

Ethics, Fairness, Responsibility, and Privacy in Data Science

CMSC 25900 (“CS 259”)

Spring 2021

The University of Chicago



**THE UNIVERSITY OF
CHICAGO**

~~Ethics, Fairness,
Responsibility, and Privacy in
Data Science~~

**Engineering for Ethics,
Privacy, and Fairness in
Computer Systems**

CMSC 25900 (“CS 259”)

Spring 2021

The University of Chicago



**THE UNIVERSITY OF
CHICAGO**

Course Overview

Course Description

This course takes a **technical approach to understanding ethical issues** in the design and implementation of **computer systems**. Tensions often arise between a computer system's utility and its privacy-invasiveness, between its robustness and its flexibility, and between its ability to leverage existing data and existing data's tendency to encode biases. The course will demonstrate how computer systems can violate individuals' privacy and agency, impact sub-populations in disparate ways, and harm both society and the environment. It will also introduce algorithmic approaches to fairness, privacy, transparency, and explainability in machine learning systems. Through hands-on programming assignments and projects, students will design and implement computer systems that reflect both ethics and privacy by design. They will also wrestle with fundamental questions about who bears responsibility for a system's shortcomings, how to balance different stakeholders' goals, and what societal values computer systems should embed.

Instructors



Blase Ur

blase@uchicago.edu

CMSC 259 instructor



Raul Castro Fernandez

raulcf@uchicago.edu

DATA 259 instructor

Instructors



Blase Ur

blase@uchicago.edu

CMSC 259 instructor

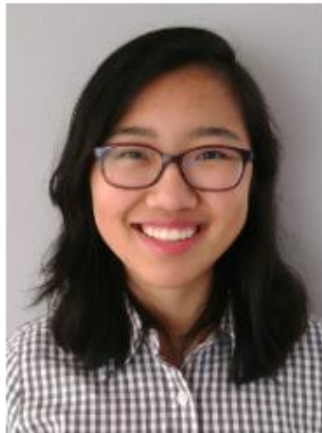


Raul Castro Fernandez

raulcf@uchicago.edu

DATA 259 instructor

Course Staff



Valerie Zhao

Primary TA



Ahmad Bamba

Course Staff



Annika Hildebrandt

Course Staff

CMSC 25900 vs. DATA 25900

CMSC 25900 (Blase): <https://classes.cs.uchicago.edu/archive/2021/spring/25900-1/>

Focus on computer systems / algorithms / engineering / implementation

9 programming assignments; 9 reading responses

Multiple languages, including Python, JavaScript, HTML, MySQL

CMSC 25900 vs. DATA 25900

CMSC 25900 (Blase): <https://classes.cs.uchicago.edu/archive/2021/spring/25900-1/>

Focus on computer systems / algorithms / engineering / implementation

9 programming assignments; 9 reading responses

Multiple languages, including Python, JavaScript, HTML, MySQL

DATA 25900 (Raul): <http://raulcastrofernandez.com/DATA-25900-Spring21/>

Focus on data science / data analysis

4 programming assignments; 6 reading responses; project; report; quizzes

Python using Jupyter Notebooks

BOTH: 50% shared lectures; close coordination

Lectures

- 2:40p – 4:00p CT on Tuesdays and Thursdays
 - **DIFFERENT ZOOM LINK FOR SHARED LECTURES**
- Synchronous to enable questions and discussion
- Feel encouraged to ask questions / respond
 - Option 1: Unmute yourself and interrupt the lecturer
 - Option 2: Message everyone on Zoom
 - Option 3: Message Blaise on Zoom
- Lectures will **not** be recorded
 - Why? To enable comfortable discussion
 - Do not record lectures or take screenshots
 - Reality: We can't enforce that no one will record / screen shot lecture

Lecture Etiquette

- We encourage you to have your video on
 - No worries at all if you prefer not to
 - Be aware of your surroundings
- Engage and be respectful as you would in person
- You can choose to change the name that shows in Zoom
 - We encourage, but don't require, you to add your preferred pronouns

Topics Covered

- User Interfaces (1 lecture; today!)
- Data (3 lectures)
- World (2 lectures)
- Machine Learning (4 lectures)
- Privacy (4 lectures)
- Lifecycles (1 lecture)
- Tracking & Surveillance (2 lectures)
- Politicized Ecosystems (1 lecture)

Course Requirements and Grading

- **9 programming assignments:** 76.5% total (8.5% each)
 - Due **Thursdays 11:59p**
 - **Synthesis** tasks are graded on **correctness and ethical practices**
 - **Exploration** tasks are graded only on **correctness**
- **9 reading responses:** 13.5% total (1.5% each)
 - Due **Mondays 11:59p** (except for the last one)
- Participation: 10%
- Work can be submitted 24 hours late with a 15 point penalty
 - Extensions for health issues, personal emergencies, wellness concerns
- Special P/F grading option for AY 2020-2021 (see syllabus)

Academic Integrity Policy

- See the syllabus for the full policy
- You may discuss high-level approaches or strategies with classmates. You may not share or show each other any code.
- At the top of each assignment write-up, list:
 - Other students with whom you discussed the assignment
 - Resources (e.g., URLs) that influenced your general solution
- You may reuse small amounts of code verbatim (e.g., from online tutorials). Rule of thumb: **< 5 lines from a source is ok**
 - You must **document** (in place, as a **comment**) the source URL
- We will pursue violations of academic integrity with the dean

Communication

- Syllabus and schedule (lectures, work due):
<https://classes.cs.uchicago.edu/archive/2021/spring/25900-1>
- Campuswire for discussion / questions / clarifications
 - **Blase will sign you up tonight**; don't ask for a course code
 - General questions about assignments / topics should be **public** as long as they don't give away key hints about the assignment
 - Questions with code / key hints should be **private**
 - **Do not email the course staff; post privately on Campuswire**
 - Use appropriate tags on Campuswire to:
 - Request an extension
 - Request a P/F grade
 - Post (appropriate) course-related memes

Assignment Submission Process

- Assignments will be distributed on Canvas
- You will submit your **code** on **Canvas**
- You will submit your **write-up (PDF)** on **Gradescope**
- Remember to list discussions / sources at the top of the PDF
- If a request for an extension was approved, list the Campuswire post number (e.g., “#154”) at the top of the PDF
- Feedback / rubric will be posted on Gradescope
- Submit regrade requests on Gradescope
- Numerical grades (but not feedback) will be synced to Canvas

Office Hours

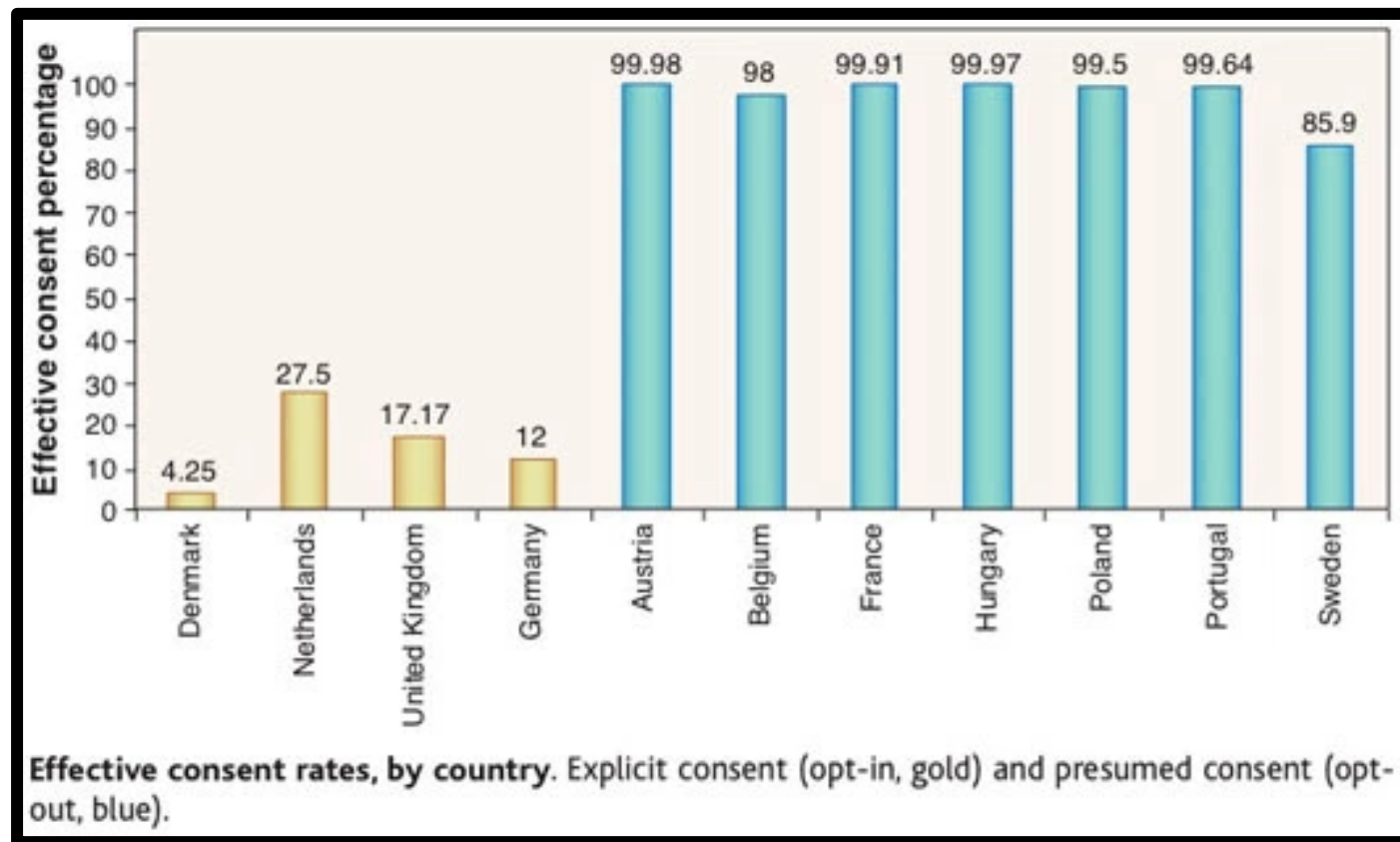
- All on the CS-only Zoom link
- Mondays 10:00a-11:00a (Valerie)
- Mondays 4:00p-5:00p (Blase)
- Tuesdays 4:30p-5:30p (Blase)
- Wednesdays 5:00p-6:00p (Valerie)

Responsible and Respectful User Interfaces

***Should a product have
an opinion?***

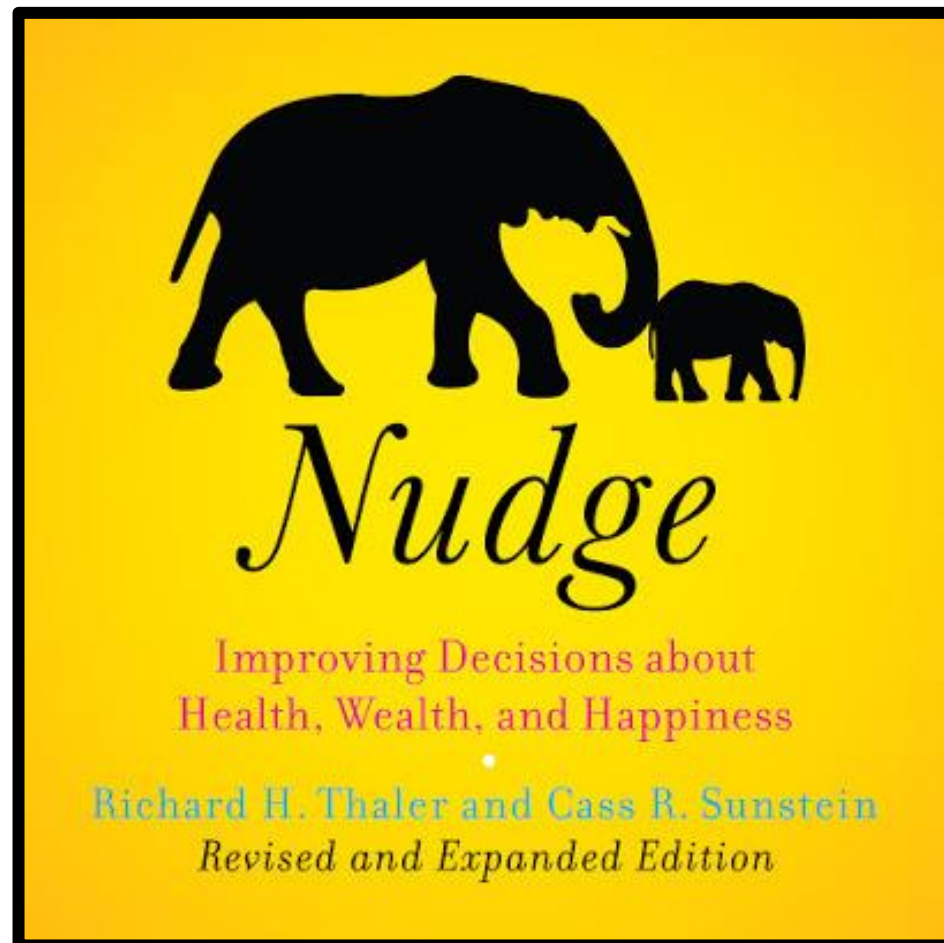
The power of defaults

The power of the default (organ donation)



Interfaces that influence

Behavioral nudging



Dark patterns

Taxonomy from www.darkpatterns.org

- Trick questions
- Sneak into basket
- Roach motel
- Privacy Zuckering
- Price comparison prevention
- Misdirection
- Hidden costs
- Bait and switch
- Confirmshaming
- Disguised ads
- Forced continuity
- Friend spam

<https://twitter.com/darkpatterns> has many examples

Examples from www.darkpatterns.org



Steve Bromley
@Steve_Bromley

"No, I don't like savings"

	50% OFF 3 Months
<p>Pro Monthly</p> <p>£11.99/mo</p> <p><i>£143.88 billed annually</i></p>	<p>Pro Monthly Plan + Discount</p> <p>£6.0/mo</p> <p><i>£101.92 billed annually / per license</i></p> <p>Save £ 17.99 per license</p>
<p>No, I don't like savings</p>	<p>Yes, Take Offer</p>



Mar 14, 2021

Examples from www.darkpatterns.org



Benedict Evans ✓

@benedictevans


You could write a book about all the dark patterns Amazon uses to promote Prime.

amazon.co.uk

WELCOME ADDRESS ITEMS WRAP DISPATCH PAY CONFIRM

Benedict Evans, we're giving you immediate access to all Prime benefits with a 30-day FREE trial

Try 30 days of Prime for free. You can cancel anytime.

		Without Prime
Prime Video	Included	Unavailable
Amazon Music	Included	Unavailable
Fast delivery on Prime eligible items	FREE	From £4.99

[Order without Prime. Decline free trial](#)

Order Now with Prime

Pay later

Examples from www.darkpatterns.org



Katie Dill

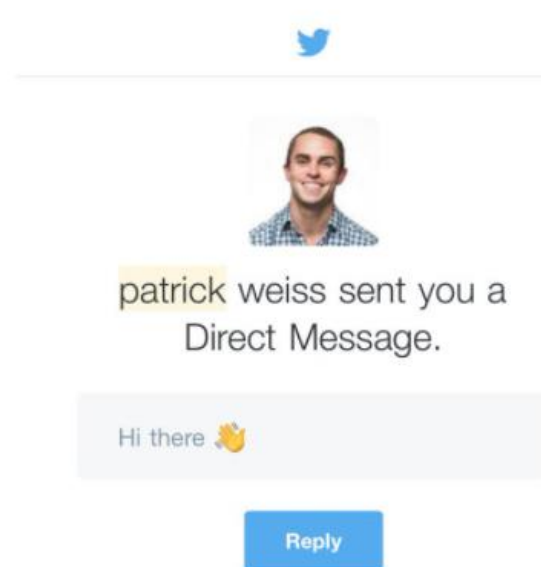
@lil_dill

When engagement metrics drive the decision..

On the left, Twitter's email with the direct message text included.

On the right, LinkedIn's email forcing me to open the app to see the message. Drives me 🍌s every time!

Twitter, I appreciate you for putting the user first on this.



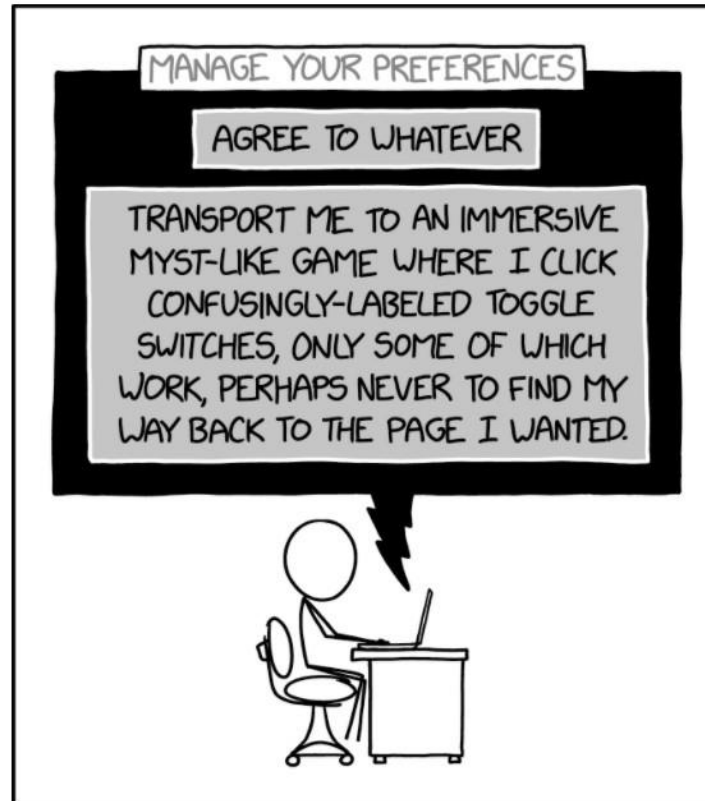
Examples from www.darkpatterns.org

The image shows two examples of dark patterns in user interfaces. On the left is a screenshot of the Christie's website's privacy preference center. It features a white modal window with a dark background. The text explains that cookies are used for personalization and offers an 'Allow all' button. Below this, there are three categories of cookies: 'Functional cookies' (with a toggle switch), 'Strictly necessary cookies' (labeled 'Always active'), and 'Performance cookies' (with a toggle switch). A 'Confirm my choices' button is at the bottom. The modal is powered by OneTrust.

On the right is a tweet from Adam Smith-Kipnis (@AdamSmithKipnis) dated March 18, 2021. The tweet discusses a dark pattern on christies.com where the 'No info' option in the cookie consent preferences does not clearly indicate that tracking is not consented to. The tweet includes the text: '@darkpatterns Cookie consent preferences from @ChristiesInc on christies.com does not say whether grey on grey, or black on white, means approve or deny. No info means tracking isn't informed or consensual. This UI gives the illusion of consent. #darkpatterns'. The tweet has 2 retweets and 11 likes. A reply from Dark Patterns (@da...) is visible at the bottom, replying to @AdamSmithKipnis and @ChristiesInc.

Meta-example from xkcd

MANAGE YOUR PREFERENCES



Are dark patterns bad?

Security “Dark Pattern”



This Connection is Untrusted

You have asked Firefox to connect securely to **grey-dev.ece.cmu.edu**, but we can't confirm that your connection is secure.

Normally, when you try to connect securely, sites will present trusted identification to prove that you are going to the right place. However, this site's identity can't be verified.

What Should I Do?

If you usually connect to this site without problems, this error could mean that someone is trying to impersonate the site, and you shouldn't continue.

Get me out of here!

- ▶ **Technical Details**
- ▶ **I Understand the Risks**

Security “Dark Pattern”

Normally, when you try to connect securely, sites will present trusted identification to prove that you are going to the right place. However, this site's identity can't be verified.

What Should I Do?

If you usually connect to this site without problems, this error could mean that someone is trying to impersonate the site, and you shouldn't continue.

Get me out of here!

▼ Technical Details

grey-dev.ece.cmu.edu uses an invalid security certificate.

The certificate is not trusted because it is self-signed.

(Error code: sec_error_untrusted_issuer)

▼ I Understand the Risks

If you understand what's going on, you can tell Firefox to start trusting this site's identification. **Even if you trust the site, this error could mean that someone is tampering with your connection.**

Don't add an exception unless you know there's a good reason why this site doesn't use trusted identification.

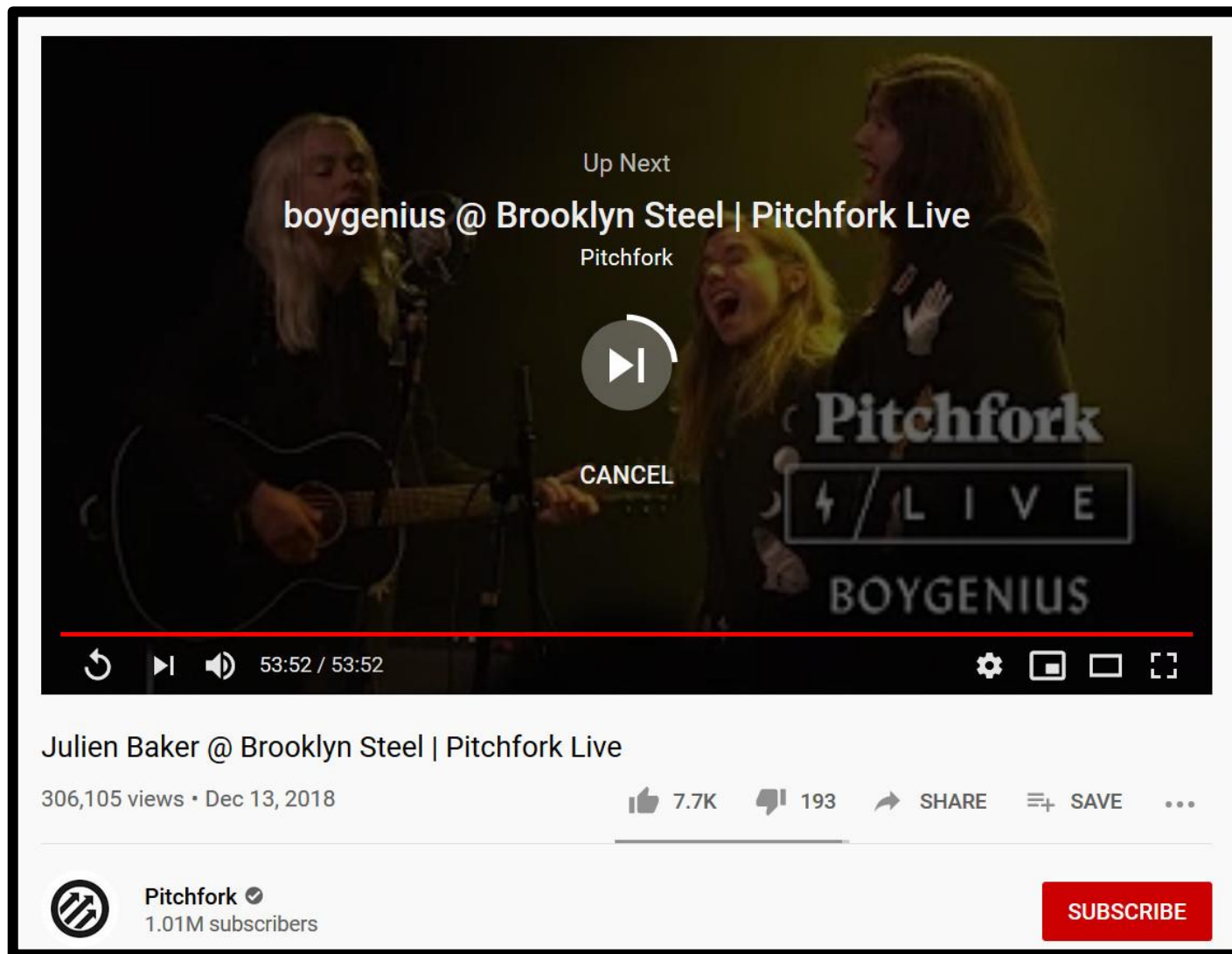
Add Exception...

Interfaces that incentivize behaviors

Ways interfaces incentivize

- Social approval (like counts / view counts / follower counts)
- Karma / leaderboards / Reddit gilding
- Creating a Fear Of Missing Out (FOMO)
- “Only 1 left in stock”
- Forced obsolescence or devices that seem to slow down
- Intentionally showing inaccurate measurements

Does auto-play capture intent? Addiction?



The image shows a YouTube video player interface. The main video is titled "Julien Baker @ Brooklyn Steel | Pitchfork Live" and has 306,105 views from Dec 13, 2018. Below the video, there is a suggestion for the next video: "boygenius @ Brooklyn Steel | Pitchfork Live" by Pitchfork. The suggestion includes a play button icon and the word "CANCEL" below it. The video player controls at the bottom show a progress bar at 53:52 / 53:52, along with icons for refresh, play/pause, volume, settings, and full screen. The Pitchfork channel logo and name are visible at the bottom left, with 1.01M subscribers and a red "SUBSCRIBE" button.

Up Next

boygenius @ Brooklyn Steel | Pitchfork Live

Pitchfork

CANCEL

Julien Baker @ Brooklyn Steel | Pitchfork Live

306,105 views • Dec 13, 2018

7.7K 193 SHARE SAVE ...

Pitchfork 1.01M subscribers

SUBSCRIBE

Data coercion by user interfaces

The screenshot displays the 'Ethics20 Project Preferences' interface. At the top, it shows 'iQ Score: Great' and 'Published' status. The main area is titled 'Default Question Block' and contains three questions:

- Q1:** 'Your name:' with a text input field. It is checked and has a gear icon for settings.
- Q2:** 'Your UChicago email address (e.g., blase@uchicago.edu):' with a text input field. It is unchecked and has a gear icon for settings.
- Q3:** 'How excited or not excited are you to do your class project in each area below?' with a horizontal scale from 'Very excited to work on it' to 'Not excited'.

On the right side, there are configuration options:

- Change Question Type:** A dropdown menu currently set to 'Text Entry'.
- Text Type:** Radio buttons for 'Single Line' (selected), 'Multi Line', 'Essay Text Box', 'Form', and 'Password'.
- Validation Options:** A dropdown menu with 'Force Response' selected and checked. Other options include 'Request Response', 'Minimum Length', 'Maximum Length', 'Character Range', 'Content Validation', and 'Custom Validation'.

A red box highlights the 'Validation Options' dropdown menu.

- Argument for not requiring: agency; privacy; legitimately missing
- Argument for requiring: preventing missing data, stopping laziness 36

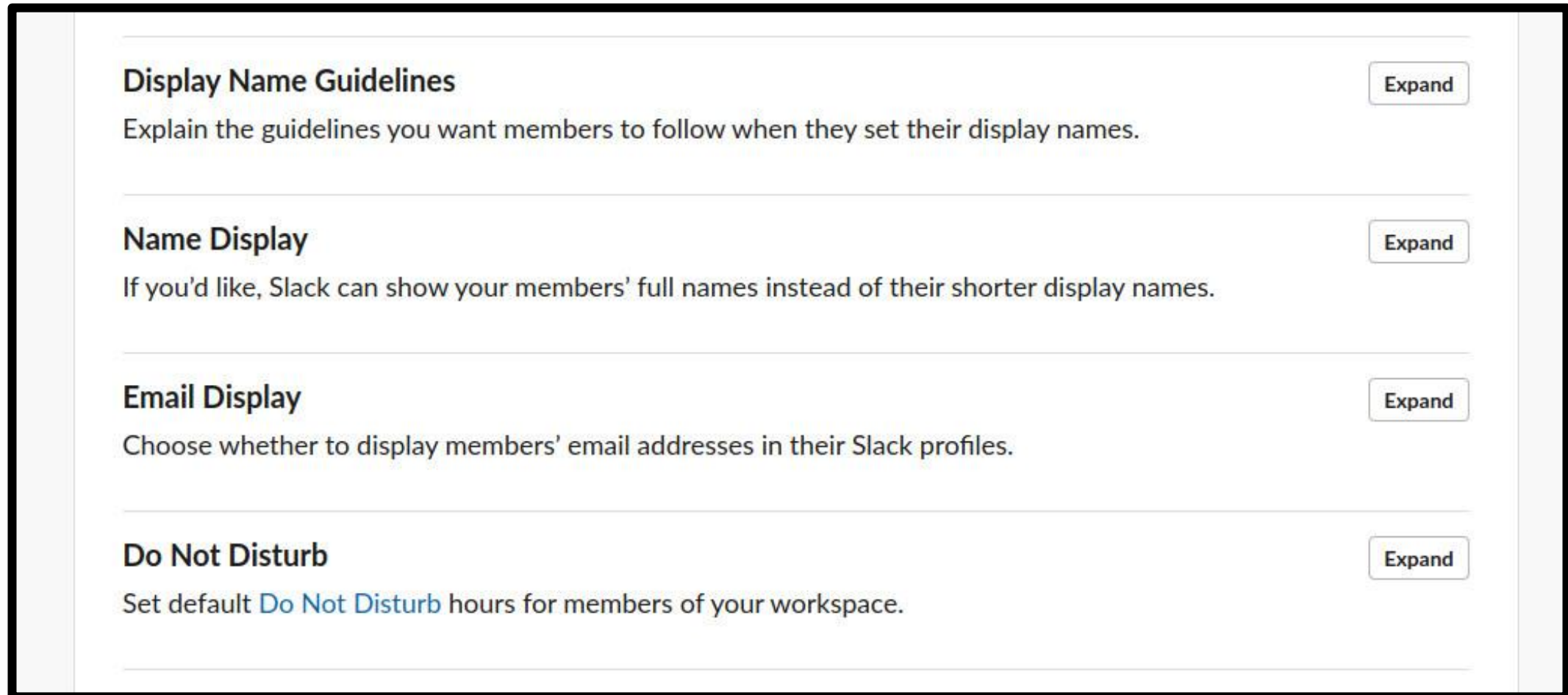
***Can we think about
data-relevant user
interfaces in terms of
power?***

Power in interfaces

- Who chooses what questions are there?
- Who chooses default options / settings / option ordering?
- Who gets to add data?
 - Facebook shadow profiles (for advertising)
 - Tag others or upload photos of others in social media
 - Grubhub, Postmates, and DoorDash add restaurants without permission (see <https://www.eater.com/2020/1/29/21113416/grubhub-seamless-kin-khao-online-delivery-mistake-doordash>)
- Who chooses which data is right?
 - Wikipedia editing wars and controversial articles

Who sets policies?

- Slack organizational policies about names



The screenshot shows a list of four organizational policy settings in Slack, each with an 'Expand' button to its right. The settings are: 'Display Name Guidelines', 'Name Display', 'Email Display', and 'Do Not Disturb'. Each setting includes a brief description of the policy.

Policy Name	Description	Action
Display Name Guidelines	Explain the guidelines you want members to follow when they set their display names.	Expand
Name Display	If you'd like, Slack can show your members' full names instead of their shorter display names.	Expand
Email Display	Choose whether to display members' email addresses in their Slack profiles.	Expand
Do Not Disturb	Set default Do Not Disturb hours for members of your workspace.	Expand

Respecting time and away-from-screen time

Expectations about types of notifications

- Email vs. text vs. in-app notification
- Are notifications *push* or *pull*?
- Does your device make a sound? Does it buzz?
- Is there an unread number on your screen?



Slack pausing notifications

Pause notifications with Do Not Disturb

If you need focus time or time away from work, you can pause your [notifications in Slack](#) with Do Not Disturb (DND). Use DND as needed to help you concentrate, and even set a DND schedule for routine times you'd prefer not to be notified.

How does DND work?

- All notifications and @mentions will be paused. Once you resume notifications, you can review everything you received while in DND.
- Members will see a 🚫 **Do Not Disturb icon** next to your name. When sending a direct message, they can override DND to notify you of urgent messages.

Pause notifications

Desktop

iOS

Android

Turn on DND

- 1 From your desktop, click your workspace name in the top left.

Acme Corp. ▾

● Sharon Robinson

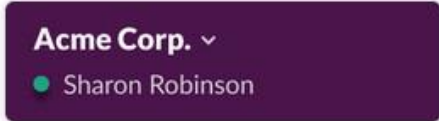
- 2 Hover over **Pause notifications**.
- 3 Choose a time frame from the menu or select **Custom** to set your own.


Slack scheduling notification pauses

Set a DND schedule

For the times you're routinely offline, you can set a Do Not Disturb schedule. With a schedule in place, your notifications will pause every day between the hours you've specified. Keep in mind that Workspace Owners and Admins can [set a default Do Not Disturb schedule](#) for their members.

Desktop iOS Android

- 1 From your desktop, click your workspace name in the top left.

- 2 Hover over **Pause notifications**, then select **Do Not Disturb schedule** from the menu.
- 3 Under **Do Not Disturb**, check the box next to **Automatically disable notifications** and select a start and end time. Your settings will automatically save, so click the **×** close icon when you're done.

 **Tip:** To set yourself to DND over the weekend, pause your notifications until the time and date you'll be back at work. There isn't currently a way to automatically pause notifications every weekend.

Online status indicators (OSIs)

- Note that users and Slack organizations can disable OSIs



Visibility of online status / audit logs

- Organizations might want to know who is clocking in/out
- Zoom lets meeting hosts know when it is backgrounded

Statistics / Data Science in three slides

Data Science

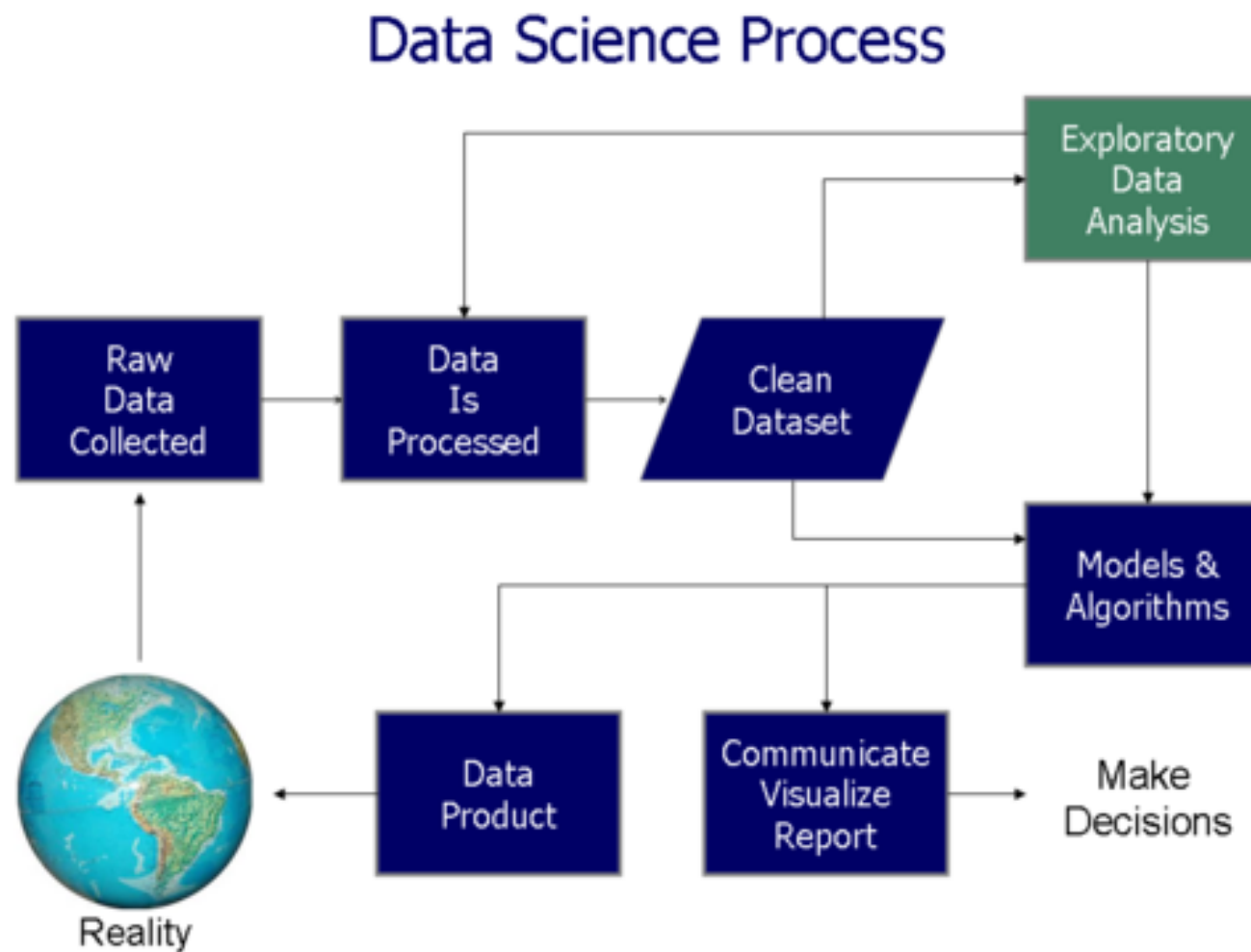


Image taken from: <https://www.kdnuggets.com/2016/03/data-science-process.html>

Descriptive Statistics

Goal: Describe the data in your sample

- Mean, weighted mean
- Histograms, skew, tail
- Correlation
- Variance and standard deviation
- Outliers
- Median
- Interquartile Range

Inferential Statistics

Goal: Generalize / make conclusions from the data in your sample

- Making inferences about the population distribution
- Confidence intervals
- Hypothesis testing (e.g., calculating a p-value)
- Building models (e.g., classifiers, predictive models)