Ethics Protocol:
A method for designing *responsibly*

6.170 Fall 2020
Serena Booth
Content: Milo Phillips-Brown and Abby Jaques

Responsibility

Responsibility

Some Choices

A. Happy to do it

B. Reluctant but would do it

C. Object to doing it and ask for an alternative task, but would do it if I had to

D. Leak information to the public, but don't resign from my job

E. Keep your job, but organize with others to stand up to leadership in the future.

F. Resign from my job rather than do it

G. Resign from my job and leak information to the public

You're an engineer at change.org.

Your users have been learning that your company is for-profit, and that your petitions are never presented to government in an official capacity. They're starting to leave your platform.

Your manager asks you to design an alternative change.org website to look and feel exactly like real government petitions using their newly purchased change.whitehouse.**org** domain.

What do you do?

Cite: Abeba Birhane

Responsibility

Some Choices

A. Happy to do it

B. Reluctant but would do it

C. Object to doing it and ask for an alternative task, but would do it if I had to

D. Leak information to the public, but don't resign from my job

E. Keep your job, but organize with others to stand up to leadership in the future.

F. Resign from my job rather than do it

G. Resign from my job and leak information to the public

You're a UROP. You're interested in going to grad school, and you had a hard time getting this position. You're hoping for a good letter.

Your grad student mentor works on a method for explaining the decisions of autonomous systems. Your grad student recently realized their system could also work for drones.

They task you with creating a website to help military personnel assess UAV decisions. They say, "they're already using the drones, this will just increase accountability."

What do you do?

Cite: Abeba Birhane

Ethics: why should I care?

Ethics: why should I care?

We claim: exploring ethics and assessing your values now will help you make better decisions in the future.

Now: less pressure, less stress, fewer sources of conflict.

The ethics protocol is a tool to help you with such decisions.

Ethics: why should I care?

We claim: exploring ethics and assessing your values now will help you make better decisions in the future.

Now: less pressure, less stress, fewer sources of conflict.

The ethics protocol is a tool to help you with such decisions.

You will use the ethics protocol for your final projects.

Ethics Protocol Goals

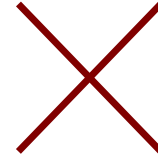- Make more informed decisions

- Justifying existing/past decisions

Ethics Protocol Goals

- Make more informed decisions
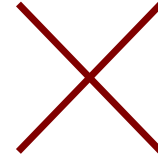
- Assess priorities, compromise

- Justifying existing/past decisions

- Search for a simple "right" answer

Ethics Protocol Goals

✓

✗

- Make more informed decisions

- Assess priorities, compromise
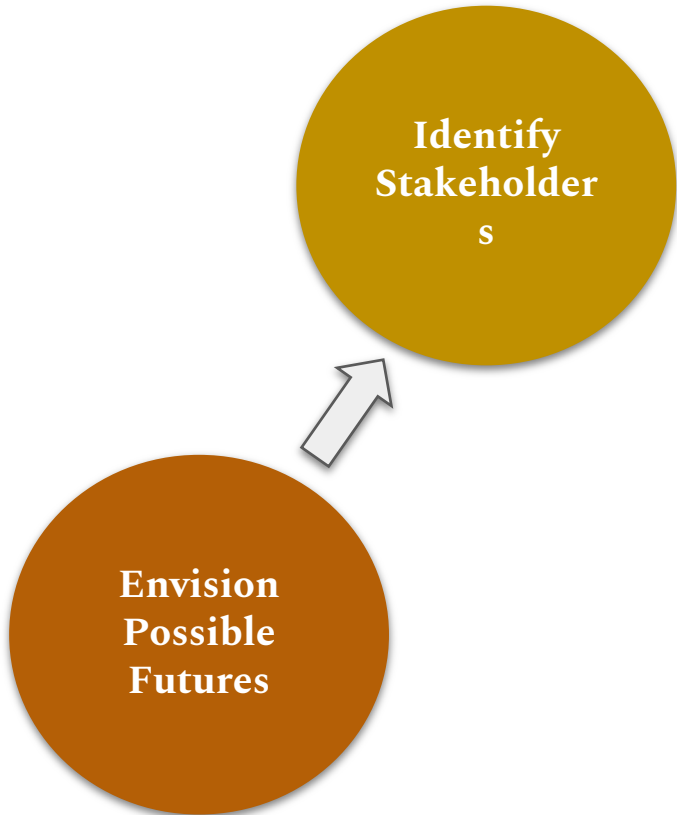
- Avoid thoughtless blunders

- Justifying existing/past decisions

- Search for a simple "right" answer

- Replacing participatory design
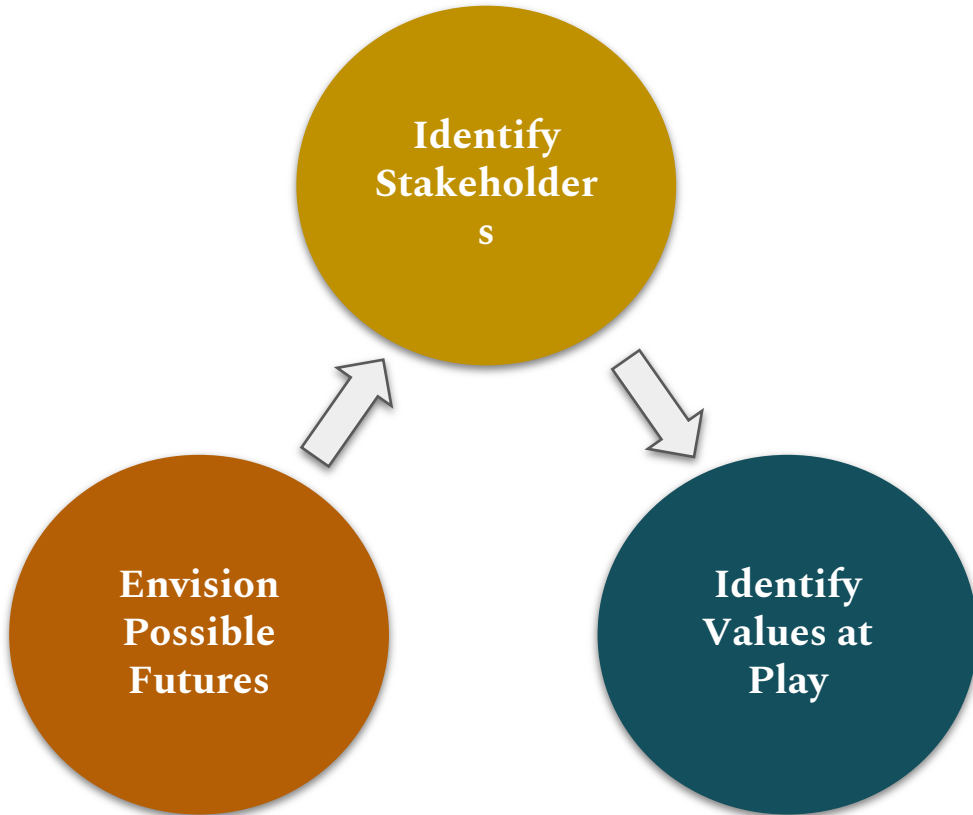
Ethics Protocol Overview

Envision Possible Futures

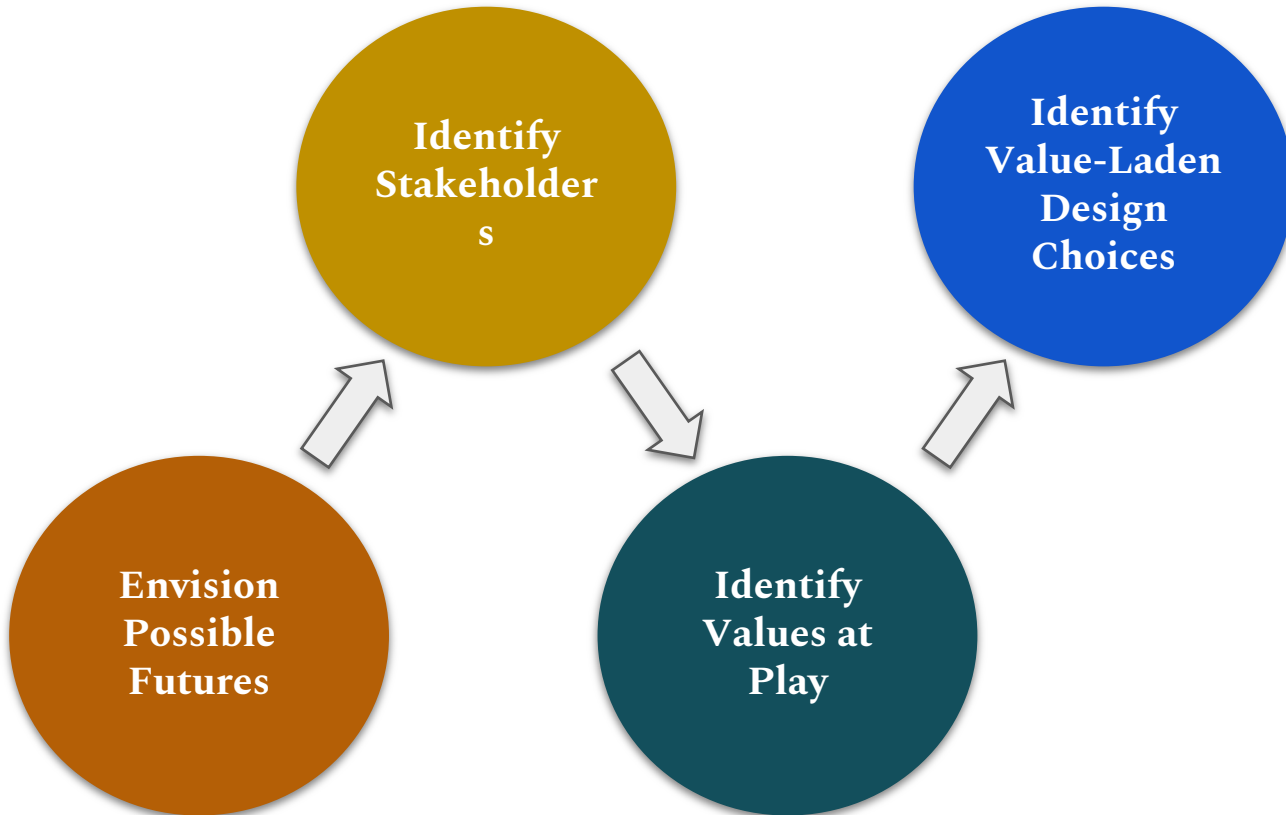# Ethics Protocol Overview

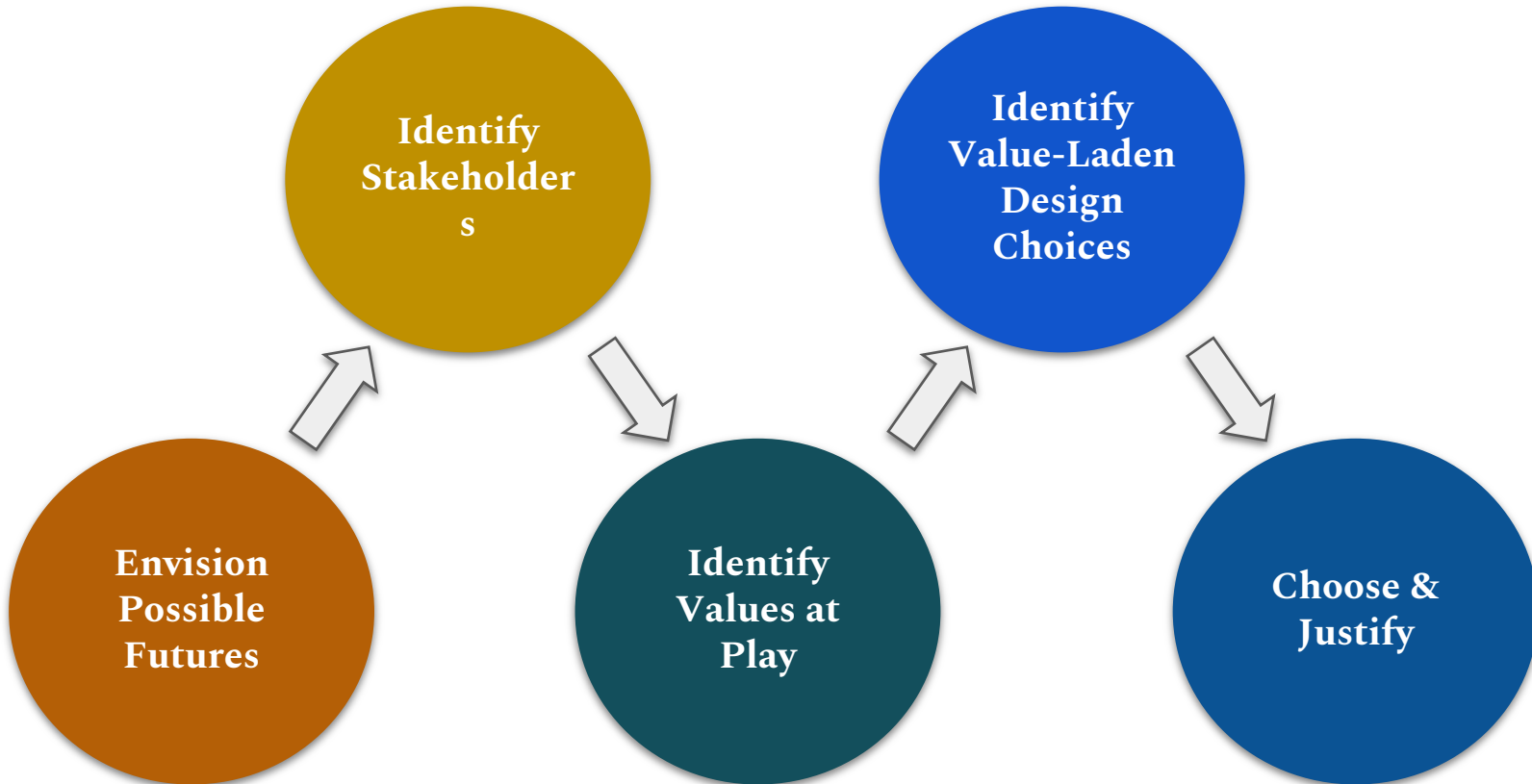**Identify Stakeholders**

Envision Possible Futures

Ethics Protocol Overview

Ethics Protocol Overview

Envision Possible Futures → Identify Stakeholders → Identify Values at Play → Identify Value-Laden Design Choices

Ethics Protocol Overview

Ethics Protocol Overview

Envision Possible Futures → Identify Stakeholders → Identify Values at Play → Identify Value-Laden Design Choices → Choose & Justify → Build
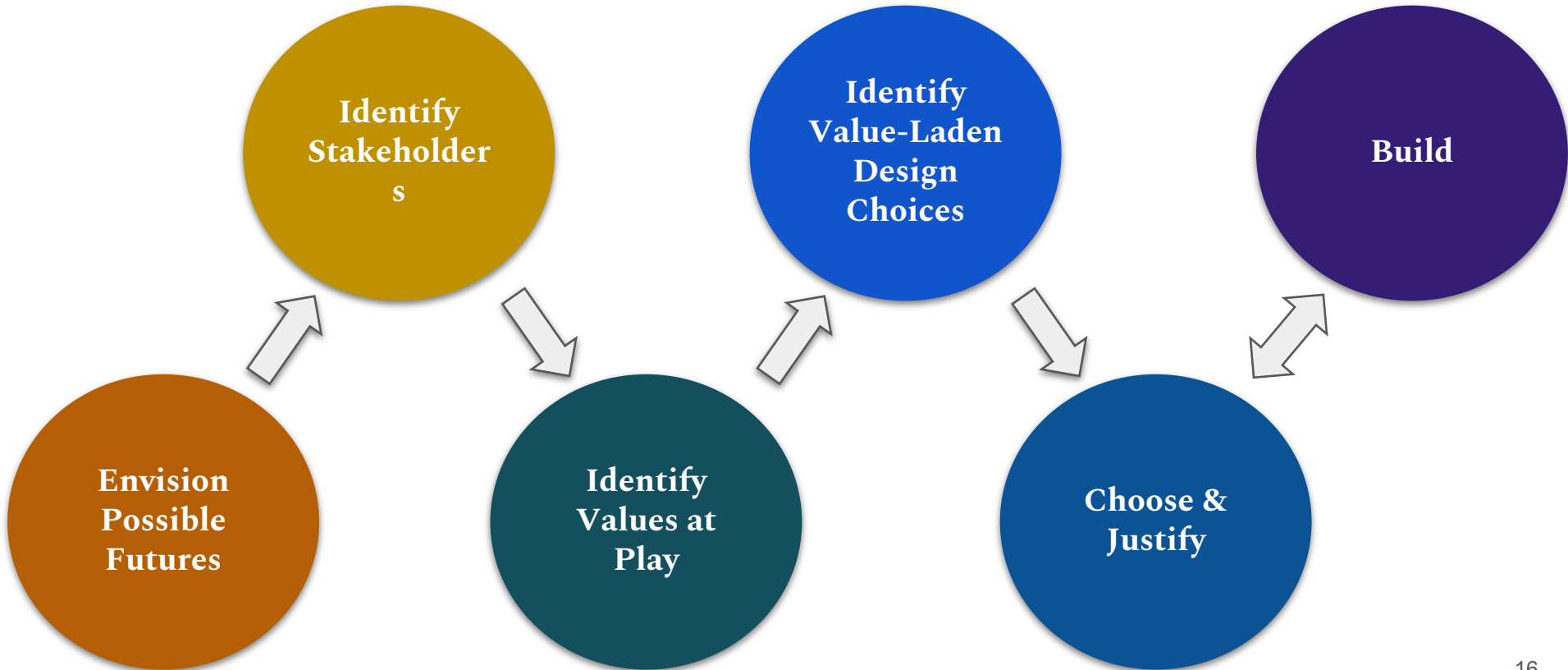
cw: COVID-19

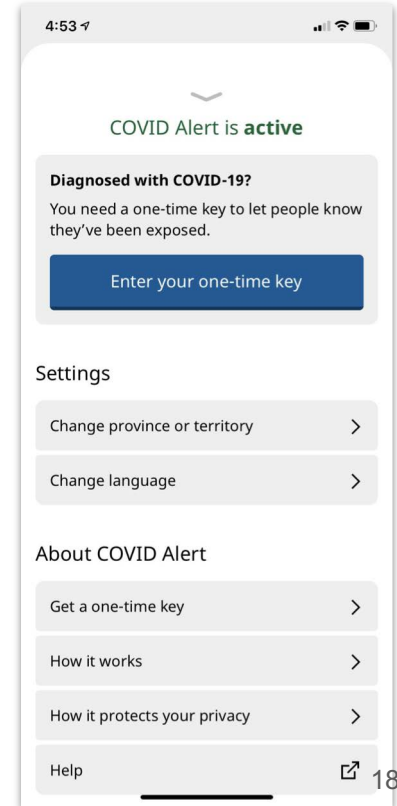# Running Example: Digital Contact Tracing

Two main features:

1. ## Contact Identification
   people who have been exposed are identified

1. ## Contact Notification
   people are notified of their potential exposure

18
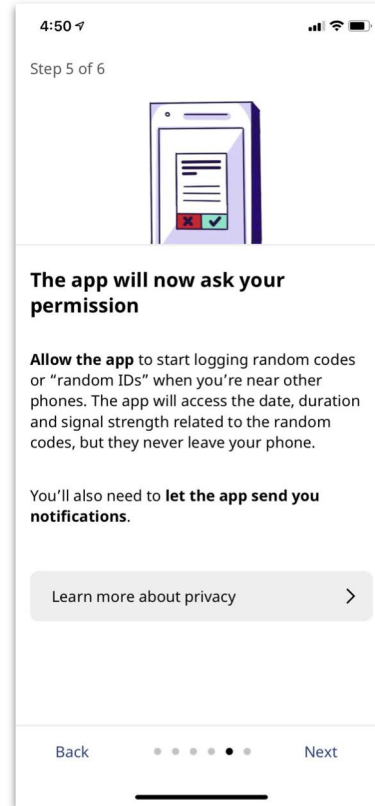
Preliminary design choices

1. Is contact tracing always "on," or is it just used at certain establishments (e.g., restaurants)?

2. Does the app use phone-to-phone communication, or do we add additional hardware beacons?

3. Do we use bluetooth or GPS or neither?

4. How is the app distributed?

5. …

Ethics Protocol Overview

Envision Possible Futures

Contact Tracing Imagined Futures

The government releases a contact tracing app. Within two months, there were zero COVID-19 cases in the country. Hundreds of thousands of lives were saved. Schools reopened, businesses flourished, and the roaring twenties kicked off. The government preemptively decided to stop collecting data from users, but maintained the app for in case of future pandemics.

Contact Tracing Imagined Futures

< Room X: Add your imagined futures>

Contact Tracing Imagined Futures

5 minutes

Contact Tracing Imagined Futures

POSITIVE
- Faster pandemic recovery

NEUTRAL
- What happens to data after contact tracing is know longer needed?
- Increased Legislation for reusing the app
- Data used for future research

NEGATIVE
- Track certain groups movements through the app (e.g. criminals)
- Overreach for other examples that don't require…
- Make assumptions about movement of certain groups

Contact Tracing Imagined Futures

- Surveillance state
  - Disproportionately harms marginalized communities
  - Protestors/whistleblowers at risk (one person traced can lead to tracking others)
- Buggy App
  - Loses User's Trust, causes mass panic, doesn't prevent spread
- App is hacked/compromised
  - Can be used for ransom/stalking/blackmail

Contact Tracing Imagined Futures

- *Privacy issues: creators of the app have access to other users' locations and may use users' information for financial gain (e.g. selling to companies), could be used in turn for individual marketing*
- *Overconfidence issue: users' behavior becomes more risky as they underestimate exposure and reduce testing frequency*
- *Government abuse: using the app to track groups of people / how they interact, potentially voter suppression*

Contact Tracing Imagined Futures

< Room 4: Add your imagined futures>

Good:

- *Limited only for pandemic*

Bad:

- *Using for other domains*
- *Collected data unintentionally leaks sensitive information or is misused/sold*

Contact Tracing Imagined Futures

Social media could be transformed if this contact tracing app takes off. Something like proximity-based socialization and connecting to others.

Surveillance from government and corporations increases. What if our location data is all sold?

Ethics Protocol Overview

Identify Stakeholders

Envision Possible Futures
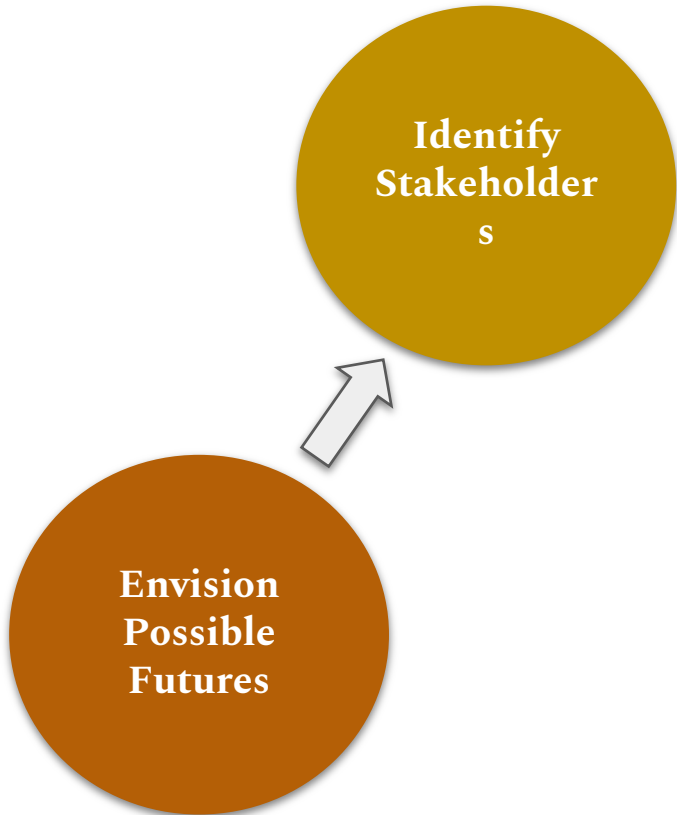
What is a **stakeholder**?

What is a **stakeholder**?

A stakeholder is *anyone or any thing* that can affect or be affected by your project.

Stakeholders are not (just):

You & your institution

Financial backers (shareholders)

Users

Why define stakeholder so broadly?

Why define stakeholder so broadly?

There are no *ethical externalities.*

That doesn't mean you're responsible for everything.

But, to make things ethically,
you need to know
what all the ethical elements are!

Stakeholder *subgroups* matter.

Contact Tracing Stakeholders

Smartphone users

Not-smartphone users

Contact Tracing Stakeholders

Smartphone users

Not-smartphone users

Black people

Hispanic or Latino people

Contact Tracing Stakeholders

Task: Brainstorm as many stakeholders as you can

5 minutes

Contact Tracing Stakeholders

- *Coronavirus(?)*
- *Low income communities*
- *Developers*
- *Governments*
- *Different Age Groups*
- *People in areas with worse connectivity*
- *People who are benefiting from the pandemic.*

Contact Tracing Stakeholders

< Room 2: Add your stakeholders>

- Healthcare workers
- Essential workers
- Businesses/potential hubs for transmission
- Maker of app

Contact Tracing Stakeholders

- Homeless & low-income
- Healthcare workers
- BIPOC
- Non-smartphone owners
- Elderly
- Disabled
- Children
- Essential workers
- Public transportation users
- Government (all levels)

43

Contact Tracing Stakeholders

- *Users*
- *Non-Users: Elderly (non-smartphone users)*
- *Government*
- *Hospitals*
- *Testing Centers*
- *Academia*
- *Big pharma*
- *Inter-Country Tracing Systems*

Contact Tracing Stakeholders

< Room 5: Add your stakeholders>

- Lower-income, anyone without a phone or areas with poor connectivity
- Rural vs. urban?
  - Proximity to other people (tracing you through other people's devices)
- Older vs younger
  - People
  - Tech
    - New tech - less hackable but built in hardware
- BIPOC

Ethics Protocol Overview

Identify Stakeholders

Envision Possible Futures

Identify Values at Play
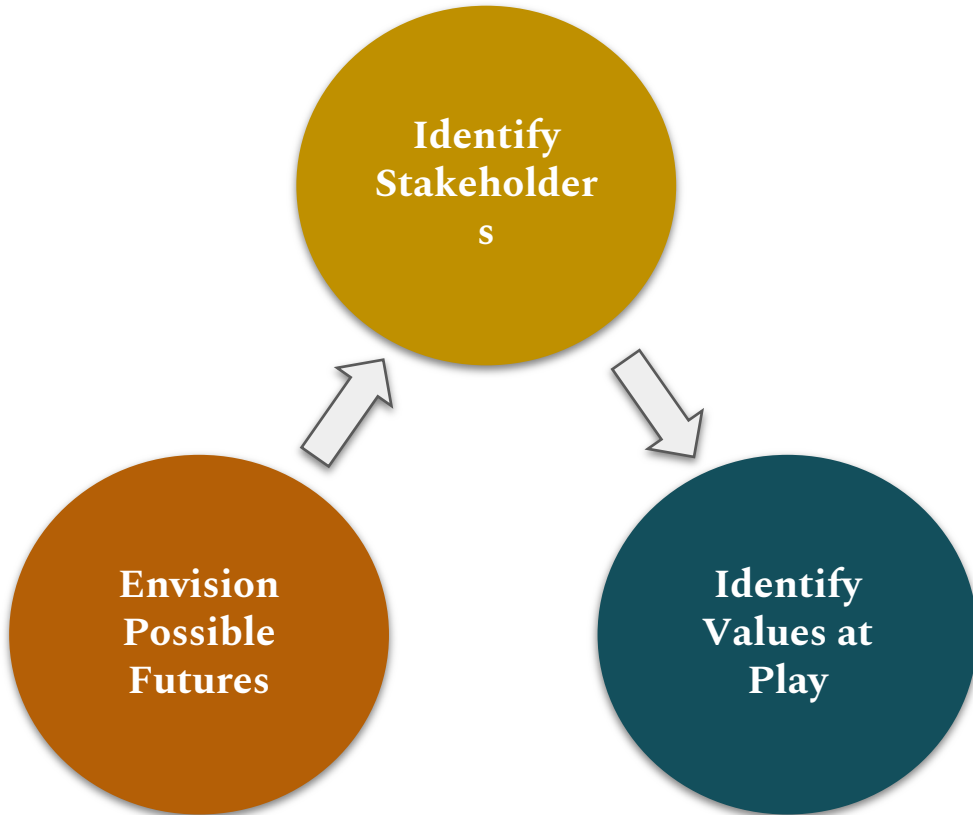
Moral Lenses

Different ways of looking at effects on stakeholders that reveal different kinds of ethical significance.

Outcome Lens

In what ways does what you're making turn out better or worse for your stakeholders?

Process Lens

How did the process treat stakeholders?

Think: autonomy, consent, transparency, participation, etc.

Structure Lens

How are the outcomes distributed among different stakeholders? What are the differences in how different stakeholders were treated by the process?
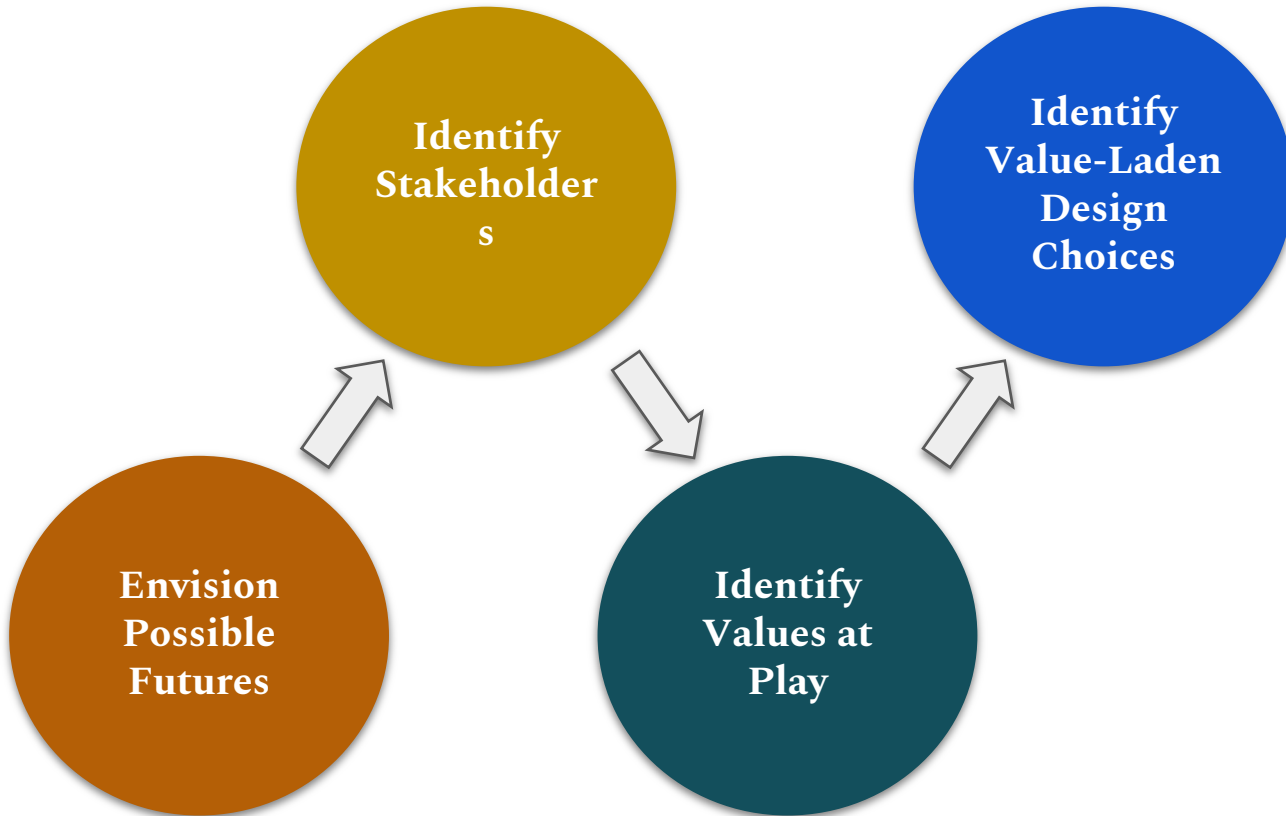
Could be called 'Justice Lens'

Outcome, Process, Structure

What's missing?

What are **your** values?

Ethics Protocol Overview

Envision Possible Futures → Identify Stakeholders → Identify Values at Play → Identify Value-Laden Design Choices

Two ways to identify value-laden design choices

Two ways to identify value-laden design choices

1. Work forward:
   Take a choice you know you'd have to make and trace out the value-laden implications

Preliminary design choices

1. How are COVID-19 cases identified?

Possible Choice	Values promoted (and for whom?)	Values demoted (and for whom?)
Option A		
Option B		
Option C		

Possible Choice	Values promoted (and for whom?)	Values demoted (and for whom?)
Users post their COVID-19 status without proof		
Option B		
Option C		

Possible Choice	Values promoted (and for whom?)	Values demoted (and for whom?)
Users post their COVID-19 status without proof		
Users post their COVID-19 status with doctor's note		
Option C		

Possible Choice	Values promoted (and for whom?)	Values demoted (and for whom?)
Users post their COVID-19 status without proof		
Users post their COVID-19 status with doctor's note		
We identify COVID-19 status by monitoring vocal patterns [src]		

Possible Choice	Values promoted (and for whom?)	Values demoted (and for whom?)
	Structure Lens: users don't have to see a doctor	
Users post their COVID-19 status without proof	**Process Lens:** the process is entirely transparent	**Outcome Lens:** bad actors can prank the system
Users post their COVID-19 status with doctor's note		
We identify COVID-19 status by monitoring vocal patterns [src]		

Possible Choice	Values promoted (and for whom?)	Values demoted (and for whom?)
	Structure Lens: users don't have to see a doctor	
Users post their COVID-19 status without proof	**Process Lens:** the process is entirely transparent	**Outcome Lens:** bad actors can prank the system

Structure Lens: high barrier to use |
| Users post their COVID-19 status with doctor's note | **Outcome Lens:** the least false positives | **Process Lens:** compromises privacy

Process Lens: users didn't consent to this data collection |
| We identify COVID-19 status by monitoring vocal patterns [src] | **Outcome Lens:** we're able to identify the most cases using this technology | **Structure Lens:** the model only uses English data. |

61

Two ways to identify value-laden design choices

1. Work forward:
   Take a choice you know you'd have to make and trace out the value-laden implications

2. Work backward:
   You've identified futures and their ethical dimensions. Which choices result in which outcomes?

Working backward

One envisioned future is that there are so many false positives, the system is incredibly ineffective and we give up using it.

One contributing design decision could be this question of how positive cases are communicated.

Preliminary design choices

1. How are COVID-19 cases identified?

2. GPS vs Phone-to-Phone Bluetooth

3. Opt-in vs opt-out?

4. Distribution mechanism?

5. Always on, or only used in certain locations?

6. …

Possible Choice	Values promoted (and for whom?)	Values demoted (and for whom?)
GPS	Outcomes: we can remove outdoors Justice: works anywhere in the world Outcomes: we're more confident you were in close proximity	Outcomes: generally worse protocol Process: collecting a lot more data
Bluetooth		
Option C		

Room 1

Possible Choice	Values promoted (and for whom?)	Values demoted (and for whom?)
Option A		
Option B		
Option C		

Room 2

Ethics Protocol Overview

Envision Possible Futures → Identify Stakeholders → Identify Values at Play → Identify Value-Laden Design Choices → Choose & Justify
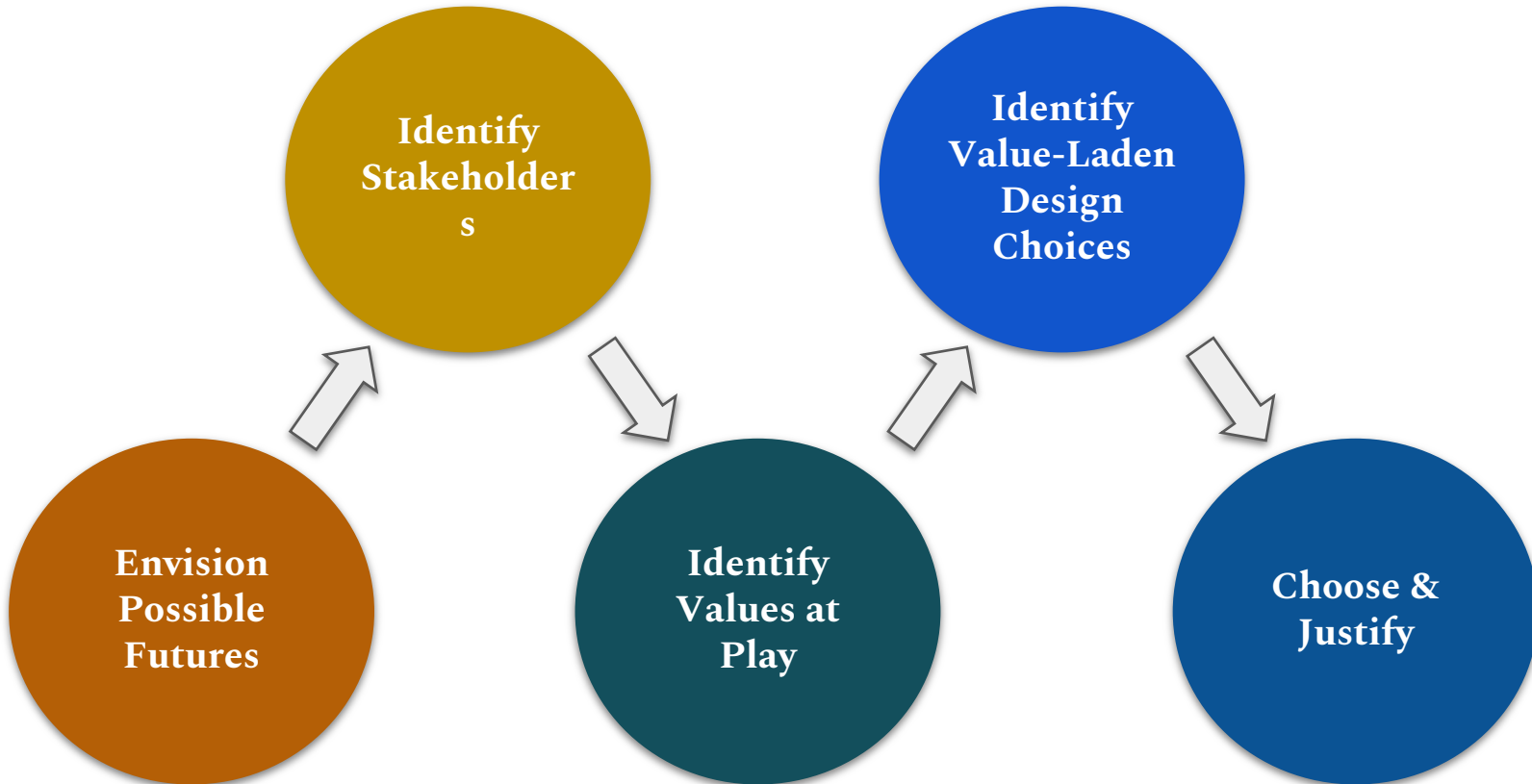
1. How are COVID-19 cases identified?

Choice: Users post their COVID-19 Status by acquiring an authorization code from their doctor.

- This prevents bad actors from manipulating the system, forcing unnecessary self-isolation, unlike without requiring authorization codes.
- Users consent, unlike with the voice monitoring solution.
- Fewer users report, since this requires the financial means/feelings of security needed to see a doctor in the US.
- Privacy is compromised since users can be associated with a specific doctor's visit.

Contact Tracing Choose & Justify
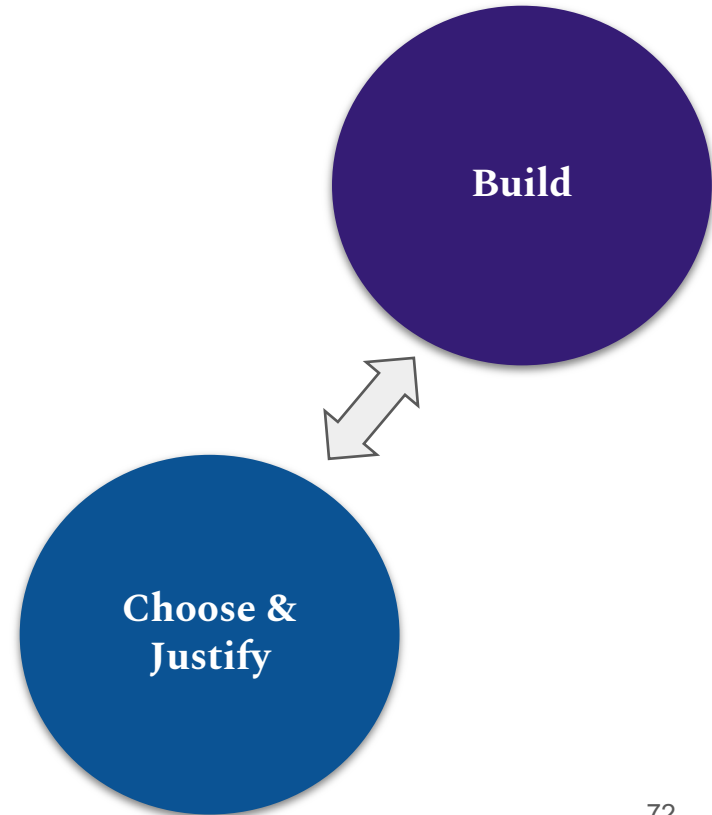
5 minutes
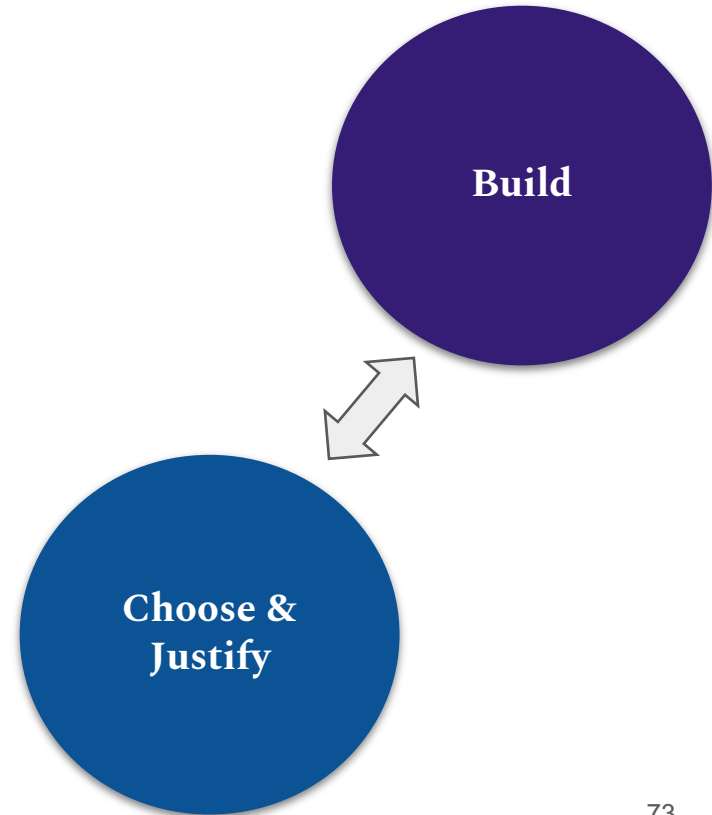
1. Question?

Choice

Justification

Ethics Protocol Overview
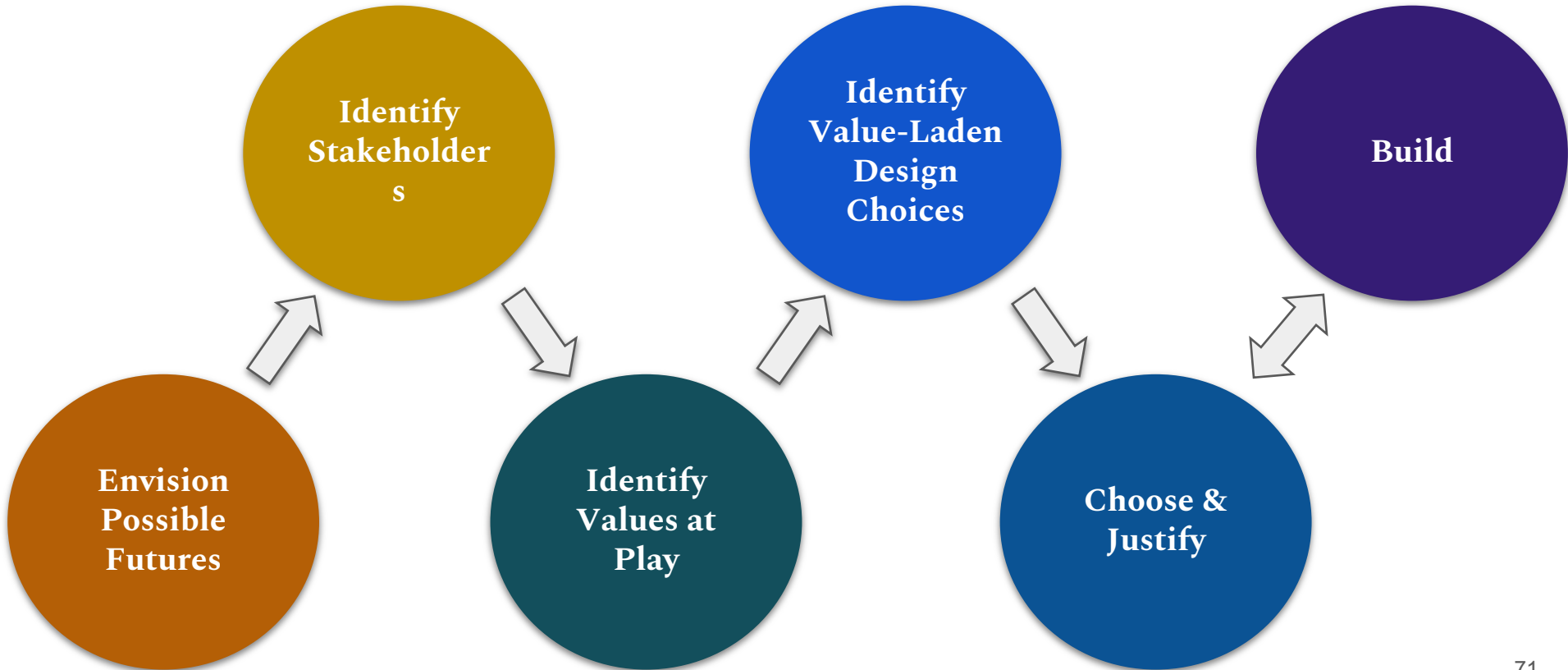
"Minor" technical decisions can have ethical consequences.

Build

Choose & Justify

Don't forget to ask:
"Should I build this?"

Fin!

Reminder: You'll use the ethics protocol
in your final projects.

Learn More Ethics!

Classes (Spring `21):

- 24.191 Being, Thinking, Doing (or not): Ethics in Your Life
- 24.03 Good Food: The Ethics and Politics of Food
- 24.231 Ethics: Systematic study of central theories in ethics
- 24.237(J) Feminist Thought

Other:

- PKG Center - internships, colloquia, funding for your ideas, lots of opportunities!
- AI Ethics Reading Group, AI Alignment Reading Group, & more

Q&A