UCLA Center for Critical Internet Inquiry

Stanford PACS
Center on Philanthropy and Civil Society
—
Digital Civil Society Lab

I, Obscura —

Illuminating deceptive design patterns in the wild

A zine by Stephanie Nguyen & Jasmine McNealy

Table of Contents

Part 1: Hello, Darkness...

You are yourself, the reasonable person that you are. You go about your day, check your phone, use your laptop, or desktop, or tablet. Read some things, play some things, use some apps, maybe watch some video or listen to audio.

On one occasion, not out of the ordinary, you come across a Tweet about a new streaming service that will carry content from Production Company X, that makes some of your favorite movies and shows, as well as new content coming soon. You visit the site where the service is offering a free introductory subscription. Register today and you can get one month free. After the intro period, you will be billed $20 per month. Like any savvy user, you set your calendar to remind you to delete your account before you are rolled over into paying membership.

When the day comes for you to end your free membership you logon and look for a way to delete your account. You try “Profile,” “Account settings,” and various other tabs to no avail. You try reading the terms of service, the privacy policy. You don’t have time to continue searching, but tell yourself you’ll come back later to find how to end your service. But you forget.

You check your credit card balance the next week and a charge of $20 from the streaming service appears on your statement. You have just experienced a kind of dark pattern.

Dark patterns are “interface design choices that benefit an online service by coercing, steering, or deceiving users into making unintended and potentially harmful decisions.”[1] At the most basic, dark patterns use design to allow organizations to get something out of individuals that they usually would not be able to absent obscure tactics. In the subscription example above the design to keep individuals from deleting their accounts, making them continue their subscriptions beyond the free period, and allowing the streaming service to collect $20 it may not have been able to earn but for the difficulty for the user in ending their membership. This is only one example of a dark pattern; there are many more with connected harms, some monetary, others emotional, even others discriminatory.

Scholars have and continue to write about the implications of dark patterns within different contexts and on different communities. Advocates continue to urge organizations to discontinue these kinds of designs, and for policymakers to prohibit and punish these deceptive practices. Regulators in the United States and around the globe are wading into the fray to hold organizations accountable for the consequences of their designs.

I, Obscura hopes to illuminate dark design patterns by telling stories. A compilation of case studies, this zine offers readers a set of dark pattern examples, along with possible design and policy solutions. These examples assist with demystifying deceptive design the possible harms to individuals, and to prompt policymakers to action.

[1] Mathur, A., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019). Dark patterns at scale: Findings from a crawl of 11K shopping websites. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1-32.

Part 2: Harm(y) of Darkness

The subscription example, where a user had to pay $20 in spite of attempting to delete their membership, represents how dark patterns can be harmful. The user lost money, so this represents a kind of financial harm. But as many different kinds of harms exist as there are kinds of dark patterns, including shaming, loss of control, and discrimination. Here are a few of the possible harms stemming from dark patterns and where you can read more about them.

Denied Choice
When a person left without the ability to make informed decisions through the product or service. E.g. A social profile is public by default, a platform disguises data collection, or obfuscates prices . *E.g.) Case study 9: Hidden cancellation fees: How dense disclaimer text on restaurant websites can create an obstruction of choice on Yelp*

Experienced Discrimination
When a platform collects user data a person may experience unfair impacts based on demographic identifiers such as age, race, and gender. *E.g.) discrimination in employment, insurance, housing, education, credit, access to opportunities etc.*

Felt Shamed
When the sharing of personal information leads to a third-party contacting the person in an unwanted way, or to that person experiencing difficulties in personal relationships due to stigma, or harassment , . *E.g.) Case study 6: Disguised advertisements: How subtle forms of advertising can create user deception and unwanted persuasion*

Felt Tricked
When a user's decision is influenced in a way that may not be advantageous to their needs through the use of words, user experience, or user interface elements that nudge them in a particular direction, like making the user feel guilty about their choices or causing emotional or psychological distress. *E.g.) Case study 4: Walking Through First Class: How games like the Barbie Dreamhouse Adventures can blur the lines between pay and play*

Lost Money
When a person's purchasing decisions are manipulated, so that they end up buying items or paying for more than intended. *E.g.) Case study 2: Pay to Progress: Making in-app purchases with Candy Crush almost a necessity.*

Lost Privacy
When a person does not willingly consent to data collection, does not have a choice to decline collection in order to use a product or service, and/or is not informed where their data is shared, the company provides data to third parties without the person's knowledge, or makes posts easy to "overshare." *E.g.) Case study 5: Deceptive Divisions: How Google Map settings can trick users into giving away personal information under the guise of feature enhancements*

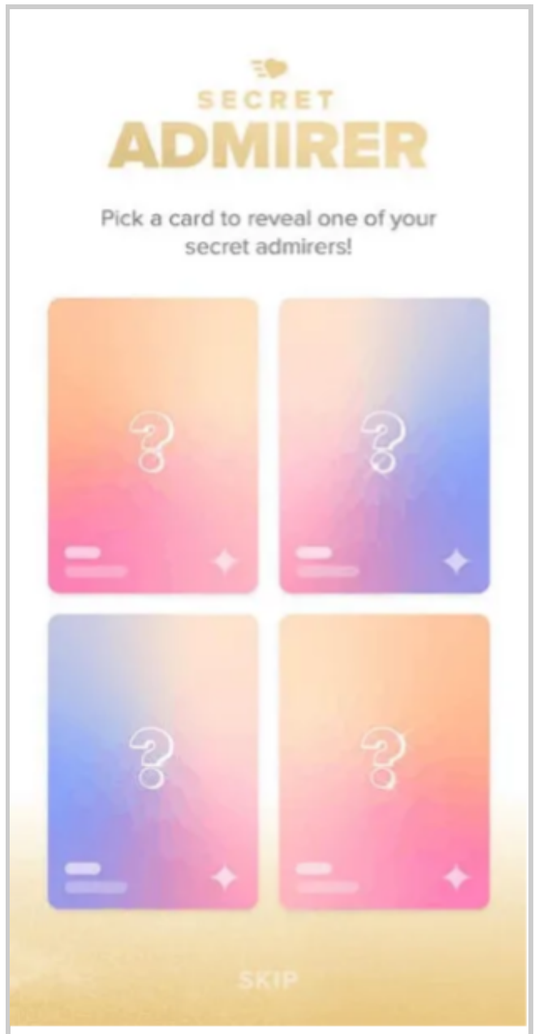
Wasted Time
When an organization deliberately makes a person's desired path more cumbersome by creating long and tedious processes that encourage the user to choose a path they do not want. *E.g.) Case study 1: Swiping Slips: How Tinder monetizes your muscle memory or Case study 3: Automatic content generation: How default design settings can contribute to online addiction through Instagram.*

Emotional Manipulation
When an organization creates a user experience that creates a feeling of urgency or panic to persuade the user to make particular choices and or continue interacting with the system in a particular way. *E.g.) Case study 8: Urgency: How creating fear and emotional nudges manipulates users into making accelerated decisions with McAfee's security products*

Part 3: Case studies

Case study 1: Swiping Slips: How Tinder monetizes your muscle memory

By: Ryan Christopher Tan



A mini-game prompting you to guess which card hides a match with someone who liked you.

Key Tinder Statistics

66 million monthly active users
Active users log in on average 4 times a day.
1.6 billion swipes per day (as of 2018)
1.5 million dates per week

45% of college students say they use Tinder mostly for confidence boosting procrastination
Tinder swiping among Gen Z users increased by 39% in first few months of coronavirus lockdown
Tinder accounts for 58% of Match Group (Tinder, Match.com, okcupid, Hinge, etc.) revenue

https://www.businessofapps.com/data/tinder-statistics/

In 1995, Gary Kremen, an American engineer and entrepreneur, had an idea that many dismissed as ridiculous. The Internet was still in its infancy when he launched match.com, the world's first online dating site. To grow its network, Kremen got all of his employees, including his girlfriend and himself to sign up for an account. His girlfriend left him for a man she met on match.com, proof that Kremen's idea was not so ridiculous after all: **you can find love online**. Kremen also proved love was profitable.

The Internet has grown rapidly since then. Online dating is now the most common way for a couple to meet in the U.S., with nearly 40% of all heterosexual relationships starting through platforms. Tinder is the largest dating app of all, with over 7.8 million users in the U.S. The app's popularity is in how it made love simple: swipe right for yes, swipe left for no. Tinder lets you swipe on up to 100 free love interests in a day, all from the palm of your hand. But does the gamification of romance come with other insidious techniques, or is all fair in love and war?

Context: Love at first swipe?

Tinder is attractive because you get to meet so many more people than you would with traditional dating methods like going to a bar. But beyond getting dates and making connections, using Tinder activates our attraction to rewards:

> *"In our technosexual era, the process of dating has not only been gamified, but also sexualised, by technology. Mobile dating is much more than a means to an end, it is an end in itself. With Tinder, the pretext is to hook-up, but the real pleasure is derived from the Tindering process. Tinder is just the latest example for the sexualisation of urban gadgets: it is nomophobia, Facebook-porn and Candy Crush Saga all in one."*
> *The Guardian*

This shopping ad displaying models appears in the same queue as your feed of potential matches.

Instant matches lead to instant gratification and each new face is a new potential for dopamine. And the deluge of enticing content is extremely effective: millennials spend around 80 minutes per day on dating apps, or 10 hours every week.

Tinder has loads of features to get you to keep swiping. At any time, you can see the number of people who have already liked you in a blurred image. Tinder will notify you when you swipe left on someone who likes you, signalling a "near miss" in the gamification. Every day, Tinder customizes new "Top Picks" along with a free "Super Like" they point out when giving a regular like to a "popular" user. Once a day, Tinder will randomly invite you to play a game that hides a "secret admirer" from you: guess the right card, and you get a match!

The potential for reward plus continuous swiping results in an experience "as lulling in its eye-glazing repetition as a casino slot machine, the chatting phase

ideal for idle, noncommittal flirting." But while swiping can ultimately lead to a potential relationship, Tinder also takes advantage of this "casino-slot machine" design to capitalize on your mistakes. In early online dating sites, you would have to intentionally submit "yes" or "no" on somebody you were interested in. Now, Tinder capitalizes on an accidental mis-swipe, punishing you for giving the wrong answer.

Dark Pattern: Swiping Slips
Tinder has created a fast-paced system that is prone to user error, an example of the dark pattern **Swiping Slips**, where if you're in a flow of swiping left on people, a lack of attention may cause you to reject someone you wanted to match with. Rather than trying to prevent or recover from user error, Tinder capitalizes on it.

Potential Harms: Monetizing Accidental Swipes
Attracted to Ads? (False Positive)
Most users are looking for matches and will swipe right on card after card, entering an inertia of repeated likes. But Tinder interrupts this stream of content with advertisements, displayed on cards similar to user profiles. Some of these ads feature attractive models, perhaps tricking an inattentive user even further. Engaging with the ad uses the exact same functionality as the core app: swipe left to skip it and swipe right to see it, meaning it's all too easy to open a link you didn't intend to.

Pay to Undo (False Negative)
Of course, the reverse could also be true. If you're in a streak of repeated left swipe "no's", you could easily accidentally swipe left on someone you are interested in. When it comes to romance in the modern day, that's a big slip to make: one wrong swipe could mean losing the love of your life! Tinder, of course anticipates this and has an undo option prominently displayed in the bottom tab. Except that undo option is a premium feature, Rewind®, and to use it you need to subscribe to Tinder Plus®.

What's love got to do with UI?

Among the 10 usability heuristics in UI design are the principles of helping users prevent and recover from errors. Tinder intentionally exploits their primary mechanic to get you to slip up for their own benefit. It would be as if Microsoft Word asked you to confirm an ad by typing, or if Facebook made you pay to edit your latest post.

As we've covered, Tinder is the leading platform in dating apps, online dating being the number one way that couples meet. Love is a core human experience that's of highest value to every user. Tinder's monopoly on romantic partners is a lot of power to hold. And they seem to know it, given how they hold the undo button hostage.

Potential Solutions
Find other ways to advertise content. An ad could be dismissed or engaged by tapping on its buttons instead of the core swipe. Platforms should not exploit the core mechanic or abuse users' muscle memory for profit.

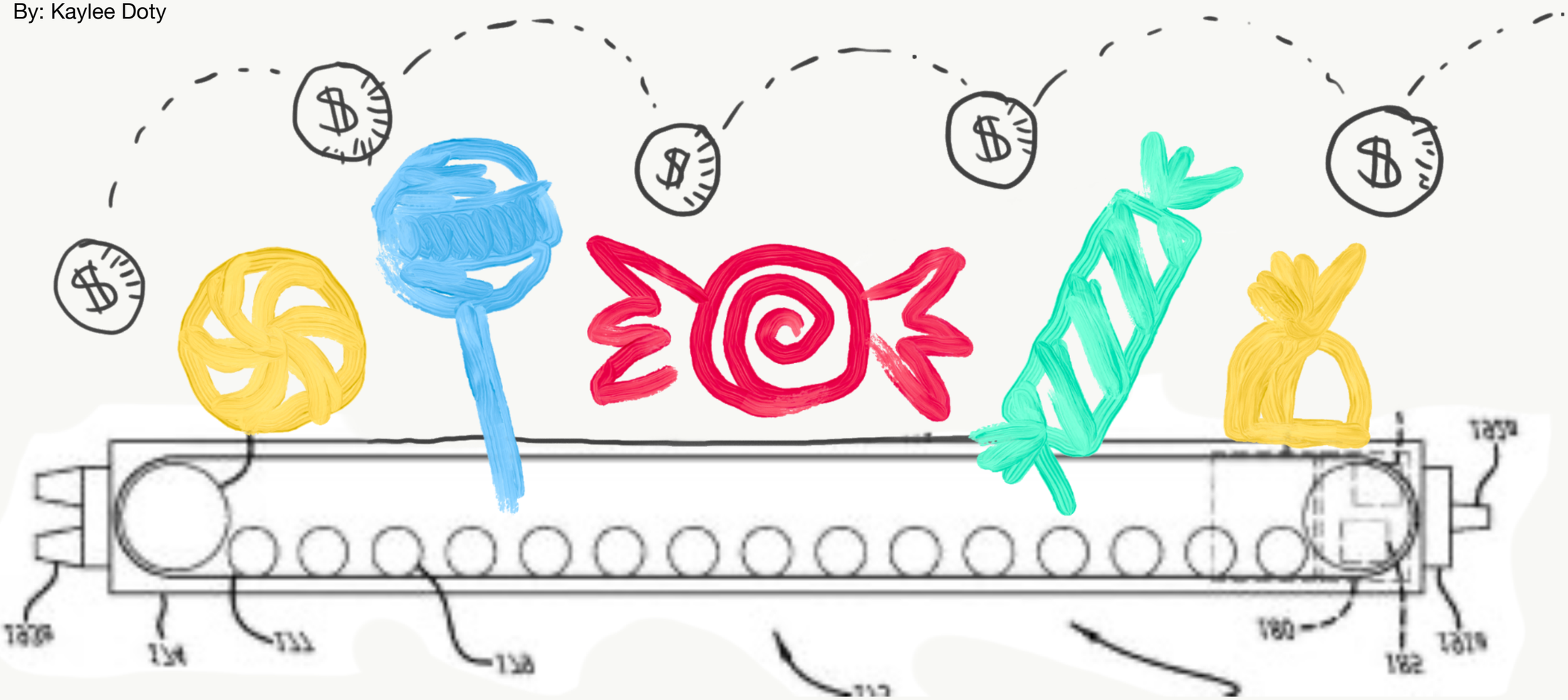
Make the undo button option free for all users. Platforms should aim to prevent and allow users to recover from errors, not abuse them. A service should be designed with the users' best interests in mind, otherwise it is acting in bad faith.

Do not rely on the gamification model. Hinge, another Match Group service, is "designed to be deleted" and takes a slower profile-based approach that is less prone to user error.

Preview from Tinder's Google Play Store page. Accessed 5/13/2021.

Case study 2: Pay to Progress: Making in-app purchases with Candy Crush almost a necessity

By: Kaylee Doty



Dark Pattern: Pay to Progress

Pay to progress is a dark pattern prevalent in mobile games where levels become so difficult that in-app purchases are almost a necessity. We have all likely heard of the popular and addictive game Candy Crush Saga, which heavily uses this dark pattern to keep users invested. The game works like this: users complete colorful puzzles in each level of the game, and as they progress the thousands of levels become increasingly difficult. This seems normal. But there is a catch. In Candy Crush, a life system that allows users five attempts at a level. Each time a user fails a level, a life is lost in-game; if the user runs out of lives, they must either wait until the lives naturally regenerate, ask a friend for lives, or pay to continue playing. Considering the highly addictive nature of the game, waiting 30 minutes for a life to regenerate can become almost unbearable. And while asking a fellow Candy Crush adventurer to send extra lives is free, this method is unreliable. As such, many users, who run out of lives but have a strong desire to beat a challenging level, default to paying real money to play the game.

Context: Nudging users to spend more money to advance in a game

Candy Crush's life system has no impact on any other aspect of game play. The requirement for users to wait a painfully long time after failing a level five times serves no purpose other than to nudge users into spending money on the game. Additionally, some of the Candy Crush levels are extremely difficult to complete the first time around; as users may be failing levels often, it is easy to see how this life system can quickly drive addicted users into steep financial consequences.

Potential harm: Financial harm

The life system in Candy Crush presents a financial harm to users. With pricey options for buying lives (replenishing five lives costs around $2.00 to $2.50 each time), users may find themselves spending large sums of money on a game that they initially thought was free to play. Additionally, the constant nudging from Candy Crush to buy more lives in game makes it easy to spend money on the game without even realizing how much it all actually costs.

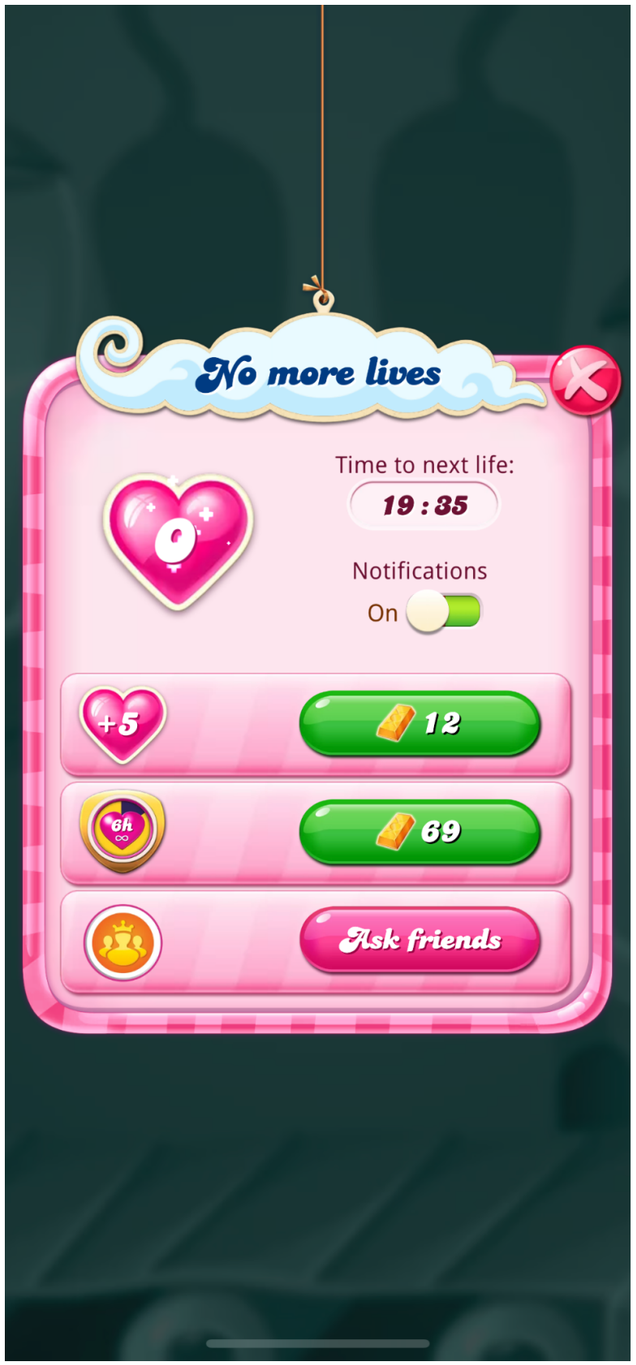
Potential harm: emotional manipulation

The financial burden is not the only harm of this life system — there is an addiction harm as well. With the requirement to wait consistent time intervals for lives to naturally regenerate, users are primed for addiction to the game, logging back on every thirty minutes or so to see if they can keep playing or not.

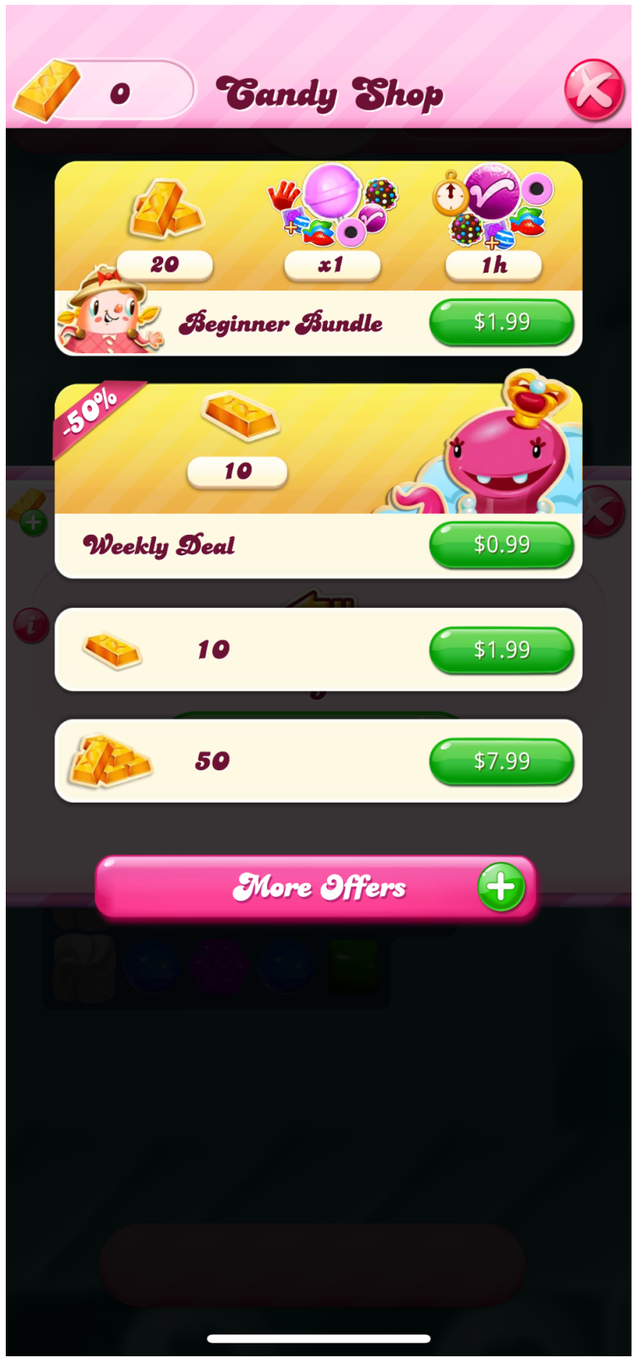
Potential Solution: Rethinking in-game consequences

Instead of having a life-based punishment for losing levels, Candy Crush could implement less significant in-game consequences, such as losing a small amount of virtual in-game money. The game already has a system of in-game currency with "gold bars," which are used to buy extra lives and boosters that help users advance in the game. By rewarding users with gold bars for completing levels and deducting some when they lose levels, Candy Crush could make the game completely free to play. Despite the positive benefits this change would have for the user, the makers of Candy Crush would lose out on significant profit. Their current game model generates about $800,000 daily from in-app purchases alone; without any incentive, it would be difficult to convince them to trade in all of this money for the mental well-being of their users.

Caption: Screenshot of Candy Crush outlining the pressure to buy additional currency to continue to the next life.



Caption: Screenshot of Candy Crush to outline the cost to the user to buy game currency within the application.



Case study 3: Automatic content generation: How default design settings contribute to online addiction through Instagram

By: Kally Zheng

Context: Increasing user engagement through related posts: In another attempt to increase user engagement, in August 2020 Instagram implemented a feature called *Suggested Posts*. Originally, when a user scrolled through a feed and reached previously viewed content, Instagram allowed users to continue to view posts from the people that user follows. Naturally, revisiting old content is less engaging than new content. However, with Suggested Posts, Instagram's main feed - by default - displays organic content in users' feeds from suggested accounts and advertisements, and appears in users' main feeds immediately after the "You're All Caught Up" notice.

Dark Pattern: Continuous Content

Suggested Posts eliminates the user's natural stopping point within their main feed and creates an infinite scrolling experience, in which the need for users to click to the next page is eliminated because content continuously loads at the bottom of the screen. The inventor of the infinite scroll Aza Razkin explains that, "if you don't give your brain time to catch up with your impulses, you just keep scrolling." "Individuals with addictions engage in behaviors that become compulsive and often continue despite harmful consequences," said the American Society of Addiction Medicine. Infinite scroll, using suggested posts, can easily increase consumption of online content. Users can open Instagram intending to only view a single video or post. Before they know it, they may waste hours scrolling through content.

Continuous Content is a dark design pattern because it automatically opts users into more appealing and addictive content. In an interview with TechCrunch, Robby Stein, the Instagram director of Product stated that, "the goal [of the suggested posts feature] is to make it clear when you're all caught up so you can decide how you want to best use your time." However, this feature does the opposite. It fails to give users a choice between an addictive infinite scroll and a natural stopping point. Instead it defaults to an endless feed of organic content.

Potential Harm: Lack of real choice for users

Instagram will infinitely generate more content like suggested posts by default, so users can continue scrolling in their main news feed. A default is, "a selection automatically used by a program in the absence of a choice made by the user." These presets make choices for users unless they choose to object. More often, these defaults make decisions for users that they are not aware of making. Defaults have a large impact on user and consumer behavior.[2] In fact, Microsoft ran a study on Word and found that, "less than 5% of the users surveyed had changed any settings at all."[3]

Defaults hold significant power. For example, in countries where laws make organ donation the default, more than 90% of people register to donate their organs. In countries where people must explicitly 'opt in,' fewer than 15% of people register.[4] This is an instance in which defaults can be used to motivate positive impacts such as increasing available organ donations.

Default design patterns like suggested posts can lead to users spending more time on Instagram, which can negatively impact mental health and increase online addiction. For example, a 2-year study funded by the NIH revealed a significant association between depression, stress, anxiety and Internet addiction, which can stem from manipulative design patterns like continuous content.[5] This study found that problems stemming from excessive internet use deserve serious attention from U.S. mental health and psychiatric communities. In fact, according to a study conducted by researchers atUSC, "the greater your level of Facebook addiction, the more significant the reduction in gray matter in the amygdala. This pruning away of brain matter is similar to the type of cell death seen in cocaine addicts."[6]

---

[2] Wang, Yue & Mo, Yiu-Wing. (2018). The Effect of Default Options on Consumer Decisions in the Product Configuration Process.

[3] Spool, Jared. "UIE Archive." *UIE Brain Sparks Do Users Change Their Settings Comments*, 2011, archive.uie.com/brainsparks/2011/09/14/do-users-change-their-settings/.

[4] Davidai, S., Gilovich, T., & Ross, L. (2012). The meaning of default options for potential organ donors. Proceedings of the National Academy of Sciences, 15201-15205.

[5] Saikia, Anku M et al. "Internet Addiction and its Relationships with Depression, Anxiety, and Stress in Urban Adolescents of Kamrup District, Assam." *Journal of family & community medicine* vol. 26,2 (2019): 108-112. doi:10.4103/jfcm.JFCM_93_18

[6] He, Q., Turel, O. & Bechara, A. Brain anatomy alterations associated with Social Networking Site (SNS) addiction. *Sci Rep* 7, 45064 (2017). https://doi.org/10.1038/srep45064

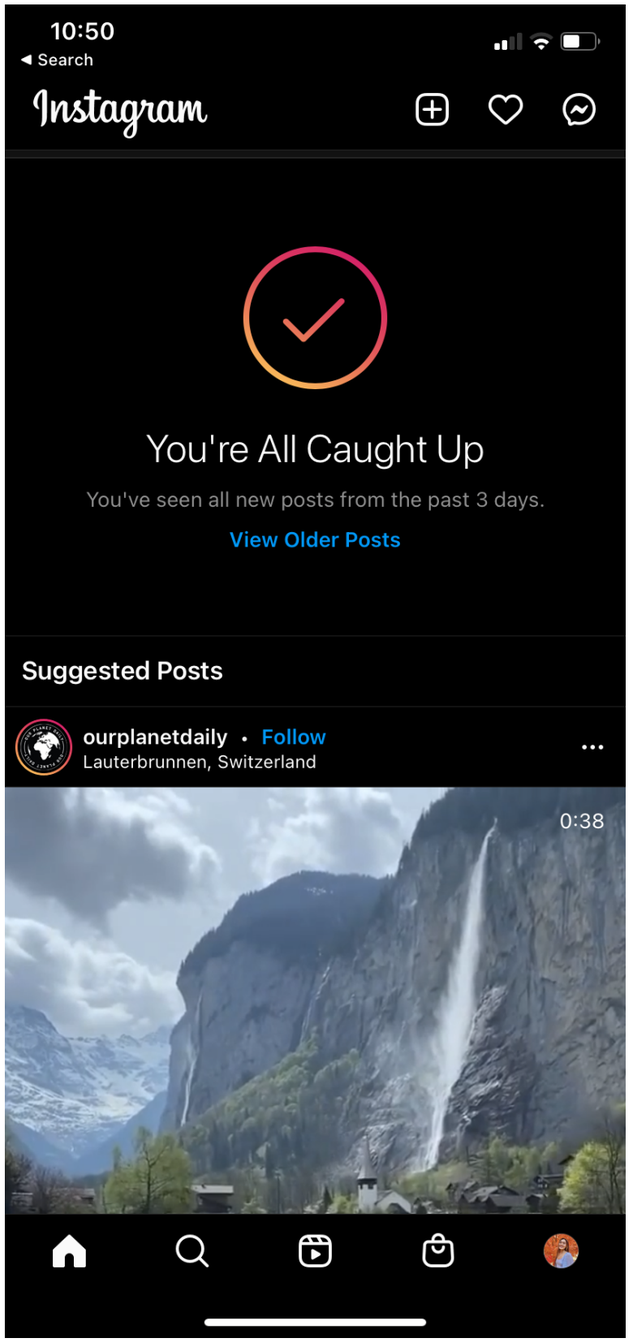
Potential Solution: Explore bills that explore effective ways to regulate manipulative patterns

There are several bills and policies that can serve as inspiration in how to regulate this manipulative design pattern. For example, Representative Josh Hawley's SMART Act outlines the banning of infinite scroll and auto-refilled content, as well as a requirement for social media to include "natural stopping points." But banning infinite scroll is difficult because it is hard to determine a specific threshold for the number of posts until a feed would be considered infinite. For example, a social media platform could argue that their feed is not infinite because after 1,000 posts a stopping point is implemented. For the user, this experience would still feel like an infinite scroll. The "You're all Caught Up" notice does in fact indicate a natural stopping point for scrolling. However, it could be a poor user experience and feel restrictive to completely stop the generation of content for users. Instead, users should be given an active, deliberate choice to either view previous content or the suggested posts.

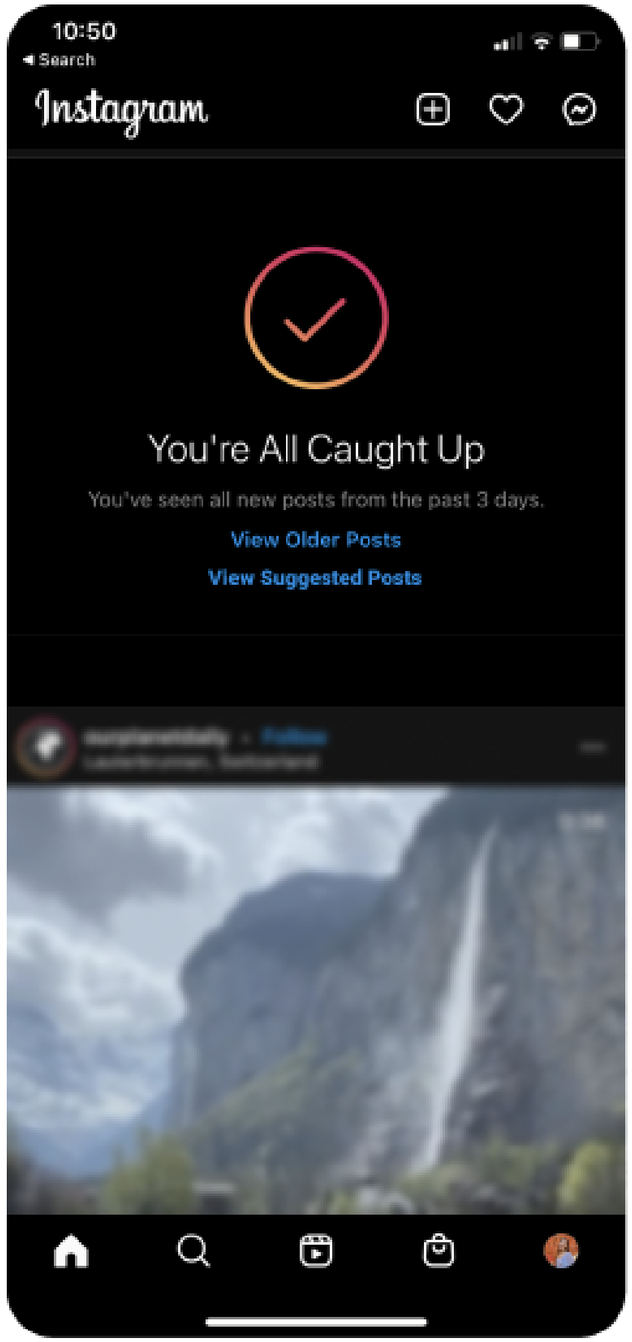
Another possible route for regulation may be found with the Canadian *Anti-Spam Legislation,* which prohibits businesses from automatically opting customers into email subscriptions. Similarly, the design of the continuous content feature should not automatically opt customers into infinite addictive content generation. The EU's GDPR also requires consent to be opt-in to a service. It defines consent as "freely given, specific, informed and unambiguous" given by a "clear affirmative action." It is not acceptable to assign consent through the data subject's silence. When users reach the You're All Caught Up notice, they should be forced to make an active decision rather than a passive decision between previous content and "Suggested Posts"

The image to the left is one potential design implementation that would protect users from online addiction. In this mockup, users must take affirmative action to see suggested posts content. Users are forced to make a decision rather than submit to the default decision as the feed will be blurred and obstructed until they make a decision. The "View Older Posts" and "View Suggested Posts" options are a similar size. The all caught up notification and screen is preserved.

Caption: Screenshot of Instagram showing the "You're All Caught Up" screen which also highlights "Suggested Posts" in the main newsfeed experience



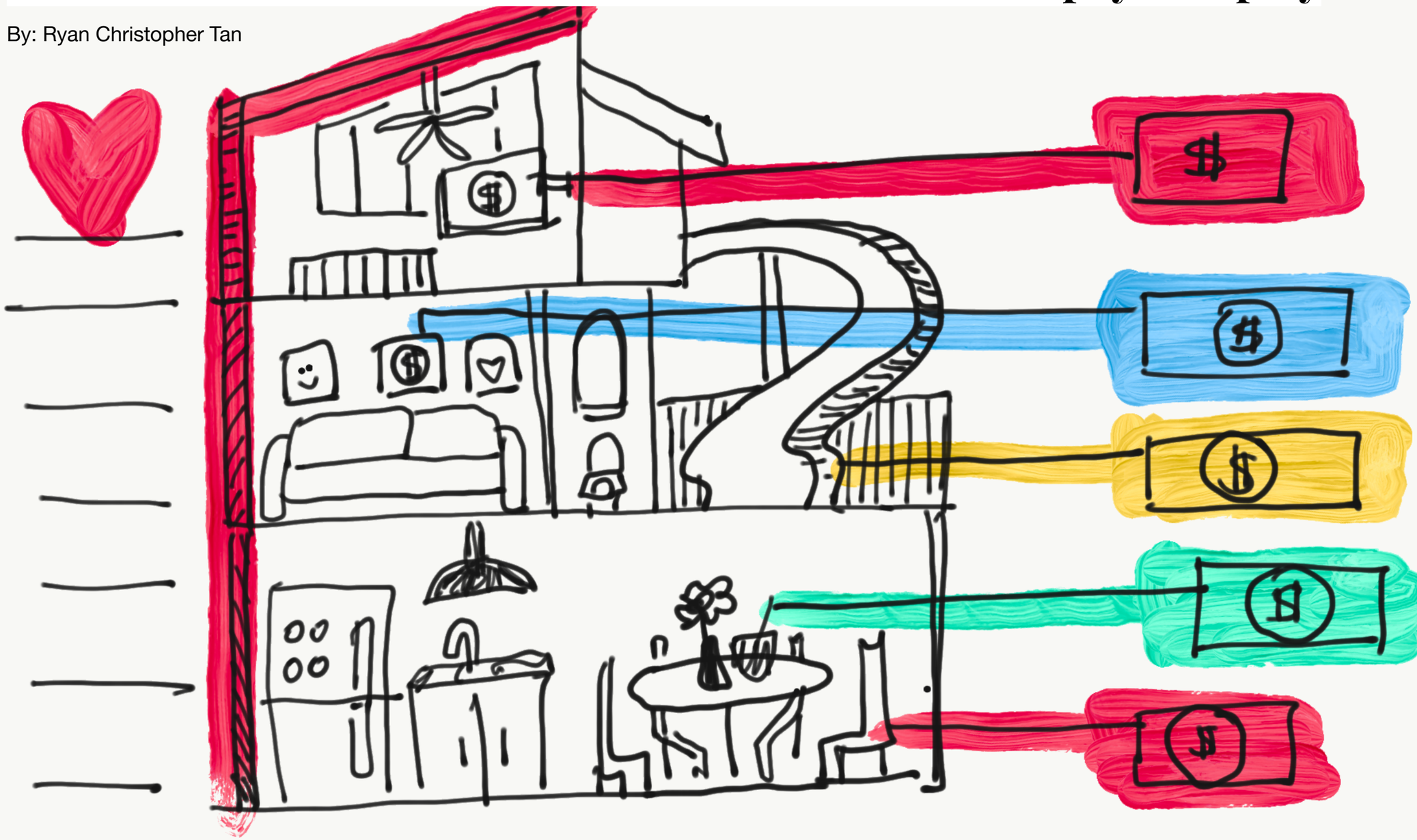
Caption: This could be improved by making the "You're All Caught up" screen be a popup modal with clearer and more visible buttons.



Instagram could go one step further by allowing users to permanently disable the suggested posts feature. In this case Instagram could keep the "You're All Caught Up" notice but allow users to choose if their feed defaults to older posts or suggested posts. By doing this, Instagram will create a better user experience. Although they will miss out on an opportunity to present promoted materials for the users who disable this feature, they will gain overall user trust.

Case study 4: Walking Through First Class: How games like the Barbie Dreamhouse Adventures can blur the lines between pay and play

By: Ryan Christopher Tan



Barbie Dreamhouse Adventures is one of the biggest kids' apps on the market right now. Barbie's Malibu Dreamhouse is packed with customizable and colorful activities, ranging from interior design to dance parties to shopping sprees. The App Store lists the game as free with in-app purchases, and appropriate for ages 4+. But how do dark patterns target children based on the unique ways they interact with the world?

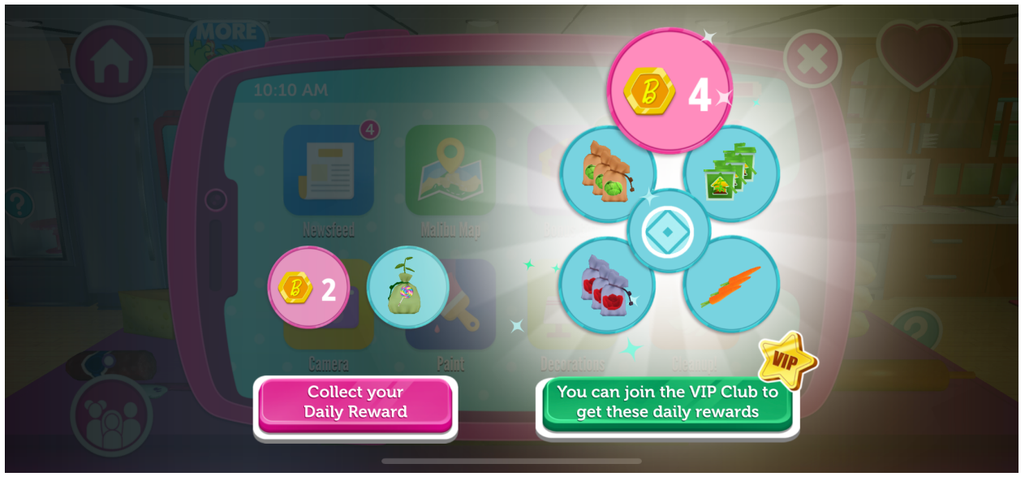


Image above: Screenshot of Barbie Dreamhouse Adventures

Dark Pattern: Walking Through First Class

Upon first launching the game, a child sees a brand new reward welcoming them. After an exciting flurry of taps breaking open a glowing present box, they finally can claim their Daily Reward, which shows not only what they won today, but what they *could have won* today if they had only been part of the VIP Club.

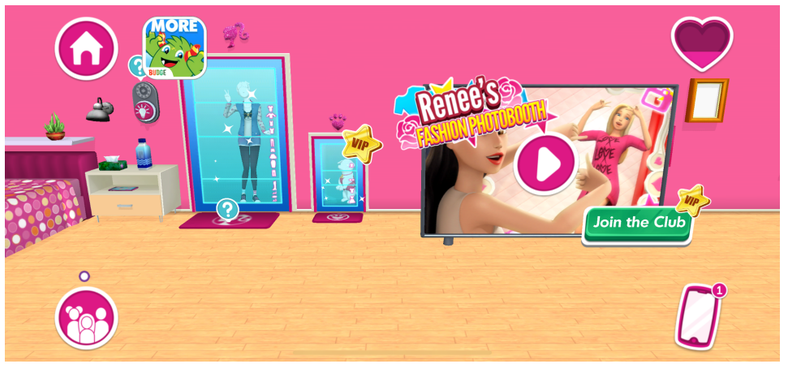


Image left: Screenshot of an explorable room in Barbie Dreamhouse Adventures

This dark pattern is called **Walking Through First Class**. It's similar to the feeling of flying in coach class and you have to shuffle through the first class cabin with their elbow room and fancy drinks. It serves as a reminder of the **VIP** status you don't have, and it's bad design to make users feel inferior just when they navigate your service. Below are some examples that highlight this dark pattern in action.

Barbie's world is built. Sure there are lots of fun activities to do for free. But if someone wants to do the big, exciting stuff that takes up most of the screen, they'll have to pay for it. Note the enticing VIP-exclusive (i.e., exclusionary) content. Adults are more likely to recognize it for what it is: an advertisement for premium in-app purchases. But these ads manifest themselves as part of the game space, not around it. Look at this ad placed against a wall in front of a rug and propped on feet to look like a playable TV:

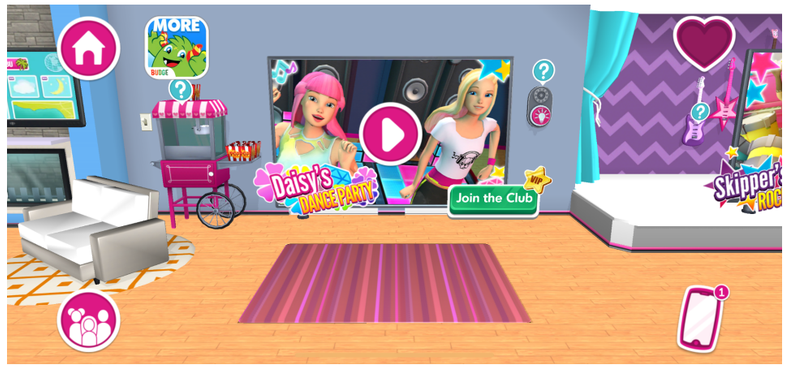


Image left:: *An ad to "Join the Club" featuring Barbie and a friend.*

A more egregious VIPing example is in the pool that the Playhouse prods you to visit. When you arrive, it's completely covered up, save for an ad of Ken, who certainly looks like he's enjoying himself in the VIP-only pool.

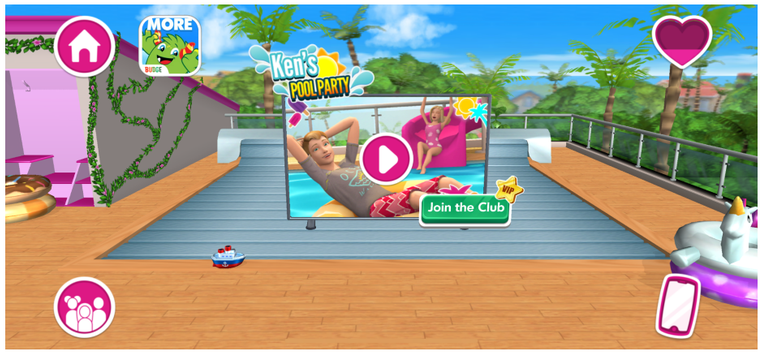


Image left: *Screenshot of the first view when you navigate to Barbie's pool*

As shown above, these ads are commonly placed throughout the game, which increases the likelihood of kids being navigated to the in-app subscription screen when they just want to explore the game.

5 Ways Children are More Susceptible to Dark Patterns:

1. Immature executive function (emotional control/attention)
2. Highly imaginative, trusting of characters
3. Susceptible to rewards
4. Indifferent to or unfamiliar with data privacy
5. Disconnect between currency and material value

FTC Dark Patterns Workshop 4/29/21

Context: Kid in a (VIP-Only) Candy Store

On dark patterns in children's apps, UX designer Chris Kernaghan writes:

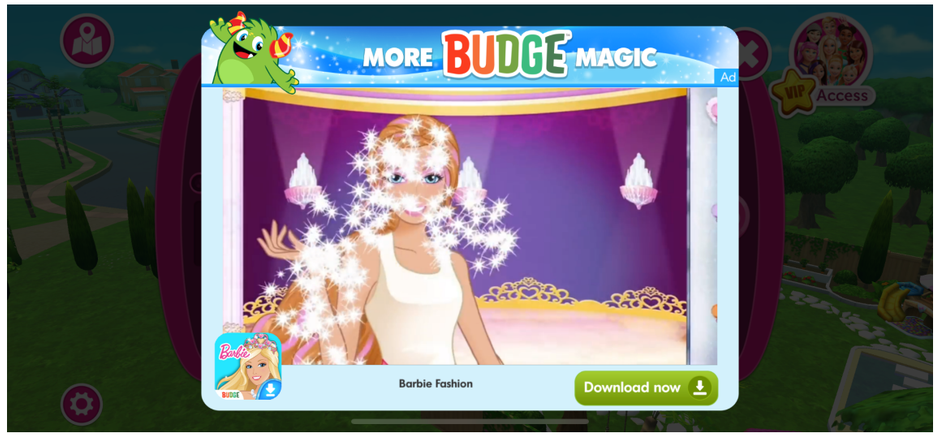
> "Young children are particularly unique in that they often can't tell the difference between what is an advertisement, and what isn't. They only see these advertisements as an extension of the fun they're already having, as they're likely designed with them in mind. They're also unlikely to critically challenge claims made, potentially resulting in a severe case of FOMO [fear of missing out]."

When dark patterns target kids, they target kids' psychology. Flashing graphics will heighten the chance that a young child will tap it, especially if they feature friendly characters. Combine this with how interactive ads are integrated nearly everywhere in the game world, and you have FOMO built in from the very first tap.

The game takes every opportunity to make VIP status much more attractive than the base game. Like a real dollhouse, you can place Barbie and her 15 friends in any of the rooms of the Dreamhouse. Well, only if you're a VIP. If not, then you can play with Barbie and her friend, singular. The player is free to scroll through all 14 of the friends they can't actually play with though!

Caption: Only Barbie and her friend can be added to the scene. Tapping the others pushes you to purchase again.

The confusing thing for a child player is that all of the game world is interactable, but not all of it is playable. Ads are represented in the game world in near-identical ways to the playable elements. More traditional ads also appear, but they take the form as an interactive game, rather than a commercial. A distracted child may not even be able to tell that this isn't a part of the game they're playing, but an ad for a different Barbie game entirely!



Caption: An in-game advertisement for another app entirely.

Harms: Accidental Purchases and Psychological Manipulation

There have been several cases of kids mistakenly buying in-app purchases, leaving parents unclear about how to unsubscribe from the service's monthly payment. Children's apps now often have purchase verification, like the simple

checks in the images below. Some of these checks border on mini-games of their own. Accidental purchases can go unnoticed; these authentifications should signal a clearer visual difference between the kid-friendly games and the adult-targeted purchasing.

Image above: In-game authentication to verify in-app purchases.

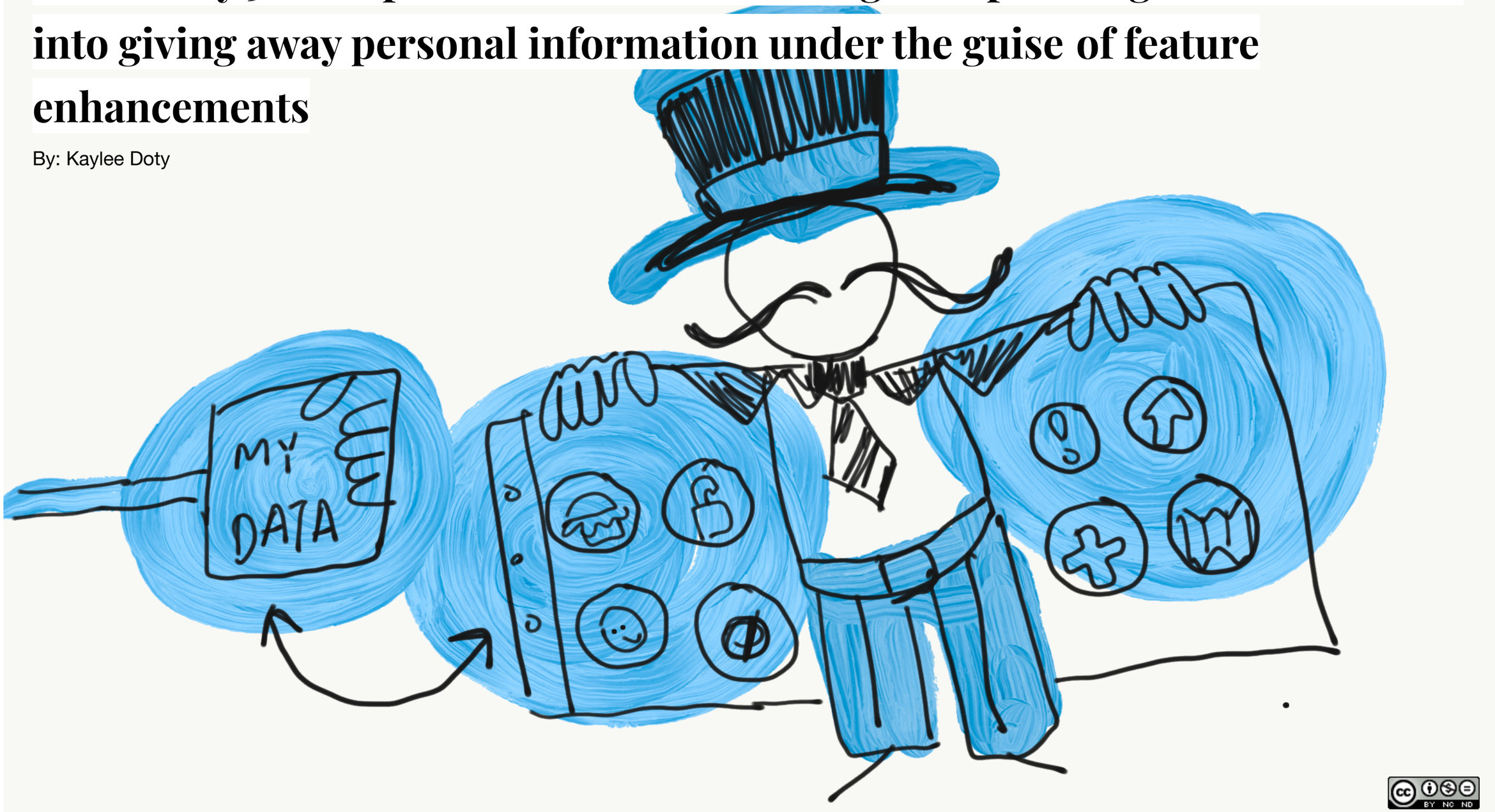
Beyond harm to parents, the greater harm is to children who are more vulnerable. These dark patterns target kids in ways that seem naive and obvious to us, but are known psychological tricks to manipulate them to engage and purchase in the app. Susceptibility to rewards means a growing numbness to daily prizes and flashing colors — this is good for short-term dopamine stimulation but bad if a child gets hooked.

Potential Solutions

The FTC emphasizes children's unique susceptibility to "unfair and deceptive marketing practices" and has settled cases of free apps illegally collecting children's personal data in free apps. In its settlement with Hyperbeard, the FTC ruled that the developer violated COPPA for failing to inform users' parents and obtain consent for their targeted advertising. We may turn to cases like these to motivate where the line is between a "fair" ad for kids, versus one that exploits their vulnerabilities.

Case study 5: Deceptive Divisions: How Google Map settings can trick users into giving away personal information under the guise of feature enhancements

By: Kaylee Doty



Dark Pattern: Deceptive Divisions

Deceptive Divisions is a dark pattern that tricks users into giving away their personal information by creating multiple settings that must be manipulated to disable one feature. Claiming that having more settings allows for a better, personalized user experience, some companies may split up one feature into multiple settings, making it difficult for users to maintain control over their data. For example, many of us would not voluntarily give a company access to record our location around the clock, but if you have the Google Maps app installed, it is likely that Google is collecting this information without your knowledge.

As a navigation app, access to your location is essential for Google Maps to function properly. What most users do not know, however, is that location tracking is enabled by default when the app is installed, allowing Google to record your location wherever you go. Even for individuals who notice this, Google has made turning off location tracking difficult. Instead of having one clear setting that disables location tracking, the user has to manipulate multiple settings to stop Google Maps from saving their location data. Specifically they must manipulate both the "Location History" and "Web & App Activity" settings. "Location History" could appear like a setting that disables location tracking,but actually allows users to view their past locations. Visiting this setting does not allow the user to request that Google stop storing their location information. Rather, users must read through dense descriptions in the “Location History” setting to figure out that the "Web & App Activity" setting—which does not reference location in the name—actually controls how Google stores and uses your location data.

Why is this a dark pattern?

Users typically assume that each setting in an app controls one distinct feature, so it is not intuitive that *multiple* settings must be manipulated to disable location tracking in Google Maps. By dividing user control over location data into multiple settings, Google makes it difficult for users to identify how to turn off location tracking and obstructs user control over their personal information. Even though Google provides lengthy descriptions of what each data setting controls, users should not have to digest large pieces of text to understand how to maintain basic privacy over their personal information. Rather, controls around data privacy should be dialed toward the user’s preferences and, at a minimum, be clear and concise.

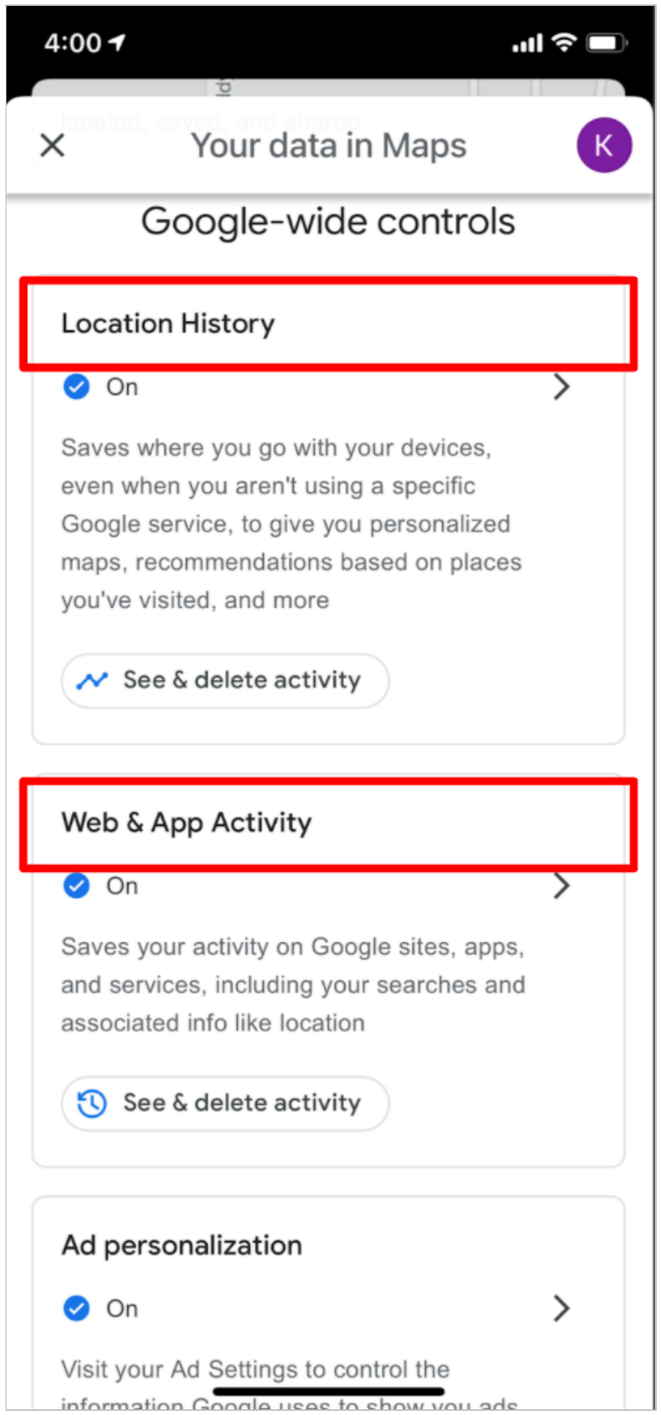
Potential harms: Overcollection of data without clear user awareness

The complex settings in Google Maps pose a privacy threat to users. They enable Google to track your every move without clear consent because the settings are turned on by default. Location information is valuable to data-collecting companies, and once your data is collected, you simply have to trust that it will not be used in malicious ways. Recent incidents in Singapore with Covid-19 contact tracing, for example, have shown that companies and authorities are not always transparent about how data is shared. Data collected for contact-tracing is highly sensitive because it contains information about the user's every movement. In the Singapore case, although officials initially stated that data collected by the contact tracing app would solely be used for notifying individuals about exposure to Covid-19, they later gave police access to this location data after millions had downloaded the app. If Google similarly shared location information with police, for example, some users may face unjustified and severe consequences in criminal investigations, just because they could not figure out how to properly disable location tracking in Google Maps.

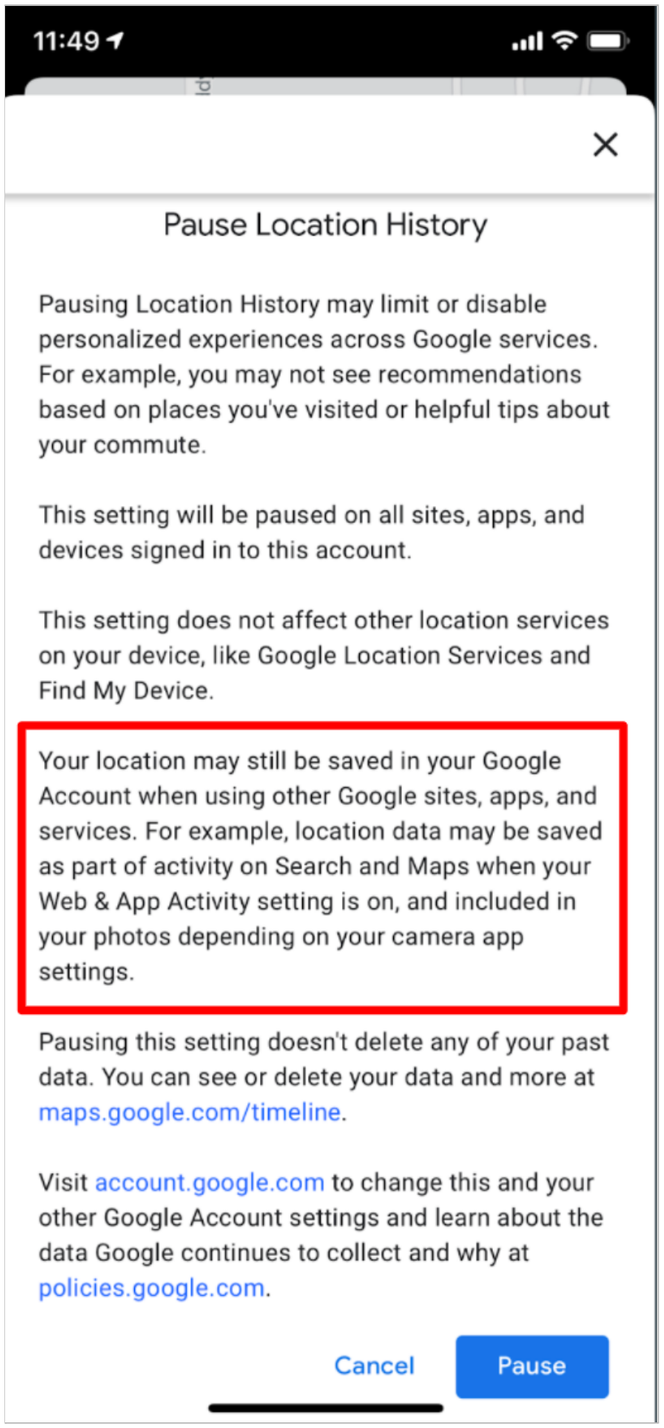
Potential solution: Combining control over location tracking into one setting

Combining "Location History" and "Web & App Activity" into one setting that allows users to fully control their location data would reduce the harm of Google's current layout. This new, merged setting should have a name that clearly identifies it as a location-related setting, such as "location data collection." Additionally, the description that accompanies this setting should be brief and concise. An example description of this setting could be, "this setting controls how we collect your location information. When toggled on, Google can record the location of your device."

Caption: Screenshot of Google Maps highlighting two separate places in the app settings where the user needs to change settings in order to disable tracking.

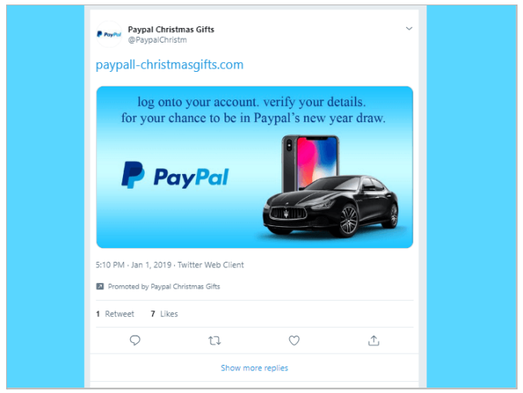


Caption: Another screenshot of Google Maps highlighting that locations may still be saved in your Google account even if they choose to "pause location history."



Case study 6: Unclear urging: How subtle forms of advertising can create user deception and unwanted persuasion

By: Kally Zheng



Context: Higher engagement, more conversation, impressions and views

In 2016 Twitter introduced a new advertising method called the "Hashflag". According to AdWeek, This product allows companies to spend 1 million dollars to have a custom emoji appear with a specific hashtag. These hashflags only exist for a limited time as set by the product's advertising campaign. In addition, they only exist on twitter.com or the twitter app. Hashtags such as "#shareacoke" help companies make their content more viewable to consumers and create higher engagement and awareness with advertising campaigns, products and business brands. They encourage conversation and as a result a higher amount of impressions and views. An article by the marketing consulting group Creative Agency highlighted that people notice these hashflags more than the average tweet. In fact, Twitter reports that the average attention increases by 10% when branded emojis are used in that advertisement. In fact, this same report explains that the median number of earned media generated is 5.3 million Tweet impressions, representing a 420% increase compared to the earned media baseline[7].

Dark Pattern: Unclear Urging

The problem with the hashflag is that it is a highly effective advertisement that is not clearly identified as an ad. This means users may be unaware that they are viewing an advertisement as well as are helping to promote ad materials. Other twitter advertisements are clearly marked as promoted materials. Because the hashflag takes up such a small portion of the screen real estate, the typical "Promoted by X" label could be difficult to see in this case.

Potential Harm: User Deception & Unwanted Persuasion

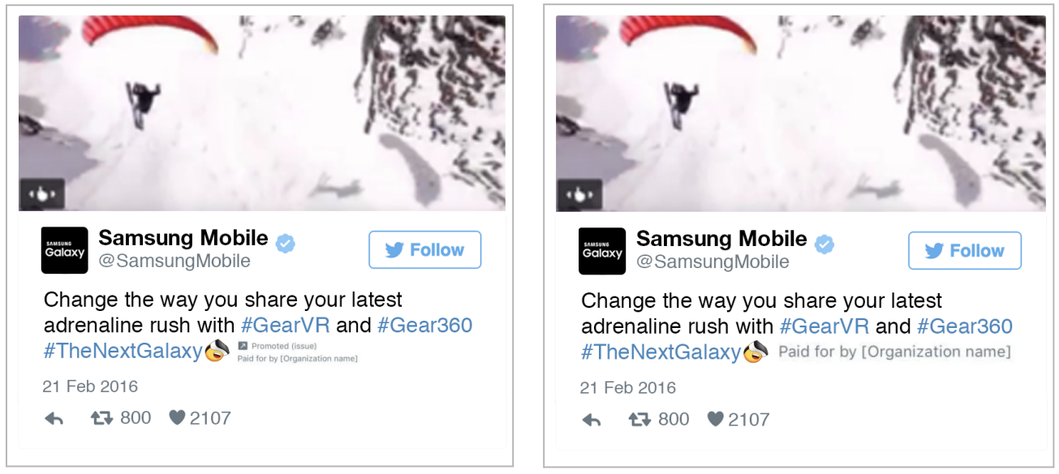
Advertisements influence consumer and user behavior. These hashtag emojis have all of the same subtle and manipulative impacts as a normal advertisement. However, they are not clearly labeled as advertisements, which creates user deception and unwanted persuasion. Twitter reports that, "By developing a branded emoji and leveraging additional ad formats, like conversational ads, Samsung was able to achieve a 75% year-over-year increase in usage of #TheNextGalaxy hashtag."

Ads are designed to make people purchase things. In a Stanford Business school experiment,[8] professors discovered that even when a user does not click on an engagement with an advertisement, they are much more likely to purchase the featured product rather than those who saw no advertising at all. Sometimes, ads can be deceptive. Navdeep Sahini, a researcher at the Stanford Business School study asserted that "the effect of advertising seems to happen through direct exposure and can result in people buying items even if people don't click on the ad itself […] consumers don't inadvertently click on an ad and buy a product without knowing they saw an ad but rather internalize it and later search for a product meaning that ad exposure has a deeply subtly and thus harder-to quantify effect." Deception of users on such a large scale is especially alarming. When considering the impact of political advertisements for example, this form of subliminal messaging could have wide scale impacts on voter behavior as well as the nation's political climate.

[7] https://marketing.twitter.com/en/insights/best-practices-for-supercharging-campaigns-with-branded-emojis

[8] Waikar, Sachin. "Disguised 'Native' Ads Don't Fool Us Anymore." *Stanford Graduate School of Business*, 8 Jan. 2018, www.gsb.stanford.edu/insights/disguised-native-ads-dont-fool-us-anymore.

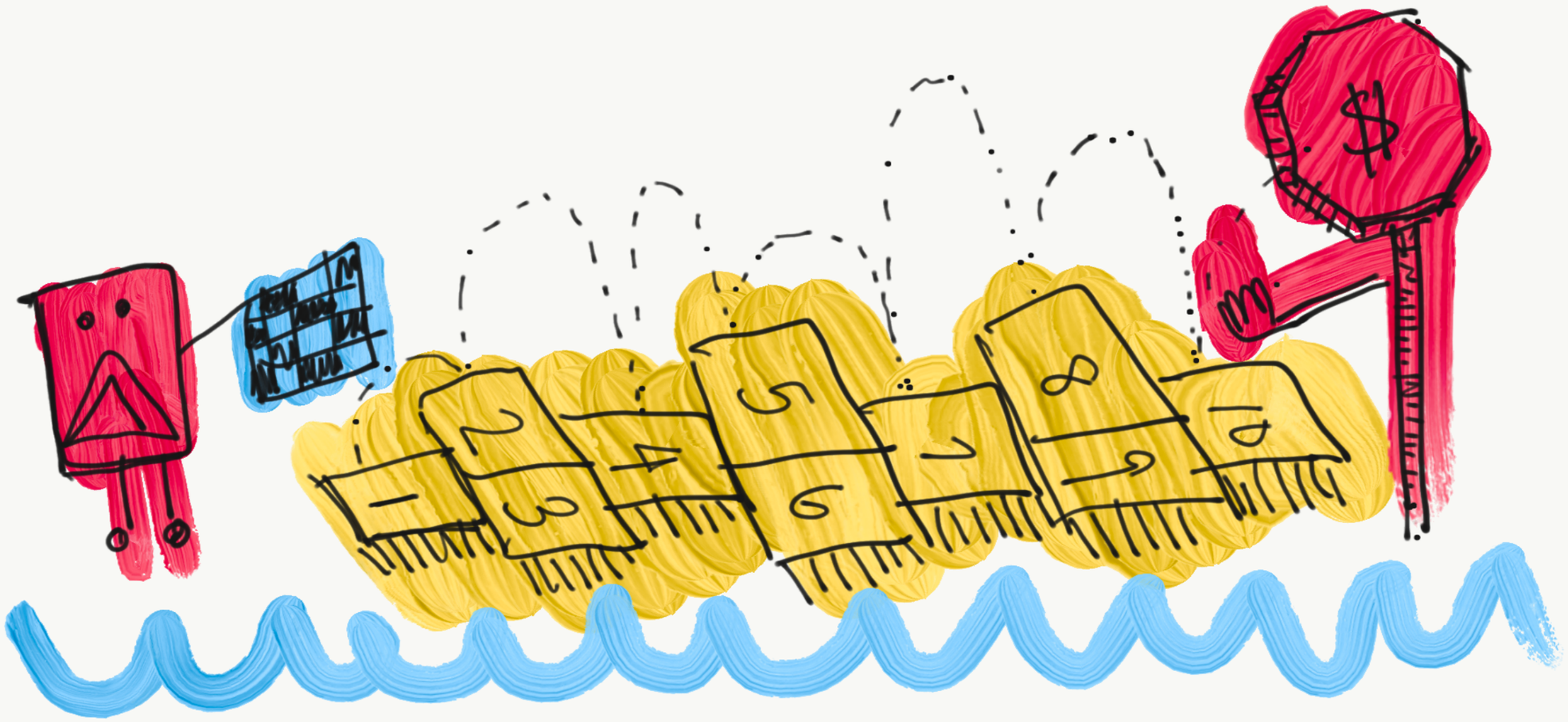
Potential Solution: Make endorsed content aware to users



The U.S. Federal Trade Commission's influencer disclosure regulations require that influencers label endorsed content clearly in a place that it is hard to miss so that viewers are aware that it is an advertisement. Digital and social media marketing practices have been strongly influenced by those FTC regulations. For example, Media Kix reported Sony and advertising agency Deutsch LA faced an FTC lawsuit over a failure to disclose that tweets were from employees of Sony's ad agency. The FTC argued that this failure to disclose the promoted nature of these tweets was deceptive and Sony was forced to settle. The Twitter hashflags are deceptive in a similar manner. Like the Sony Tweets, users are unaware that hashflags are promoted content. Users should be aware that they are contributing to advertisement campaigns when they view and use the emoji charged hashtag. The Twitter hashflags should be labeled in a similar manner to influencer disclosures. Twitter should clearly label all tweets that utilize the emoji hashtag as advertisement content. The label's should take the burden off of the consumer to determine if material is an advertisement. The following images show a potential design implementation of the suggested solution.

Case study 7: Hidden costs: How Adobe Acrobat sinks your time and energy to pressure you to pay at the finish line.

By: Ryan Christopher Tan



Unlike other examples of hidden costs, Acrobat doesn't just subtly get your foot in the door; they actively nudge you to seal the deal and purchase the product! They fully acknowledge the unexpected charges, which will deter most users, so it's not enough to just get you to the finish line. Instead, they pivot their strategy: get you near completion to the finish line and then once you're there, pressure you into paying after sinking in your time and energy.

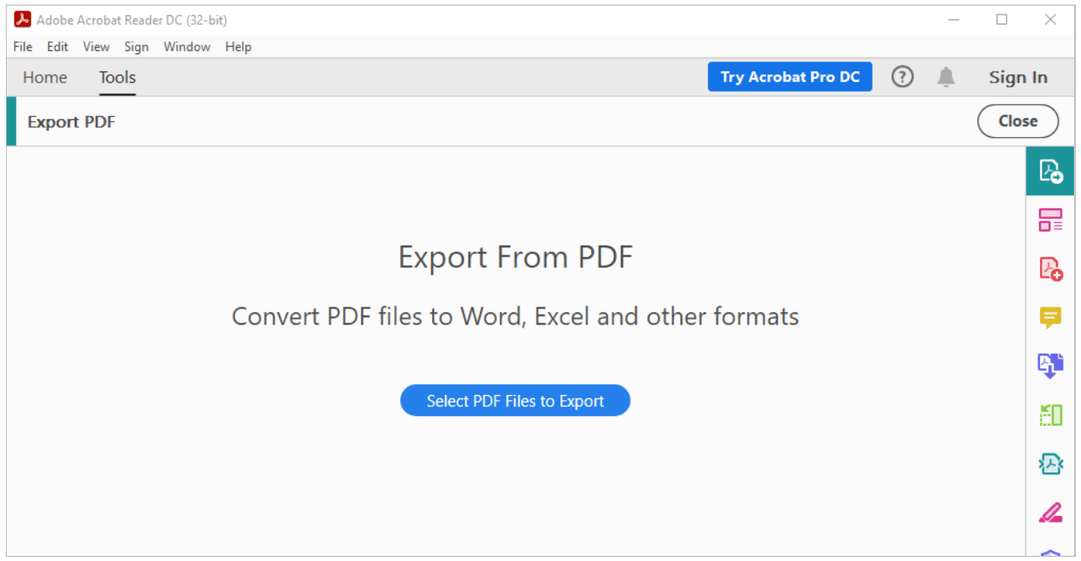


Image above: The free version of Reader's "Export" page, even though you can't actually use it unless you subscribe.

Dark Pattern: Hidden Costs: *Adobe Acrobat Reader* is a free service that allows you to "view, sign, collaborate on and annotate PDFs". *Acrobat Pro* includes paid services, like editing and converting PDFs into other formats, including Word or Excel. The application doesn't distinguish between these services in the free version, with the right sidebar displaying both categories together.

For many who work with PDFs, Reader's most valuable asset is its ability to convert PDFs into easily editable, user-friendly formats. The free version actually lets you upload as many files as you want to line up to export, even letting you select from all of the conversion options, and indicate the language of the original document. Once you've customized your preferences, a bold blue button invites you to Export to Word. Then, the user is taken to an online store. There are at least four sentences egging the user on that their hard work has almost paid off to complete the task and export to a certain file type:

- "Don't retype it."
- "Easily convert it to Microsoft Word."
- "You're just one step away."
- "Convert your document in 60 seconds or less."

This is an example of **hidden costs**, where free apps will apparently promise functionality to users only to hold it behind a paywall at the last step. Users have already made it this far, so they are less deterred by the paywall than if they had known it going in.

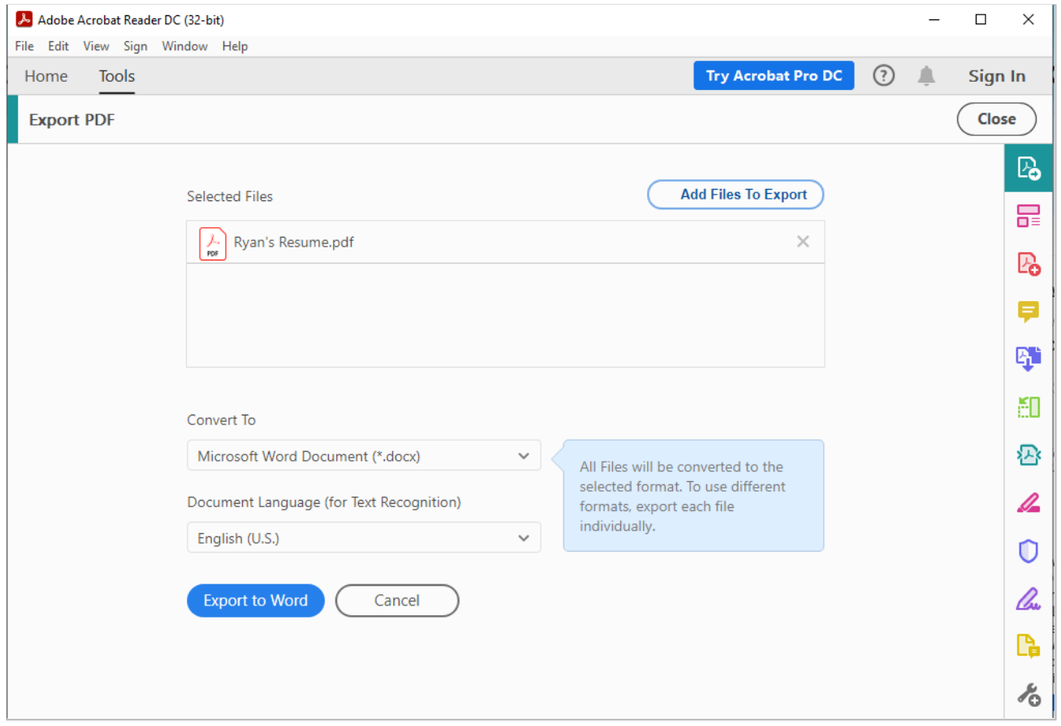


Image above: Adobe Acrobat Reader's customizable options for a feature we haven't paid for yet.

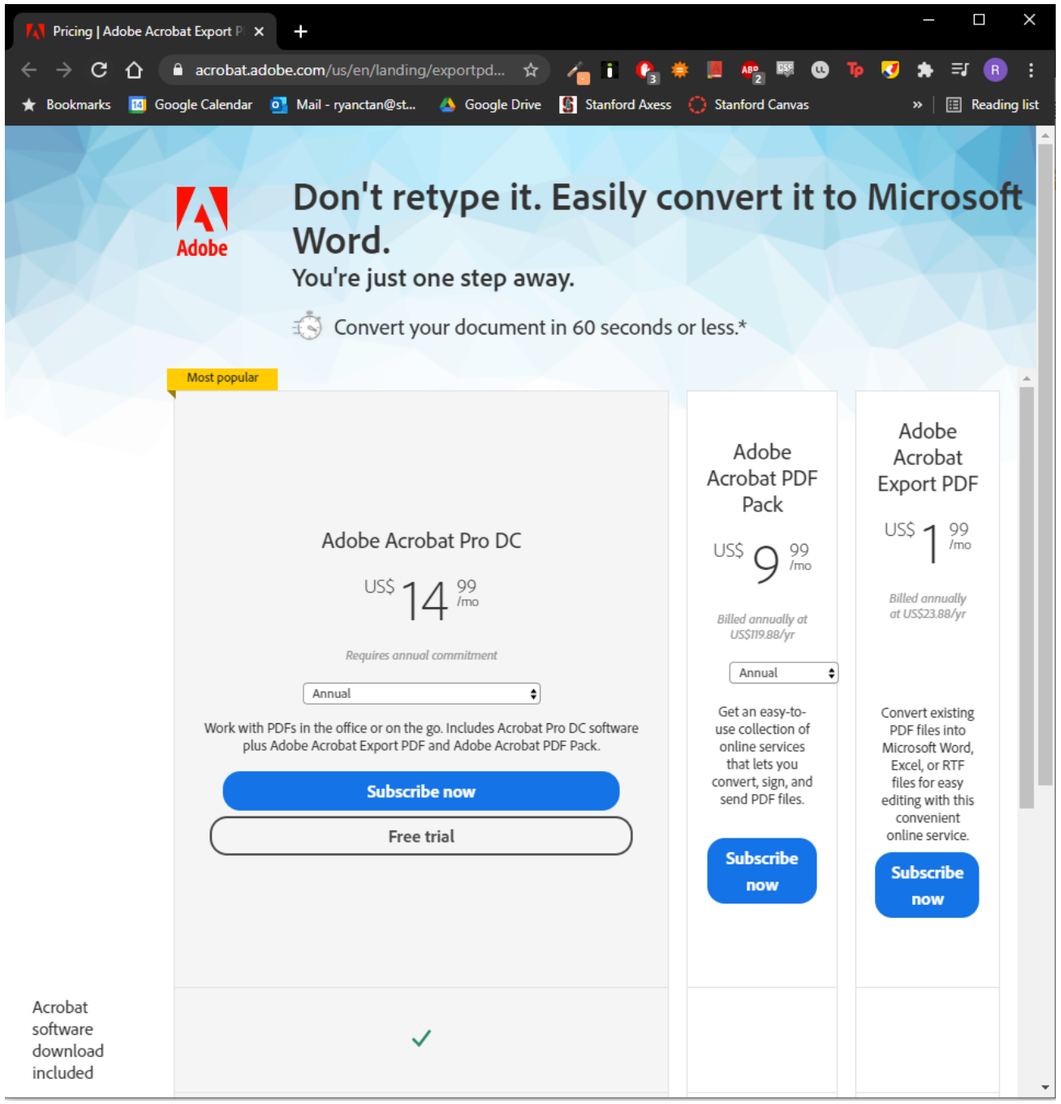


Image above: Acrobat Pro's store page, launched after confirming "Export to Word".

Context: Hidden costs in online retailer services

Unsurprisingly, hidden costs most often emerge in retailer services. The shopping process can be time consuming and tedious, having to select your item, input personal information and confirm payment information, such that by the end you're worn down and more willing to acquiesce to any tacked-on charges.

Interestingly in the Acrobat example, the hidden fees aren't quietly slipped in your bill. The Adobe Store page pitch to switch over to Acrobat Pro is instead sudden and disruptive. Their tactic relies on intentionally placing the fee notice page at the finish line to push you to pay.

Potential Harms: Wasted time & Unclear choices

The human-facing harm of hidden costs is in how the organization misdirects users into making a transaction. In this case, the user does not actively have clear choices to make. And what's especially egregious about this example is how blatantly they point out your wasted time and energy should you choose not to pay. The language implies that you would be harming your own interests for not "easily" taking that last step, being "one step away" from the finish line!

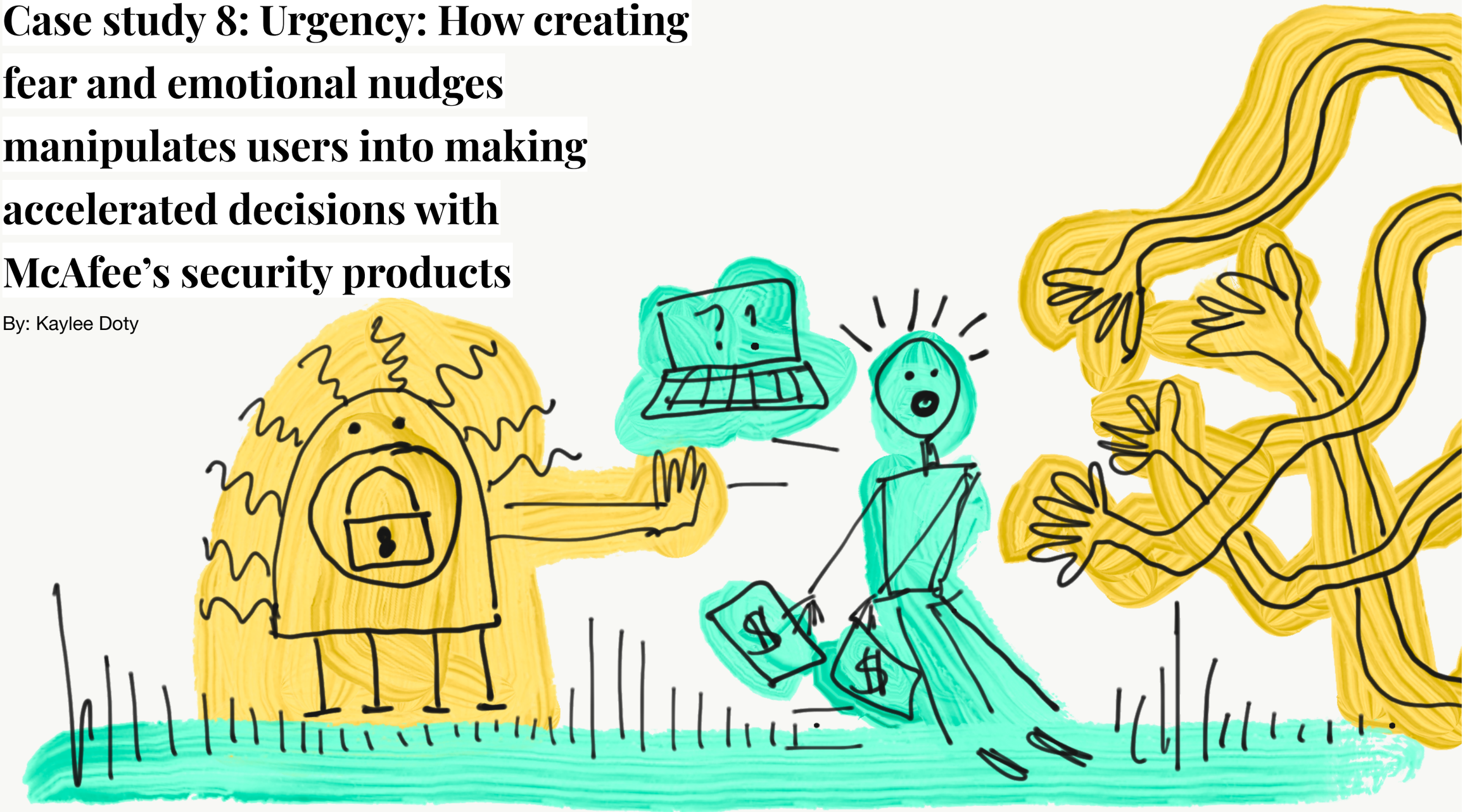
Potential solution: Clearly indicate which features require payment

Free services should clearly indicate what and how much a user can expect to get without paying, and which options need to be paid for. In terms of design implementation, this can be done by notifying users that a feature requires payment before they complete any steps of the process.

In addition, users should be aware of language that is manipulative or nagging, shaming you to take that one additional step when you really don't want to. Look out for messages that take advantage of small commitments, like completing a form, and pressure you into sealing the deal with just one easy click.

Case study 8: Urgency: How creating fear and emotional nudges manipulates users into making accelerated decisions with McAfee's security products

By: Kaylee Doty



To keep our information safe it's important to watch for potential viruses and malware as we use technology . Many people turn to antivirus software that routinely runs safety checks to protect their devices against such threats. But who would've thought that some of these antivirus companies would be employing dark patterns?

Dark Pattern: Urgency: The dark pattern **Urgency**, which uses fear, fake scarcity, and other emotional nudges, to manipulate users into making accelerated decisions. In practice, imagine you're browsing the web on your computer when a window unexpectedly pops up from McAfee's Antivirus software, prompting a security scan of your device. As the test runs you see green check marks populating next to different security categories, indicating that there are no security issues on your device. However, the next time you glance over at your computer, you see a flash of red on the screen (pictured below). "How did I get a virus?!" "Is my information safe?" You take a closer look at the text next to the red icon that reads "McAfee Web Protection: OFF." There was no threat after all, McAfee was simply advertising one of their other products.

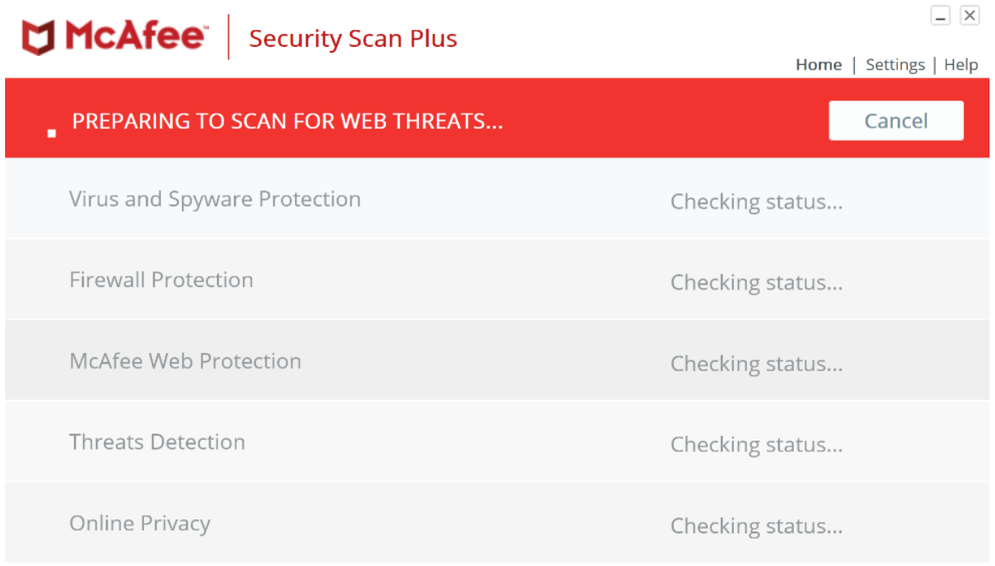


Image above: Screen of McAfee security scan before results populate.

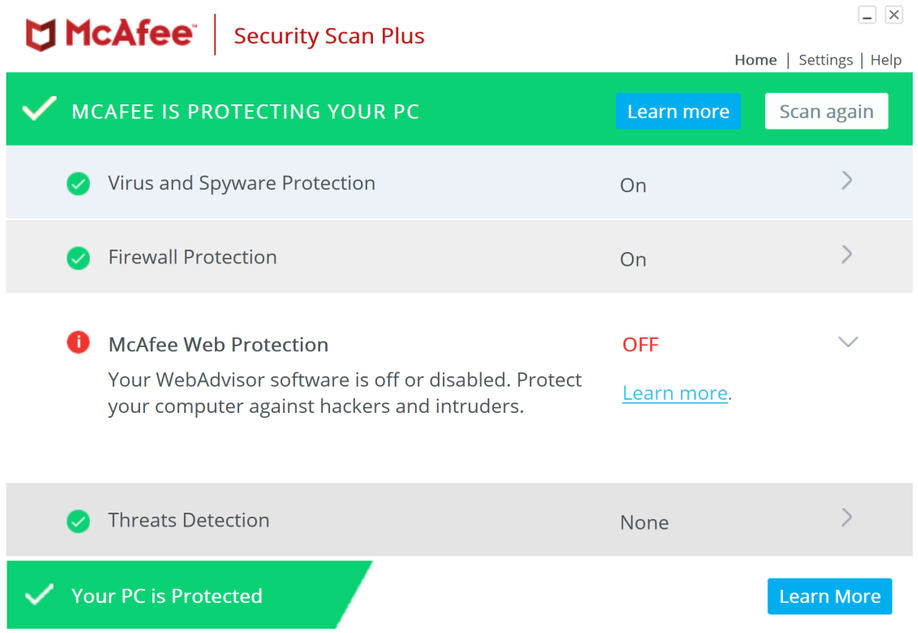


Image above: Completed security scan with McAfee AntiVirus software.

Potential harms: Tricking users & causing alarm

Understanding that people in western countries like the United States tend to associate green with 'good' and red with 'bad' or 'harm,' McAfee strategically uses these colors to mislead users into thinking that a threat has been found on their computer. This creates a false sense of urgency for some users to protect their computer and download potentially unwanted software, without carefully considering the options. McAfee's design is reminiscent of scareware, a scamming tactic that tricks users into downloading harmful software. While McAfee is a legitimate and well-known antivirus company, the tactics they employ are similar—but less harmful to users, which is why this is an example of urgency, not scareware.

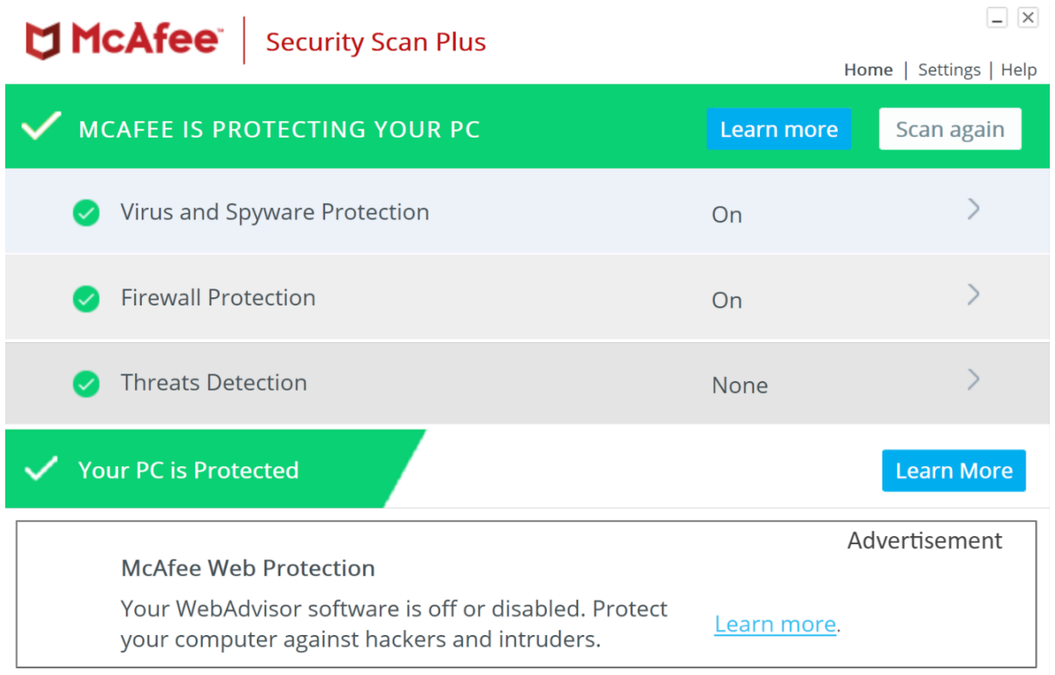
Is it inherently malicious for McAfee to encourage customers to protect their devices with antivirus software? Protecting against threats is generally a helpful service. However, the tactics companies use to nudge customers towards downloading this software potentially presents an emotional harm to users. It

may bring about unnecessary stress and panic. When using antivirus software, people generally want reassurance that their devices are protected, and are on the lookout for anything that can be perceived as a security threat. Considering the color contrast of this design choice, it seems that McAfee wants to draw people's attention towards their WebAdvisor software.[9] For individuals who are not tech-savvy or familiar with how antivirus scans work, seeing a red icon in the middle of a security scan goes beyond merely drawing one's eye to an area on the screen, and could instigate decisions that may cause loss of money or time. Should companies be allowed to emotionally manipulate or nudge their customers into trying new products?

In addition to emotional manipulation, this design choice could accelerate the decision-making process for customers to download the WebAdvisor software, without considering the potential risks of doing so. While this software is actually free, upon downloading it the user automatically allows McAfee to collect information about the websites they visit and the online searches they make. This design pattern is deceptive and harmful because, without the additional pressure to protect their devices, people may not normally consent to giving away all of their search data.

Potential solution: Improve the design to reduce unnecessary urgency or panic

McAfee could change the color of the text advertising the WebAdvisor software to something more neutral like gray. This would lessen the negative emotional response we may have to the color red used next to the WebAdvisor advertisement. Additionally, this ad could be clearly labeled as such and sectioned off from the other security checks to avoid confusion over what is a security threat and what is not. A example model of the redesigned McAfee window is included in the *image below: Re-designed version of McAfee Security Scan interface*



[9] Delwin T. Lindsey, Angela M. Brown, et al. 2010. 'Color Channels, Not Color Appearance or Color Categories, Guide Visual Search for Desaturated Color Targets'. In *Psychological Science* 21(9):1208-1214.

Case study 9: Hidden cancellation fees: How dense disclaimer text on restaurant websites can create an obstruction of choice on Yelp

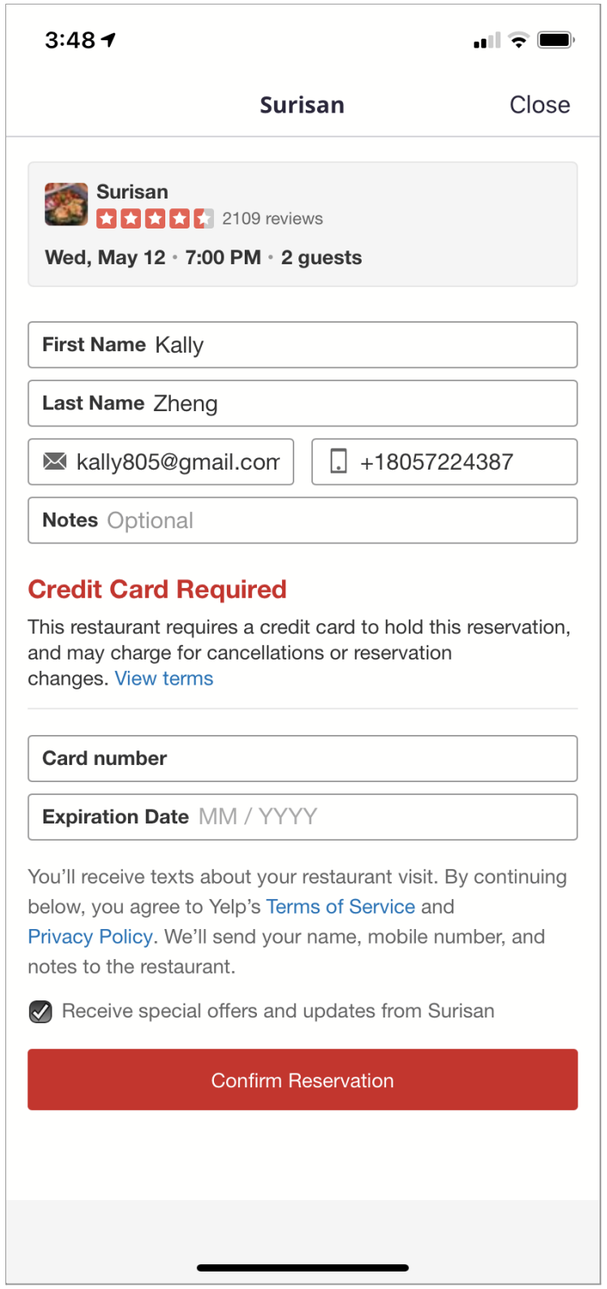
By: Kally Zheng

Context: Holding restaurant patrons accountable
Restaurants have adopted the practice of credit card holds for reservations and cancellation fees. In an interview with the *Washington Post*, the general manager of well-known restaurant Emelie, Elizabeth Schetter explained that restaurants are enforcing this new policy because some customers were making multiple reservations at the same restaurant then forgetting to cancel the ones they weren't using. As a result, many other guests are unable to find availability and restaurants experience a financial loss, as reserved seats are left unfilled. According to Yelp, “About 25% of restaurants require credit card (CC) hold, but some restaurants only require CC holds at peak times and days of the week” to create a reservation.

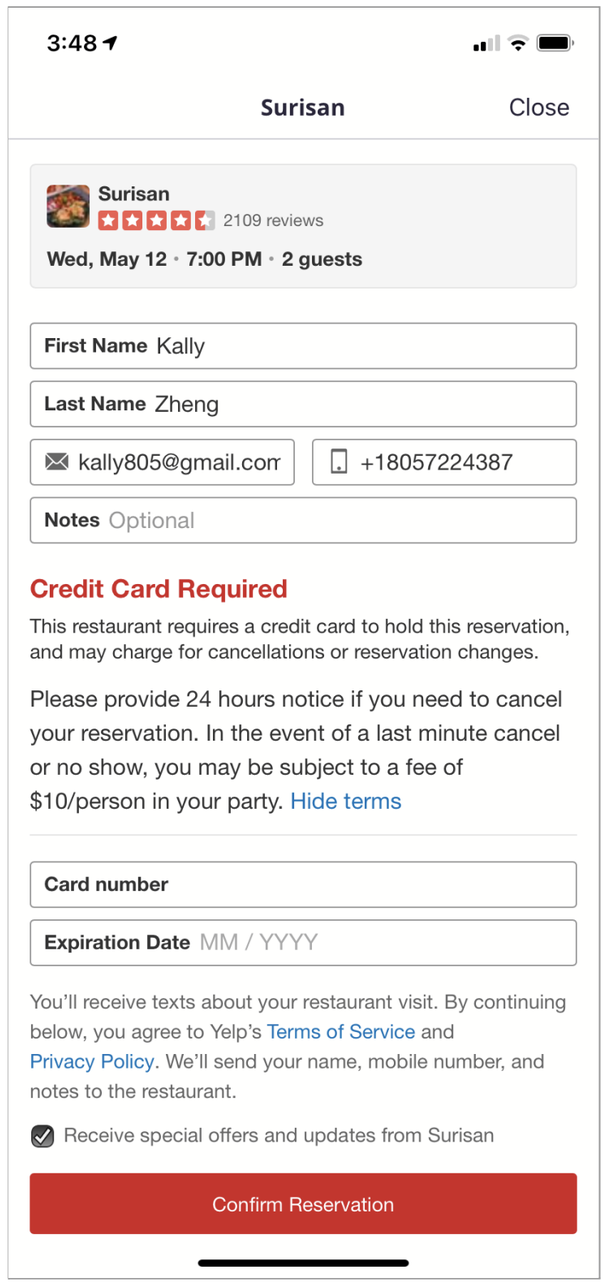
Dark Pattern: Dense Disclaimer
When a user makes a reservation by phone, the restaurant explains the reservation and cancellation policy. A restaurant employee will state the amount of the cancellation fee and obtain verbal confirmation of agreement to the terms before recording the credit card information of the client. In contrast, the Yelp reservations product design conceals the terms and conditions. The reservations page indicates when a credit card is required but does not clearly state the cancellation fee or policy around cancellation fees. It only states that the restaurant may enforce one. The user must select the “View Terms” hyperlink in order to understand that there are penalties to missing a reservation. The interface design makes it easy for users to overlook the terms and enter the credit card information without understanding that they could be charged for forgetting to cancel or cancelling a reservation too late. Even worse, there is no reminder to cancel the reservation. A user could easily make a reservation, enter their credit card information and be charged without understanding why. Informed consent requires that individuals be made aware of consequences. The Yelp reservations product is an example of how design can facilitate an organization obtaining permission without transparency, and violates the requirements of informed consent.

Potential Harm: Unexpected Charges and Lost Money
In this case the most direct harms are unexpected charges and lost money. An *Eater* article by Ryan Sutton reports significant cancellation and no-show charges for restaurants in major cities, “Per Se in New York charges no shows $175 per person while Masa levies a $200 fine. And if you think that's tough, Meadowood in Napa Valley reserves the right to charge truant diners who booked the chef's counter $500 apiece.” In general these policies require a 24 hour notice to avoid the cancellation charge. However some restaurants even extend this to require 3-or even 7-days notice. *SFGate* reports that restaurants such as Saison have implemented a 7-day cancellation policy with a $248 penalty for missing this deadline. Some restaurants such as French Laundry in Yountville, California charge $100 per person for cancellation. This is a significant amount of money to lose especially when the terms and conditions of the agreement were not clear.

Potential Solution: Ensure users are clearly notified of data processing and usage: The GDPR mandates that for consent to be informed and specific, the data subject must be clearly notified what kind of data will be processed, how it will be used, and the purpose of the processing operations. In the case of credit card reservations this means that the user must be clearly notified of the cancellation policy and how their credit card information could be used. And according to recent legal cases, by hyperlinking instead of using express text, an organization might not be creating a contract with users. In fact, in *Compass iTech, LLC v. eVestment Alliance, LLC* (2016), the court concluded that the text hidden in the nested hyperlink was not part of the agreement that users signed off on.



The Yelp reservation credit card hold policy should not be concealed in a nested hyperlink (image left).



Instead it should be displayed clearly. Yelp can inform their users better by displaying the terms and conditions without hyper linking them and give users a reminder to cancel the reservation before the deadline is met. In addition, the cancellation policy language should be simple and easy to understand. The design could be as simple as only having the second expanded policy screen below (image left)

Part 4: Looking for the Light

Based on the case studies, the authors outline a variety of solutions we highlight and list below:

Category 1: Improve product or service design

Design app features (buttons, options, product incentives) with the users' best interests in mind
Case Study 2 details how an online game incentivizes users to buy in-app purchases to advance their position in the game. One suggestion is to lessen the punishments for losing levels and virtual in-game currency and not push online sales to advance in the levels. This would directly impact an organization's potential profits.

Limit the use of design elements that provoke additional urgency or panic
Case Study 8 illustrates that instead of using panic-inducing color palettes like red, which can provoke unnecessary negative emotional responses, more neutral colors like gray or blue. Ads should be clearly labeled so that the user can better assess their purchasing options.

Instead of disguising advertisements, make them more clearly ads
Case Study 1 highlights how ads are embedded in the core experience of the dating app, making it more likely for the user to accidentally engage with the ad. This should be changed so that the user demonstrates some affirmative interest in engaging with the ad, like tapping on it, instead of using the app's core user motions.

Category 2: Policy or regulatory interventions

Empower state and federal regulators and policymakers
Case Study 4 and Case Study 6 outline the role of the U.S. Federal Trade Commission in enforcing rules on "unfair and deceptive marketing practices" and influencer disclosure regulations. The Agency has settled cases against the creator of apps that were illegally collecting children's personal data asserting that the developer violated COPPA for failing to inform the kids' parents and obtain consent for their targeted advertising.

Explore and improve legislation that regulates manipulative patterns
Case Study 3 outlines Josh Hawley's SMART Act, which bans infinite scroll and automatic content refill as well as a requirement for social media to include "natural stopping points" which end consumption after a certain amount of content. But bans like this require more research because determining a specific threshold for infinite is complex.

Another possible route for regulation may be found with *the Canada Anti-Spam Legislation,* which prohibits businesses from automatically opting customers into email subscriptions. Similar legislation aimed at design features of dark patterns could prohibit organizations from automatically opting users into infinite addictive content generation.

The EU's GDPR also requires an organization to get consent before connecting a user service, defining consent as "freely given, specific, informed and unambiguous" and given by a "clear affirmative action." It is not acceptable to assign consent through the user's silence. Therefore, when users reach a certain point on their screens they should be forced to make an active decision.

Category 3: Create user empowerment and product clarity

Inform people about deceptive technology, but enforce transparency requirements t.
Case Study 7 the author outlines how users should be aware of general language that is manipulative or nagging. The onus should not solely be on the users to recognize complex design patterns, the organization must make user options clear and can remind users where they can make decisions.

Clarify the implications of in-app decisions
Case Study 7 outlines why free services should clearly indicate the boundaries of free content. This can be done by clear notification to users that a feature requires payment before they complete any steps of the process.

Ensure users are clearly notified of data processing and usage
Case Study 9 details the GDPR mandate that user consent be informed and specific, and that the data subject must be clearly notified what kind of data will be processed, how it will be used, and the purpose for processing. In the case of credit card reservations this means that the user must be clearly notified of the cancellation policy and how their credit card information could be used. The Yelp reservation credit card hold policy in the case study should not be concealed in a nested hyperlink. Instead it should be displayed clearly. Yelp can inform their users better by displaying the terms and conditions without hyperlinking them, and by reminding users to cancel the reservation before the deadline. Cancellation policy language should be simple and easily understandable.

Part 5: Hit the Switch!

A core aim of this project is to **make deceptive patterns more tangible and understandable to a broader, less-tech savvy audience**. The dark patterns highlighted above show the diversity of the types of tactics, harms, shapes and sizes they could manifest in products and services. We believe there is an opportunity for continued investigation into more types of design patterns and negative outcomes across industries.

More work to highlight adverse harms, especially for overburdened, underserved communities. Researchers, advocates, and technologists must be intentional to focus on harms that often adversely impact marginalized and vulnerable communities including communities of color, the elderly, the disable, and those with limited understanding of English. While educating individuals and communities about the risks of being deceived by dark patterns, the burden of avoiding dark patterns should not be on people. Dark patterns can be seen as a symptom of the larger problem of the power imbalance between organizations including corporations, nonprofits, and civil society organizations and individuals, which makes it all but impossible for people to exercise total control over their personal data. It also makes the use of consent a terrible barometer for whether data collection, use, and access can happen. Instead, impacts should be at the central consideration for policymakers and regulators who must act to protect the vulnerable from these kinds of designs and activities.

Foundations for future classes focused on deceptive patterns in technology. On a meta-note, we piloted a series of small discussion sections for a course that focused on design, policy and technology. We believe that this could be a model for future classes or seminars at various universities, and focused on better outlining methods that can help researchers do investigative work at scale to prove that harms are happening across platforms. The discussion sections and seminars could be particularly interesting for students and instructors interested in public interest technology, as they focused on considering risk, impact, and solutions. These are areas of considerable importance in this era of rapid innovation and deployment of technology without full investigation into the human-costs of letting systems loose in the wild.

Use data and research for impact. There are publicly available databases that show evidence of potential harms that digital products and services that have been used by people have caused throughout the decades. It is critical for researchers to use these existing resources as a way to have baseline knowledge or understanding.

- The Dark Patterns Tip Line
- The Consumer Financial Protection Bureau's Consumer Complaint Database
- The Federal Trade Center's Consumer Sentinel
- The Federal Communications Commission's Consumer Complaint Data Center
- Princeton's Center for Internet Tech Policy's Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites
- Darkpatterns.org's curated Airtable list of dark patterns
- #DarkPatterns hashtag feed on Twitter
- Reddit's r/asshole design

Researchers should work with regulatory bodies like federal regulatory agencies and state attorney general offices to focus on the patterns of behaviors that organizations are demonstrating, and the kinds of patterns being reported. Policymakers should use these as evidence of the need for stronger policies to protect individuals. These sources can also be used to forecast and prevent the next kinds of deceptive practices organizations may invent.

Part 6: About the team

Co-Authors & Editors

- **Stephanie T. Nguyen** is a research scientist and designer focused on HCI + UX design, policy and technologies that impact underserved and overburdened populations. She is a public interest technologist at the U.S. Digital Service under the Biden Administration helping to reunify migrant unaccompanied children at the southern border. She is currently on the Scholarly Council for UCLA's Center for Critical Internet Inquiry and was previously at Consumer Reports, MIT Media Lab and the Obama White House serving on projects across CMS, Department of Education and the State Department. She holds a MPP at Harvard Kennedy School with a Bachelor of Science degree in Digital Media Theory & Design at the University of Virginia. https://www.stephanienguyen.co/
- **Jasmine E. McNealy** is an associate professor in the Department of Media Production, Management & Technology, in the College of Journalism and Communications at the University of Florida, where she studies information, communication, and technology with a view toward influencing law and policy. She is also the Associate Director of the Marion B. Brechner First Amendment Project and a Faculty Associate at the Berkman Klein Center for Internet & Society at Harvard University. https://jasminemcnealy.com

Student Collaborator Bios

- **Kaylee Doty** is a Computer Science major passionate about the intersections of tech in music and linguistics. She will complete her B.S. in 2024.
- **Ryan Christopher Tan** is a Computer Science major interested in digital humanities and game design. He will complete his B.S. in 2022.
- **Kally Zheng** is pursuing a masters in Computer Science with a concentration in Human Computer Interaction. She will be completing her M.S in 2021.

Faculty, Staff, Mentors

- **Professor Lucy Bernholz** is a Senior Research Scholar at Stanford University's Center on Philanthropy and Civil Society and Director of the Digital Civil Society Lab. She has been a Visiting Scholar at The David and Lucile Packard Foundation, and a Fellow at the Rockefeller Foundation's Bellagio Center, the Hybrid Reality Institute, and the New America Foundation. She is a co-editor of Philanthropy in Democratic Societies (2016, Chicago University Press) and the volume Digital Technology and Democratic Theory (2021, University of Chicago Press). She writes extensively on philanthropy, technology, and policy on her award winning blog, http://philanthropy2173.com.
- **Ashley Lee** is a postdoctoral research fellow in the Digital Civil Society Lab at Stanford University. Her research focuses on technology, social movements, and surveillance in comparative perspective. She completed a PhD in Culture, Communities, and Education at Harvard University as a Weatherhead Center for International Affairs Graduate Fellow. She holds a BS in Computer Science from Stanford University.
- **Professor Jeff Ullman** is the Stanford W. Ascherman Professor of Computer Science (Emeritus). His interests include database theory, database integration, data mining, and education using the information infrastructure.

Context & Opportunity: The Dark Pattern Zine

In partnership with UCLA's Center for Critical Internet Inquiry & Stanford's Digital Civil Society Lab, we want to bring together a community of students, fellows, mentors and faculty to compile a series of case study examples to expand our current definition of the most harmful deceptive patterns in industries beyond social media and retail companies. From discrimination in credit score reporting to contract worker fairness in food delivery apps, scholars, researchers and consumer advocates investigate and study how technology impacts society in

many ways. In partnership with UCLA's Center for Critical Internet Inquiry and Stanford's Computer Science Course, "Bridging Policy and Technology Through Design," we created this zine that explores the dark patterns in connected products and services that impact overburdened, underserved populations. We compiled these examples by engaging with a community of researchers and practitioners to outline, illustrate and explain the dark patterns through case studies in a simple, down to earth format.

Goal: Our goal with this work is to better highlight, educate, and engage practitioners and researchers on these patterns in industries ranging from financial services to smart home devices to children's related technology. We also wanted to pilot an academic seminar that would discuss deceptive designs, centered around a tangible output where students could creatively apply their research, technical, design, and policy skills toward a shareable deliverable. The zine will be hosted online and available for download to the public.

Many thanks: Thank you to graduate and undergraduate students, Kaylee Doty, Ryan Christopher Tan and Kally Zheng who worked with us to research and create their own dark patterns. Thank you to Lucy Bernholz for the opportunity to mentor and work with students on this project and to UCLA's Center for Internet Inquiry, Vanessa Rhinesmith and Dr. Stacy Wood, for supporting the vision of this work for the semester.


End