

Cornell DLI Manipulation Workshop, May 2023

1. My talk

1.1. Themes to hit

- "Design"
 - Building on the shoulders of giants
 - STS
 - Values-in-Design
 - As we so often have, legal scholars can serve as a last-mile force, turning these insights into action.
 - Also the legacy of "code is law"
 - From the article: For too long, legal scholars have focused solely on the descriptive version of this maxim: code operates like law, so we ought to study the way code constrains human behavior alongside law. Our study of friction deploys a more proactive, less studied, prescriptive version of the maxim: We should focus on enacting laws that nudge, shape, harness, or ban particular forms of code to combat false news and other information age harms.
- Participatory design
 - Stage 1 of agenda: what industry has done "self-regulatory"
 - Stage 2 of agenda: how, why, and when we can pull these levers
 - Deep focus on why the cloistered priesthood model is illegitimate
 - These tend to be people who
 - aren't diverse
 - are terrible people
 - are motivated solely by economic efficiency
 - Given the complexity and increasing importance of these opaque systems, we must demand a seat at the design table.
- Interventions to date that have tried to deal with manipulative design
 - Dark patterns
 - Naming and shaming
 - Regulatory prohibitions
 - Twitter's "are you sure you want to post that?"
 - 40% of those presented with the prompt click the link
 - Some (percent unstated) users decline to post

- WhatsApp Frequently Forwarded
 - 80% of images appear for two days or less
 - 60% of those for less than 1000 minutes
 - Conspiracy theories have been spread
 - India: 225M use WA
 - May 2017, 7 men beaten to death by crowds falsely believing them to be child and organ traffickers
 - 2018, at least 24 other people killed in similar circumstances
 - Calls to reform
 - July 2018, the Indian government called on WA to assume "accountability and responsibility"
 - July 19: WA announced changes to its forwarding policy.
 - Similar time period: Brazil and its presidential election
 - One candidate: \$3M for disinfo campaign on WA
 - Later: same in elections in Nigeria and India—"WhatsApp Elections"
 - COVID misinfo lead to more restrictions
 - Key claim: lowered spread of viral messaging by 70%.
- Others?
 - Netflix auto-play example
 - Twitter detecting "hateful" speech and warning users before they can reply.
 - 34% of posts were not posted after prompt
 - Instagram's "Take a Break" feature
 - TikTok did something similar
 - In Josh Hawley's SMART Act
 - Social Media Addiction Reduction Technology (SMART) Act
- What do we hold constant, untouchable products of permissionless innovation?
 - Grover Norquist's despicable quote, rehabilitated!
 - Smaller social networks
 - Internet Balkanization

1.2. Specific pieces of legislation

- Almost all of this from this very good summary by FPF: <https://fpf.org/blog/the-future-of-manipulative-design-regulation/>
- Their framework:
 - It all starts with the DETOUR Act

- Then three types:
 - Those that focus on dark patterns and consent
 - CCPA, CPA, CT
 - ADPPA
 - EU DSA
 - Those that go beyond consent
 - Heavy focus on the rights of vulnerable and marginalized peoples
 - California Age Appropriate Design Code (CAADC)
 - Those focused on addiction and other narrower forms of pernicious behavior

1.2.1. Deceptive Experiences To Online Users Reduction Act (DETOUR) Act

- introduced in Congress in 2018 by Senators Warner (D-VA) and Fischer (R-NE)
 - but never passed into law.
- The DETOUR Act would forbid websites, platforms, and services from acting
 - “to design, modify, or manipulate a user interface
 - with the purpose or substantial effect of
 - obscuring,
 - subverting, or
 - impairing
 - <object>
 - user autonomy,
 - decision-making,
 - or choice
 - to obtain
 - consent or
 - user data.
 - This is omitted from CCPA, ADPPA, CPA, etc.

1.2.2. California Age Appropriate Design Code (CA AADC)

- prohibits services targeted at young people from using “dark patterns”
 - to steer youth into sharing personal data or
 - to act in a way that a “business knows, or has reason to know, is materially detrimental to” a child’s physical or mental health or “well-being.”
- could be construed as applying to design features far beyond consent,

- algorithmically-selected content,
- music and video feeds, and
- other core features of child-directed services.

1.2.3. Social Media Addiction Reduction Technology (SMART) Act

- would outlaw auto-refreshing content, music, and video feeds without “natural stopping points,” and

1.2.4. S.1629

- regulate the use of loot boxes, which can encourage addictive, gambling-like behavior in gaming.

1.3. Probably Not?

- Governance seams
- Consent as:
 - Voting
 - Labor organizing

Author: Paul Ohm

Created: 2023-05-12 Fri 07:39

[Validate](#)