts that you are connected on Messenger or have a particular setting turned off (promoting frequent checking behaviors and perceptions of reward without delivering gratification).



Previous Notes:

Internal at facebook

- Work with Margaret Stewart and other design ethics leaders to understand what design changes would be in line with the goals of reducing low value rewards and improving user control
- Refine a definition of what aspects of features promote "addictive"-like behaviors to target in the audit, specifically:
    - Which features constitute unpredictable, low-value rewards
        - e.g., certain notifications
    - Which features or settings promote low-control or automatic behaviors
        - e.g., autoplay
    - Which features encourage frequent behaviors that are pressured or unwanted
        - e.g., read receipts
- Perform a design audit of the platform with key external behavioral experts, with internal designers and researchers, to identify features that may lead to habitual and unwanted / low-value behaviors
- Use our refined definition to identify and understand the function of those features in creating "addictive"-like behaviors (i.e., a habit loop that is having negative impacts on the person)
- Follow up with audit with user research on certain features to verify the hypothesized functions (?)
    - e.g., test potential drivers of habitual, repeated behaviors by creating variants on the platform and examining associated behavior changes (like changing notification badge color)
- Iterate on the audit following qualitative user research in September-October and quantitative research in (? October) if new aspects of the platform are identified as problematic
- Consider expanding or adding to well-being principles to incorporate new "anti-addiction" principles
    - E.g., "Behaviors that are frequent should also be valuable."

with an external third party

- Share and iterate on our definition of characteristics of features that promote "addictive"-like behaviors
- Engage with a third party to conduct an independent audit and / or review of the features we have identified
- Form an alliance where the third party can vouch for the thoroughness and relevance of our approach for targeting the "addiction" claims

Partnering with Tristan Harris on this - Pros & Cons

- Pros:
    - He has lots of thoughts in this space - he said it's the area he's best positioned to inform, it'd be helpful to learn from him.
    - Gives him a chance to be heard
    - Helpful to deeply partner with him on area he's most informed about to further our relationship
    - He's likely connected to other experts in this area we could leverage and learn from
- Possible concerns:
    - NDA?
    - What if we don't agree on much or his claims aren't substantiated by other further research (eg, color red on badges)
    - We, from the research side, do agree with his advice but product teams are slow or unwilling to change things
    - He has even more inside info about what's bad with FB
    - Ideally working with an external expert could on this could make them see we're taking this seriously and they become vocal advocates of ours we can point people to. Likely won't happen with him

- Things we can do to help mitigate concerns:
  - Setting expectations: Be very explicit about us wanting to carefully vet/further research anything suggested so he knows we won't necessarily take whatever he says at face value and run with it.
  - Identify at the outset that getting product teams to make big changes may be tricky, use this as a challenge for him - how can he help us convince product teams?
  - Work with additional design firms and experts in addition to him. Have him co-create an expert panel with us (we can both nominate and veto people). This ensures more balanced thinking, other people that aren't us to possibly disagree with him, and builds relationships with other allies to advocate for us.
- Overall this is really risky - high risk but high reward. Are the potential gains to our relationship with him and broader reputation worth the potential harm this can do/are we already in a bad enough spot with him that possible gains outweigh harms?

Contact Support

Conversation History

Quipbot#08 left the thread Oct 12 at 10:25 am

Quipbot#08 Oct 12 at 10:20 am

=====IMPORTANT=====

This Quip file has been copied to Google Workspace (https://docs.google.com/document...) as part of the Quip2Google migration efforts. Quip is moving to Read-Only. Please work in the new Google file moving forward.

Learn more

https://fb.workplace.com/groups...

Questions?

Visit the Quip2Google Migration Feedback Group (https://fb.workplace.com/groups...).

Note

New edits in this Quip file will NOT be automatically synced to Google Workspace.

Quipbot#08 added Quipbot#08 to the thread Oct 12 at 10:25 am

Quipbot#08 added you to a document Oct 12 at 10:20 am

Jonny[redacted] opened your conversation with Elena Davis, Gretchen [redacted], Crystine [redacted] and Jennifer Guadagno Aug 20, 2018 at 6:47 pm

---

Gretchen [redacted] added Jonny [redacted] to the thread Aug 20, 2018 at 2:18 pm

---

Gretchen [redacted] added you to a document Aug 20, 2018 at 2:12 pm

---

Crystine [redacted] Aug 17, 2018 at 12:17 am

Here is the WIP doc for the dark pattern work that defines some of the sub-categories: Dark Pattern spotting

---

Crystine [redacted] Aug 17, 2018 at 12:14 am

I've been working with Shali [redacted] on dark patterns, there is an option where make dark patterns (or some other term) the parent category that includes subtypes

---

Crystine [redacted] opened your conversation with Elena Davis, Gretchen [redacted], Jennifer Guadagno and Jonny [redacted] Aug 17, 2018 at 12:13 am

---

Crystine [redacted] edited Aug 17, 2018 at 12:28 am

Jennifer Guadagno Aug 16, 2018 at 10:02 pm

We just went through Joe's "Designing for Values" workshop and it wasn't good

---

Jennifer Guadagno edited Aug 16, 2018 at 10:17 pm

- Tristan Harris (and others in his camp: eg, Justin Rosenstein)
  - Lots of pros and cons here - all included at very end of doc, but summary is biggest risk is that he points out things that we don't agree with or are unfounded and we don't make changes he's suggesting
- IDEO
- MIT Media Lab
- Academics specializing in addiction and technology (eg, UK researchers: Daria Kuss and Mark Griffiths - informational meeting set up in mid-September). Others?
- ~~Joe Edelman - value based design (part of Tristan's camp)~~

Partnering with Tristan Harris on this - Pros & Cons

- Pros:
  - He has lots of thoughts in this space - he said it's the area he's best positioned to inform, it'd be helpful to learn from him.
  - Gives him a chance to be heard
  - Helpful to deeply partner with him on area he's most informed about to further our relationship
  - He's likely connected to other experts in this area we could leverage and learn from
- Possible concerns:
  - NDA?
  - What if we don't agree on much or his claims aren't substantiated by other further research (eg, color red on badges)
  - We, from the research side, do agree with his advice but product teams are slow or unwilling to change things
  - He has even more inside info about what's bad with FB
  - Ideally working with an external expert could on this could make them see we're taking this seriously and they become vocal advocates of ours we can point people to. Likely won't happen with him

- Things we can do to help mitigate concerns:
  - Setting expectations: Be very explicit about us wanting to carefully vet/further research anything suggested so he knows we won't necessarily take whatever he says at face value and run with it.
  - Identify at the outset that getting product teams to make big changes may be tricky, use this as a challenge for him - how can he help us convince product teams?
  - Work with additional design firms and experts in addition to him. Have him co-create an expert panel with us (we can both nominate and veto people). This ensures more balanced thinking, other people that aren't us to possibly disagree with him, and builds relationships with other allies to advocate for us.
- Overall this is really risky - high risk but high reward. Are the potential gains to our relationship with him and broader reputation worth the potential harm this can do/are we already in a bad enough spot with him that possible gains outweigh harms?

Elena Davis Aug 16, 2018 at 6:04 pm

So interesting - looks like he's part of the Tristan coalition too. Have we spoken to him previously about this space?

Elena Davis Aug 16, 2018 at 6:04 pm

So interesting - looks like he's part of the Tristan coalition too. Have we spoken to him previously about this space?

Gretchen [redacted] Aug 16, 2018 at 4:36 pm

Justin Rosenstein also has strong opinions on this. he's part of the alumni network.

Gretcher[redacted] Aug 16, 2018 at 4:36 pm

Justin Rosenstein also has strong opinions on this. he's part of the alumni network.

Gretcher[redacted] opened your conversation with Elena Davis, Crystine[redacted] Jennifer Guadagno and Jonny[redacted] Aug 16, 2018 at 4:35 pm

---

Gretchen [redacted] edited Aug 16, 2018 at 4:35 pm

Jennifer Guadagno Aug 10, 2018 at 11:31 pm

I see this as being a part of the definition/criteria section above? I think it's captured there but thinking if it needs to be added elsewhere?

Jennifer Guadagno Aug 10, 2018 at 11:31 pm

I see this as being a part of the definition/criteria section above? I think it's captured there but thinking if it needs to be added elsewhere?

Jennifer Guadagno Aug 10, 2018 at 11:29 pm

oh yea, the sentence structure doesn't make sense. Internal designers and researchers will work alongside external experts, is how i'm seeing it

---

Jennifer Guadagno edited Aug 11, 2018 at 12:17 am
Leads POCs: Elena Davis, & Jennifer Guadagno,
[+ Crystine [redacted] depending on degree to which we want to combine with dark patterns work]

- How closely to work with Margaret's team on this? Where are they at, are goals aligned?
- Do for Instagram as well?
- Include Instagram as well?
- Who should we partner with externally? Tristan or others? (see pros/cons below)
- See more about "addiction" narrative here: https://fb.facebook.com/groups/ResearchInsights/permalink/2042922635756341/

Refine Partner with design to refine a definition of what aspects of features promote "addictive"-like/problematic behaviors to target in the audit and how they're different from/related to dark patterns. Specifically, features are problematic if they promote:
Perform a design audit of the platform with key external behavioral experts, and with internal designers and researchers, to identify features that may lead to habitual and unwanted / low-value behaviors

- Prioritized, validated list of concerning product experiences that should be fixed or improved
- Take prioritized list into a design sprint to mock up fixes or alternative features to address main problems

---

Elena Davis Aug 10, 2018 at 11:10 pm

also shouldn't the sprint be specified in the plan? I imagine we won't complete the audit during a sprint - or did you think we would?

Elena Davis Aug 10, 2018 at 11:07 pm

(maybe it fits in with the "setting expectations"?)

Elena Davis Aug 10, 2018 at 11:07 pm

(maybe it fits in with the "setting expectations"?)

Elena Davis Aug 10, 2018 at 11:06 pm

Did you incorporate this content somewhere already? I don't think I see it yet...

Elena Davis Aug 10, 2018 at 11:06 pm

Did you incorporate this content somewhere already? I don't think I see it yet...

Elena Davis Aug 10, 2018 at 11:05 pm

that specifies design elements as well as behavioral impacts?

Elena Davis Aug 10, 2018 at 11:04 pm

If we are doing the definitions ourselves based on research and not design input, this would be a first step before the "align with internal teams." But because it comes second, that made me think that we would involve the design teams in this process. And I thought that would be as part of the upcoming sprint?

Elena Davis Aug 10, 2018 at 11:02 pm

specify "Who should we partner with externally?"

Elena Davis edited Aug 10, 2018 at 11:07 pm

- See more about "addiction" narrative here: https://fb.facebook.com/groups/ResearchInsights/permalink/2042922635756341/
- Share and iterate on our definition of characteristics of features that promote "addictive"-like/problematic behaviors. Establish how that's different from/related to dark patterns and specify whether we're after just former or both (tbd on (TBD from our internal discussions).
- Engage with a third party to conduct an independent audit and/or and / or review of the features we have identified
- Academics specializing in addiction and technology (eg, UK researchers: Daria [redacted] and Mark [redacted]) - informational meeting set up in mid-September). Others?

Elena Davis Aug 10, 2018 at 11:01 pm

I just didn't understand the sentence structure. Should be "Perform a design audit with external experts AND internal designers"?

Elena Davis Aug 10, 2018 at 11:00 pm

Sorry to have made this more confusing. I thought that the sprint was to define this space including in design terms in order to launch the audit. Were you thinking we would firm up definitions BEFORE syncing with the design teams?

Jennifer Guadagno Aug 10, 2018 at 10:07 pm

Hmm I saw seeing this as a partnership and for us all to be working together, at the same time on this. What did you have in mind

---

Jennifer Guadagno Aug 10, 2018 at 10:06 pm

Is this how you were envisioning this section or as something else?

Jennifer Guadagno Aug 10, 2018 at 10:05 pm

You have some of these already listed below, and can likely be expanded or added to. We should have our own POV and suggestions going into this, and of course have it as a point of discussion for whoever we partner with as well

---

Jennifer Guadagno Aug 10, 2018 at 10:04 pm

These definitions are then the criteria around which the audit is performed (eg, identify features that promote/convey x, y, z)

Jennifer Guadagno Aug 10, 2018 at 10:04 pm

Hmm I'm not sure what you mean here - part of sprint or part of audit? I see this as happening first. Us defining the criteria of problematic use, that's different from dark patterns, etc.

Jennifer Guadagno Aug 10, 2018 at 10:01 pm

Good one, added him below

Jennifer Guadagno Aug 10, 2018 at 10:01 pm

good point, incorporated

Jennifer Guadagno Aug 10, 2018 at 10:01 pm

good point, incorporated

Jennifer Guadagno edited Aug 10, 2018 at 10:07 pm
Design Partners: TBD (possibly (ideally someone from Jason [redacted] Z1 team, possibly someone from FB Design System on Product Foundations) Foundations?)

- How closely to work with Margaret's team on this? Where are they at, are goals aligned?
- Do for Instagram as well?
- Who should we partner with? Tristan or others? (see pros/cons below)

- These may lead to feelings of being manipulated, feeling lack of control related to certain behaviors, and feelings of dependence on checking or... on Facebook that could be related to lower well-being - and could be fueling the subjective, colloquial experiences of feeling "addicted" to Facebook.
- See more about "addiction" narrative here:

ALIGN WITH INTERNAL TEAMS
Work with Margaret Stewart and other design ethics leaders to understand what design changes would be in line with the... rewards and improving user control. Work with Crystine [redacted] and team to understand how this work fits into and can benefit dark pattern work. See if this should be expanded to Instagram as well.
CREATE CLEAR DEFINITIONS DEFINITIONS/CRITERIA
Refine a definition of what aspects of features promote "addictive"-like/problematic behaviors to target in the audit and how they're different from/related to dark patterns. Specifically, features are problematic if they promote:

- Which features constitute unpredictable Unpredictable, low-value rewards (e.g., certain notifications)

- Which features or settings promote low-control Low-control or automatic behaviors (e.g., autoplay)
- Which features encourage frequent Frequent behaviors that are pressured or unwanted (e.g., read receipts)
- Define other criteria
    - He has even further more inside info about what's bad with FB
        - Overall this is really risky - high risk but big rewards high reward. Are the potential gains to our relationship with him and broader reputation worth the potential harm this can do/are we already in a bad enough spot with him that possible gains outweigh harms?
- IDEO
- Academics specializing in addiction and technology (eg, UK researchers: Daria [redacted] and Mark [redacted]). Others?
- Joe [redacted] - value based design (part of Tristan's camp)

Elena Davis Aug 8, 2018 at 10:48 pm

Part of sprint?

Elena Davis Aug 8, 2018 at 10:46 pm

I feel like it's fair to acknowledge somewhere that not all habits/automaticity are bad. A lot are just efficient. Simply because something is "addictive"-like doesn't mean it's a) actually having that effect on people, and b) is unwanted by the majority of users. It seems like we can identify top offenders and work to make them less bad, if not remove them altogether.

Elena Davis Aug 8, 2018 at 10:46 pm

I feel like it's fair to acknowledge somewhere that not all habits/automaticity are bad. A lot are just efficient. Simply because something is "addictive"-like doesn't mean it's a) actually having that effect on people, and b) is unwanted by the majority of users. It seems like we can identify top offenders and work to make them less bad, if not remove them altogether.

Elena Davis Aug 8, 2018 at 10:43 pm

We'd need to do ours first if our plan to engage externally includes sharing our work-in-progress

---

Elena Davis Aug 8, 2018 at 10:43 pm

is this an "and"? Do you imagine them sequentially? I thought we'd do internal first...?

Elena Davis Aug 8, 2018 at 10:33 pm

further - potential gains to our broader reputation if we are seen as making these changes?

Elena Davis Aug 8, 2018 at 10:33 pm

further - potential gains to our broader reputation if we are seen as making these changes?

Elena Davis Aug 8, 2018 at 10:33 pm

Joe [redacted] here?

Elena Davis edited Aug 8, 2018 at 10:48 pm
There is growing concern in the public that technology companies are deliberately manipulating users by designing addictive products that can harm well-being and subvert self-control... seem rewarding or reinforcing without delivering real value, and what are the types of habitual FB behaviors that people find unwanted and difficult to change. ?
Perform a design audit of the platform with key external behavioral experts, with internal designers and researchers, to identify features that may lead to habitual and unwanted / low-value behaviors
Possible external experts to partner with:

- We, from the research side, do agree with his advice but product teams are slow or unwilling to change things
- Ideally working with an external expert could on this could make them see we're taking this seriously and they become vocal advocates of ours we can point people to. Likely won't happen with him

- Overall this is really risky - high risk but big rewards. Are the potential gains to our relationship with him worth the potential harm this can do/are we already in a bad enough spot with him that possible gains outweigh harms?

- Prioritized, validated list of concerning product experiences that should be fixed or improved
- Well-being engagement model and design principles (link) , selected principles copied below): Davina Baum and Jennifer Guadagno

Image

Jennifer Guadagno edited Aug 8, 2018 at 10:11 pm

"Addictive"-Like "Addiction"/Problematic Use Design Audit Plan [WIP]

Leads: Elena Davis, Jennifer Guadagno, Crystine [redacted]

Design Partners: TBD (possibly someone from Jason [redacted] Z1 team, possibly someone from FB Design System on Product Foundations)

[WIP]
Outstanding questions

- Is "addiction"/ problematic use audit the same as dark-pattern work? Should they be combined or kept separate, with addiction/problematic use audit informing work happening on dark patterns and vice versa? - Elena/Crystine/Jennifer to figure out
- How closely to work with Margaret's team on this? Where are they at, are goals aligned?

- Research on reward learning and habit formation processes point to certain behavior patterns that promote repeated, automatic behaviors that may begin to operate outside of an individual's control.
- We hypothesize that unpredictable, low value rewards and cues that promote frequent, automatic, undesired behaviors are contributing to reward learning and habit formation in a way that is not aligned with users' intentions or preferences.

- These may lead to feelings of being manipulated, feeling lack of control related to certain behaviors, and feelings of dependence on checking or going on Facebook that could be related to lower well-being - and could be fueling the subjective, colloquial experiences of feeling "addicted" to Facebook.
- In keeping with the engagement model of the well-being team, we have defined guidelines for how teams throughout the company can be incorporating well-being when designing and building products. Some of these guidelines are particularly relevant for identifying and reconsidering aspects of the platform that are currently promoting "addictive"-like behaviors. (See guidelines in Related work section below).

relevant Well-being design principles

In keeping with the engagement model of the well-being team, we have defined guidelines for how teams throughout the company can be incorporating well-being when designing and building products. Some of these guidelines are particularly relevant for identifying and reconsidering aspects of the platform that are currently promoting "addictive"-like behaviors.

- Give people the most value in the time they have
  - Do: recognize that people have different intentions when coming to Facebook
  - Don't: design to increase growth metrics without also benefiting the individual
- Provide the transparency needed for people to make decisions with intentionality and clarity
  - Do: make it clear what will happen as a result of their actions
  - Do: provide defaults that are good for people's well-being, while empowering them to control their Facebook experience
  - Don't: impose behaviors on people that are against their best interests
- Create an environment that has a positive impact on people, and that is easy to participate in

Research Plan and Goals

ALIGN WITH INTERNAL TEAMS

Work with Margaret Stewart and other design ethics leaders to understand what design changes would be in line with the goals of reducing low value rewards and improving user control. Work with Crystine[redacted] and team to understand how this work fits into and can benefit dark pattern work.

CREATE CLEAR DEFINITIONS

Refine a definition of what aspects of features promote "addictive"-like/problematic behaviors to target in the audit and how they're different from/related to dark patterns. Specifically:

- Which features constitute unpredictable, low-value rewards (e.g., certain notifications)
- Which features or settings promote low-control or automatic behaviors (e.g., autoplay)
- Which features encourage frequent behaviors that are pressured or unwanted (e.g., read receipts)

DESIGN AUDIT

Perform a design audit of the platform with key external behavioral experts, with internal designers and researchers, to identify features that may lead to habitual and unwanted / low-value behaviors

How we'll work with external expert:

- Share and iterate on our definition of characteristics of features that promote "addictive"-like/problematic behaviors. Establish how that's different from/related to dark patterns and specify whether we're after just former or both (tbd on our internal discussions).
- Engage with a third party to conduct an independent audit and/or review of the features we have identified
- Form an alliance where the third party can vouch for the thoroughness and relevance of our approach for targeting the "addiction" claims

Possible external experts to partner with:

- Tristan[redacted]
  - Pros:
    - He has lots of thoughts in this space - he said it's the area he's best positioned to inform, it'd be helpful to learn from him.
    - Gives him a chance to be heard

        - Helpful to deeply partner with him on area he's most informed about to further our relationship
        - He's likely connected to other experts in this area we could leverage and learn from
    - Possible concerns:
        - NDA?
        - What if we don't agree on much or his claims aren't substantiated by other further research (eg, color red on badges)
        - We, from the research side, do agree with his advice but product teams are slow or unwilling to change things
        - He has even further inside info about what's bad with FB
        - Ideally working with an external expert could on this could make them see where taking this seriously and they become vocal advocates of ours we can point people to. Likely won't happen with him
    - Things we can do to help mitigate concerns:
        - Setting expectations: Be very explicit about us wanting to carefully vet/further research anything suggested so he knows we won't necessarily take whatever he says at face value and run with it.
        - Identify at the outset that getting product teams to make big changes may be tricky, use this as a challenge for him - how can he help us convince product teams?
        - Work with additional design firms and experts in addition to him. Have him co-create an expert panel with us (we can both nominate and veto people). This ensures more balanced thinking, other people that aren't us to possibly disagree with him, and builds relationships with other allies to advocate for us.
    - Overall this is really risky - high risk but big rewards. Are the potential gains to our relationship with him worth the potential harm this can do/are we already in a bad enough spot with him that possible gains outweigh harms?
- IDEO
- Academics specializing in addiction and technology (eg, UK researchers: Daria Kuss and Mark Griffiths). Others?

EXPLORE/TEST ISSUES RAISED IN AUDIT

Work with relevant product teams to follow up audit with user research, tests, lit reviews/gathering scientific support on certain issues raised in audit.

- Eg., test potential drivers of habitual, repeated behaviors by creating variants on the platform and examining associated behavior changes (if badge color being red is raised as an issue test effects of changing notification badge color)

GOALS

- Prioritized, validated list of concerning product experiences that should be fixed or improved
- Consider expanding or adding to well-being principles to incorporate new "anti-addiction"/ problematic use principles
    - E.g., "Behaviors that are frequent should also be valuable."
- Inform and help move along work on dark patterns (Crystine [redacted] team and Margaret Stewart's team)

- Well-being engagement model and design principles (link): Davina [redacted] and Jennifer Guadagno
    - Give people the most value in the time they have
        - Do: recognize that people have different intentions when coming to Facebook
        - Don't: design to increase growth metrics without also benefiting the individual

- Provide the transparency needed for people to make decisions with intentionality and clarity
  - Do: make it clear what will happen as a result of their actions
  - Do: provide defaults that are good for people's well-being, while empowering them to control their Facebook experience
  - Don't: impose behaviors on people that are against their best interests
- Create an environment that has a positive impact on people, and that is easy to participate in
  - Don't: create features designed for good that can come off as self-serving

* Dark Pattern spotting (link): Crystine [redacted]
* Margaret's Ethical Design work

Additional Thoughts & Info

Q: How is an "addictive"-like design element similar vs. different from a "dark pattern"?

A: "Dark pattern" is a UX term referring to interactions that are deceptive, or that trick you into doing something you didn't want to do.

These include misleading or manipulative features or settings that are designed to encourage a particular behavior that is not necessarily in the user's best interests. These can include single-instance behaviors.
"Addictive"-like design elements are those that promote repeated or poorly regulated behaviors that may have a negative impact on the individual over time as they continue behaviors that aren't in their best interests.

A dark pattern deviates from user expectations; an "addictive"-like design results in a behavior pattern that is not wanted or valuable for the user over time (the function of the design is to promote repeated maladaptive or low-control behaviors).

> Q: How is an "addictive"-like design element similar vs. different from a "dark pattern"?
>
> A: "Dark pattern" is a UX term referring to interactions that are deceptive, or that trick you into doing something you didn't want to do. These include misleading or manipulative features or settings that are designed to encourage a particular behavior that is not necessarily in the user's best interests. These can include single-instance behaviors.
>
> "Addictive"-like design elements are those that promote repeated or poorly regulated behaviors that may have a negative impact on the individual over time as they continue behaviors that aren't in their best interests.
>
> A dark pattern deviates from user expectations; an "addictive"-like design results in a behavior pattern that is not wanted or valuable for the user over time (the function of the design is to promote repeated maladaptive or low-control behaviors).

Previous Notes:

- Work with Margaret Stewart and other design ethics leaders to understand what design changes would be in line with the goals of reducing low value...
  - Which features encourage frequent behaviors that are pressured or unwanted
    - e.g., read receipts
- Perform a design audit of the platform with key external behavioral experts, with internal designers
- accompanied by researcher input
- and researchers, to identify features that may lead to habitual and unwanted / low-value behaviors
- Use our refined definition to identify and understand the function...

- adding to well-being principles to incorporate new "anti-addiction" principles
    - E.g., "Behaviors that are frequent should also be valuable."

Jennifer Guadagno renamed "Addictive"-Like Design Audit Plan to "Addiction"/Problematic Use Design Audit Plan [WIP] Aug 8, 2018 at 9:01 pm

---

Jennifer Guadagno opened your conversation with Elena Davis, Gretchen [redacted] Crystine [redacted] and Jonny Thaw Aug 8, 2018 at 7:29 pm

---

Elena Davis enabled link sharing - View/Edit Aug 8, 2018 at 6:11 pm

---

Elena Davis added Jennifer Guadagno to the thread Aug 8, 2018 at 6:11 pm

---

Elena Davis added you to a document Aug 8, 2018 at 6:04 pm

Elena Davis edited Aug 8, 2018 at 6:06 pm

"Addiction" Feature "Addictive"-Like Design Audit Plan

[WIP]

Context

:

There is growing concern in the public that technology companies are deliberately manipulating users by designing addictive products that can harm well-being and subvert self-control. The prevalence of problematic (general) internet use ranges between 1.5-8.2% in the US and Europe (link), and academics are debating whether 1. FB is truly "addictive", and 2. what standardized definitions we should use to identify FB addiction or problematic FB use. Regardless of the clinical definitions here - given that currently there is NO medically defined FB addiction disorder - what are the real people problems in this space? The well-being team has reframed the addiction narrative to focus on the ways that FB may be contributing to use patterns that people find difficult to control despite negative impacts in their lives, and to identify and correct those contributing factors. Specifically, what are the ways that FB can seem rewarding or reinforcing without delivering real value, and what are the types of habitual FB behaviors that people find unwanted and difficult to change.

Background for the audit

Research on reward learning and habit formation processes point to certain behavior patterns that promote repeated, automatic behaviors that may begin to operate outside... feeling lack of control related to certain behaviors, and feelings of dependence on checking or going on Facebook that could be related to lower well-being - and could be fueling the subjective, colloquial experiences of feeling "addicted" to Facebook.

relevant Well-being design principles

In keeping with the engagement model of the well-being team, we have defined guidelines for how teams throughout the company can be incorporating well-being when designing and building products. Some of these guidelines are particularly relevant for identifying and reconsidering aspects of the platform that are currently promoting "addictive"-like behaviors.

- Give people the most value in the time they have
    - Do: recognize that people have different intentions when coming to Facebook
    - Don't: design to increase growth metrics without also benefiting the individual

- Provide the transparency needed for people to make decisions with intentionality and clarity
    - Do: make it clear what will happen as a result of their actions
    - Do: provide defaults that are good for people's well-being, while empowering them to control their Facebook experience
    - Don't: impose behaviors on people that are against their best interests
- Create an environment that has a positive impact on people, and that is easy to participate in
    - Don't: create features designed for good that can come off as self-serving

Research Plan and Goals

Internal at facebook

- Work with Margaret Stewart and other design ethics leaders to understand what design changes would be in line with the goals of reducing low value rewards and improving user control
- Refine a definition of what aspects of features promote "addictive"-like behaviors to target in the audit, specifically:
    - Which features constitute unpredictable, low-value rewards
        - e.g., certain notifications
    - Which features or settings promote low-control or automatic behaviors
        - e.g., autoplay
    - Which features encourage frequent behaviors that are pressured or unwanted
        - e.g., read receipts
- Perform a design audit of the platform, with designers accompanied by researcher input, to identify features that may lead to habitual and unwanted / low-value behaviors
    - Use our refined definition to identify and understand the function of those features in creating "addictive"-like behaviors (i.e., a habit loop that is having negative impacts on the person)
- Follow up with audit with user research on certain features to verify the hypothesized functions (?)
    - e.g., test potential drivers of habitual, repeated behaviors by creating variants on the platform and examining associated behavior changes (like changing notification badge color)
- Iterate on the audit following qualitative user research in September-October and quantitative research in (? October) if new aspects of the platform are identified as problematic
- Consider expanding or adding to well-being principles to incorporate new "anti-addiction" principles
    - E.g., "Behaviors that are frequent should also be valuable."

with an external third party

- Share and iterate on our definition of characteristics of features that promote "addictive"-like behaviors
- Engage with a third party to conduct an independent audit and / or review of the features we have identified
- Form an alliance where the third party can vouch for the thoroughness and relevance of our approach for targeting the "addiction" claims

Related work within the company

Well-being engagement model and design principles (link)

- (?With a third party?) Review features and default settings on the platform to identify unnecessary examples of frequent low-value reward signals, automatic unwanted behaviors that contribute to low perceptions of control, and problematic domains that come from user interviews and/or the problematic use quantitative analysis

What Facebook features are implicated in:
Frequent, pressured or unwanted behaviors (e.g., read receipts)
Low-control or automatic behaviors (e.g., autoplay)
Unpredictable, low-value rewards (e.g., certain notifications)

Principles:
Things that are frequent should be wanted and valuable

> Q: How is an "addictive"-like design element similar vs. different from a "dark pattern"?

> A: "Dark pattern" is a UX term referring to interactions that are deceptive, or that trick you into doing something you didn't want to do. These include misleading or manipulative features or settings that are designed to encourage a particular behavior that is not necessarily in the user's best interests. These can include single-instance behaviors.

> "Addictive"-like design elements are those that promote repeated or poorly regulated behaviors that may have a negative impact on the individual over time as they continue behaviors that aren't in their best interests.

> A dark pattern deviates from user expectations; an "addictive"-like design results in a behavior pattern that is not wanted or valuable for the user over time (the function of the design is to promote repeated maladaptive or low-control behaviors).

Examples of cases that may be both "dark patterns" and "addictive"-like: All Roads Lead to Watch (since it seems like it could promote automatic or unintentional behaviors when coupled with autoplay), and low value or "fake" notifications like alerts that you are connected on Messenger or have a particular setting turned off (promoting frequent checking behaviors and perceptions of reward without delivering gratification).

Elena Davis edited Aug 7, 2018 at 11:39 pm

What Facebook features are implicated in:
Frequent, pressured or unwanted behaviors (e.g., read receipts)
Low-control or automatic behaviors (e.g., autoplay)
Unpredictable, low-value rewards (e.g., certain notifications)

Principles:
Things that are frequent should be wanted and valuable

Elena Davis edited Aug 2, 2018 at 6:31 pm

"Addiction" Feature Audit

Context:
Research on reward learning and habit formation processes point to certain behavior patterns that promote repeated, automatic behaviors that may begin to operate outside of an individual's control. We hypothesize that unpredictable, low value rewards and cues that promote frequent, automatic, undesired behaviors are contributing to reward learning and habit formation in a way that is not aligned with users' intentions or preferences. These may lead to feelings of being manipulated, feeling lack of control related to certain behaviors, and feelings of dependence on checking or going on Facebook that could be related to lower well-being.

- (?With a third party?) Review features and default settings on the platform to identify unnecessary examples of frequent low-value reward signals, automatic unwanted behaviors that contribute to low perceptions of control, and problematic domains that come from user interviews and/or the problematic use quantitative analysis
- Work with Margaret Stewart and other design ethics leaders to understand what design changes would be in line with the goals of reducing low value rewards and improving user control

---

Elena Davis created the document Aug 2, 2018 at 6:32 pm