Data Ethics: Choices and Values

Veronica A. Rivera, Ph.D.

McCoy Family Center for Ethics in Society - HAI

Original slides and content by Kathleen Creel, Diana Acosta-Navas, and Benjamin Xie

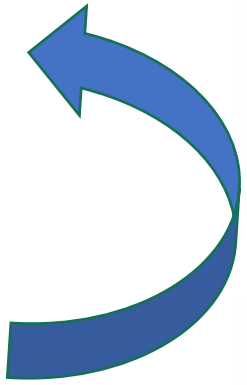
We use data to inform our decisions

- Evidence-based
- Impartial
- Reliable

What can we learn from a data set?

- Patterns
- Correlations
- Distributions
- …

- Values



QUALITY

5.0

DIFFICULTY

1.0

CS101 😎 **AWESOME** **May 21st, 2015**

Attendance: **Not Mandatory** Grade: **A** Textbook: **Yes** Online Class: **Yes**

This class was awesome. A beginner like me that has never done anything further than facebook on a computer, Professor was very clear and easy to listen to. I very much enjoyed the lectures and how easy it was to learn from such a great teacher. Thank you for all that you do

2 1

Data is intrinsically values-laden

VALUES IN DESIGN

PROBLEM FORMULATION

DATA COLLECTION

DATA INTERPRETATION

What are values?

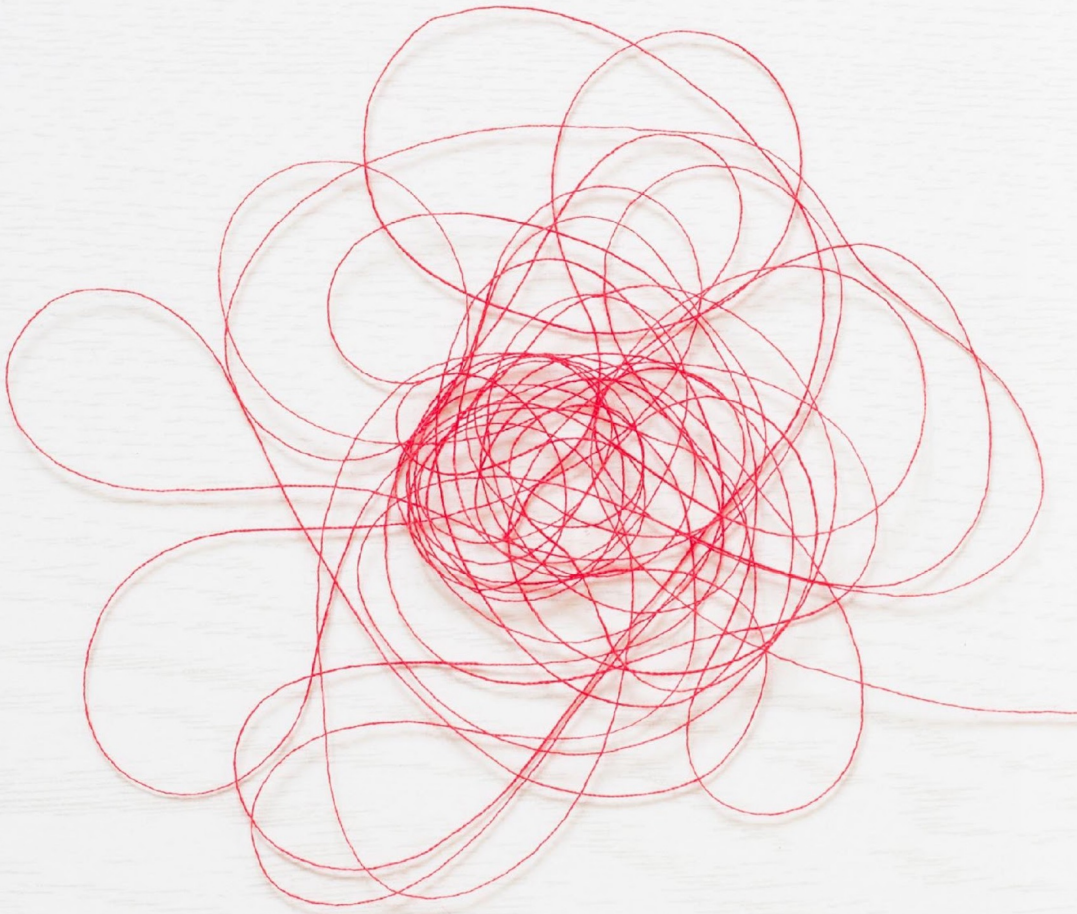
- *value (n)*: an individual or community's belief about what matters
- Values express **what we care about**
  - Efficiency
  - Privacy
  - Truth
  - Security
  - Beauty
  - Fairness
  - …
- Values reveal our **assumptions** about the world, people interacting with our designs, and how our choices affect them

Intentionality of values

- Explicit values: Values that designers **intend** for their products to embody

- Collateral values: Values that crop up as side effects of design decisions and the way users interact with those designs. These values are **not intentionally** designed into the system.

PROBLEM FORMULATION



Problem Formulation Statements

- Formulating a problem means treating the **desired solution as good or worthy** of being done.
- **Why** should we care about solving this problem?
- **Who** can agree that this is a problem worth solving?
- Who would **benefit** from its solution?

Example: Unsubscribe feature

1. People don't like our app/service & it should be easy for them to unsubscribe
   - Potential design: Unsubscribe button on app homepage. Single click and user is unsubscribed.

2. The user doesn't like our app/service, we want to understand why so we can make it better for future users
   - Potential design: Unsubscribe button on app homepage → 10 minute questionnaire → talk with a representative

Example: Course evaluations

1. Students have different learning needs. We want to help students decide which courses best fit their learning style

> **9 - What would you like to say about this course to a student who is considering taking it in the future?All comments are subject to Stanford's Terms of Use for Sites. Answers to this question will be viewable by other students, as well as instructors.**

2. We want to promote and hire good teachers

> **5 - Overall, how would you describe the quality of the instruction in this course?**

3. We want teachers to have the info they need to improve their classes

> **11 - What was your favorite topic from CS 355? Why?**

> **12 - What was your least favorite topic from CS 355? Why?**

DATA COLLECTION

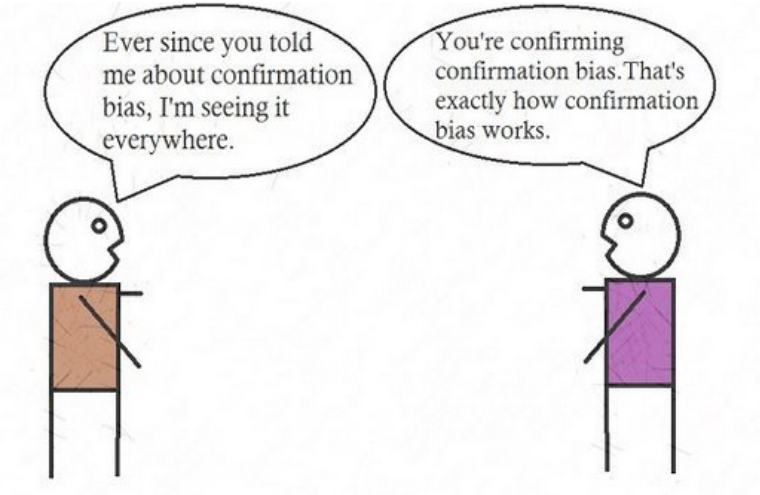
What is data bias: Two definitions

- Statistical: Difference between measured and “true” value
- Social: Human-created biases, such as stereotypes, that arise through embedding of values.

Examples of social biases in data

Confirmation bias

When we favor information that confirms or strengthens our beliefs or values



Source: https://www.statice.ai/post/data-bias-types

Selection bias

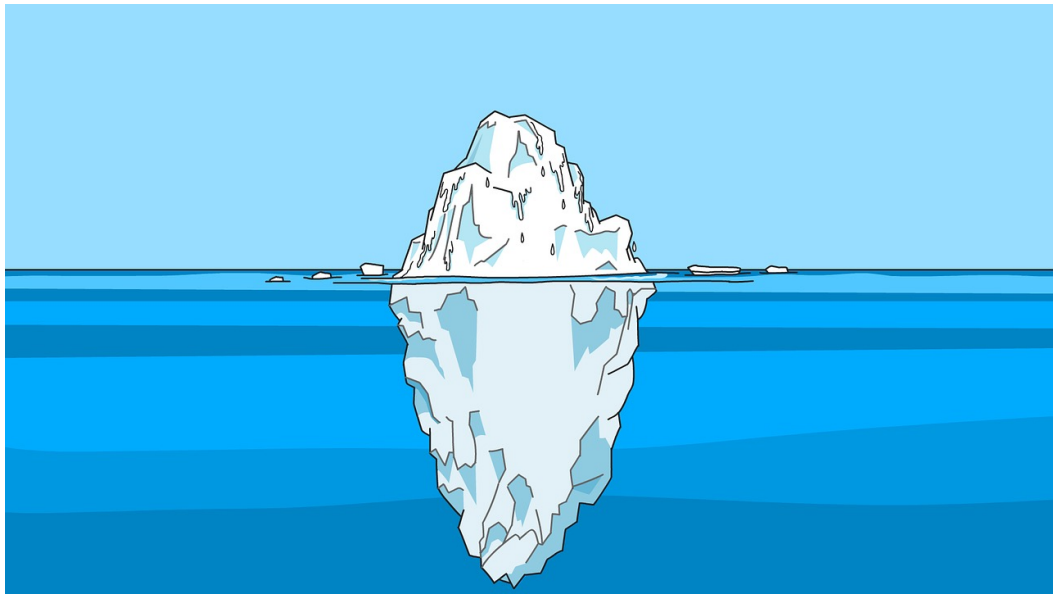
When the selection of study participants or data is not randomized, so the sample is not representative of the entire population intended to be analyzed

Survivorship bias

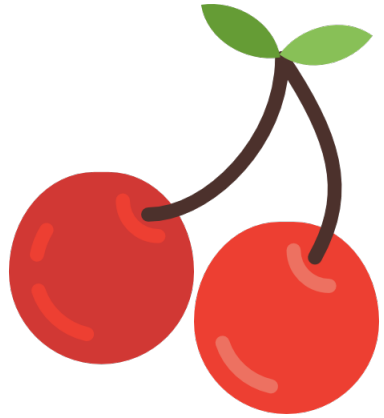
A form of selection bias in which ”winners” are overly focused on in a sample

Example: confirmation & selection bias

Confirmation bias



Selection bias



DATA INTERPRETATION

Descriptive and normative terms

QUALITY

5.0

DIFFICULTY

1.0

CS101 😎 **AWESOME** **May 21st, 2015**

Attendance: **Not Mandatory** Grade: **A** Textbook: **Yes** Online Class: **Yes**

This class was awesome. A beginner like me that has never done anything further than facebook on a computer, Professor Nick was very clear and easy to listen to. I very much enjoyed the lectures and how easy it was to learn from such a great teacher. Thank you for all that you do

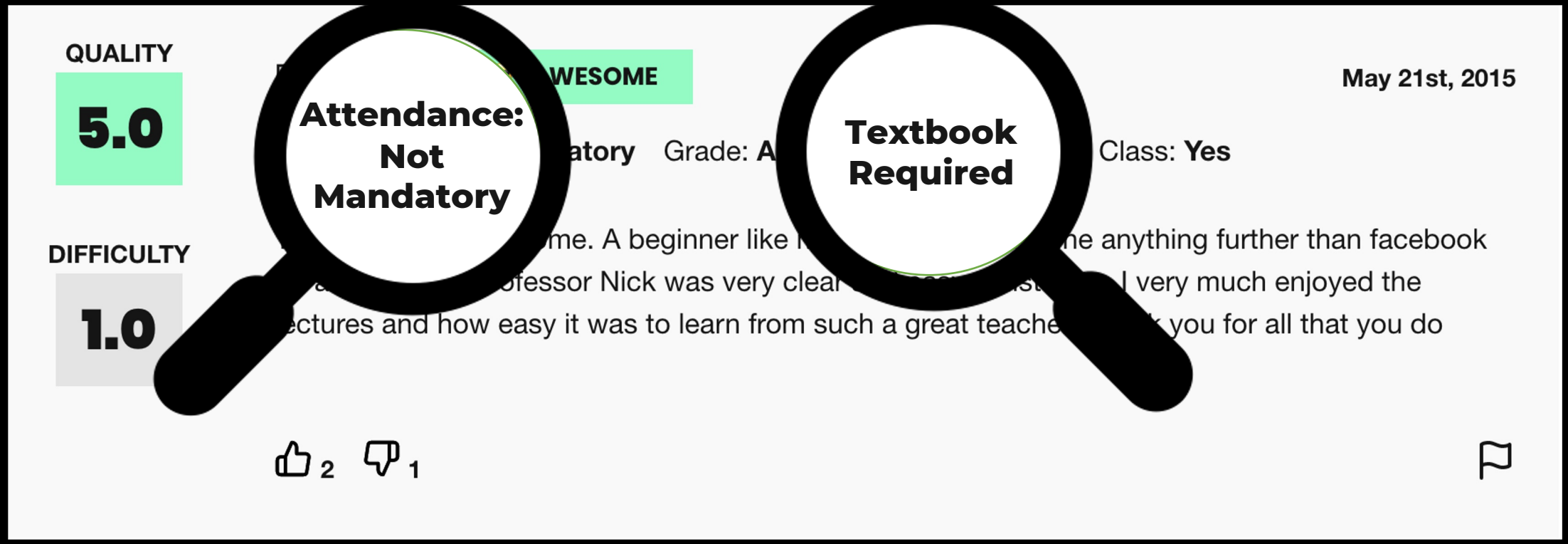
2 1

Descriptive vs. Normative Language

Descriptive language

- Statements of fact
- What people did
- What happened

- “Lectures are 90-minutes long”
- “Assignments take more than two hours to finish”
- “Sections are mandatory”

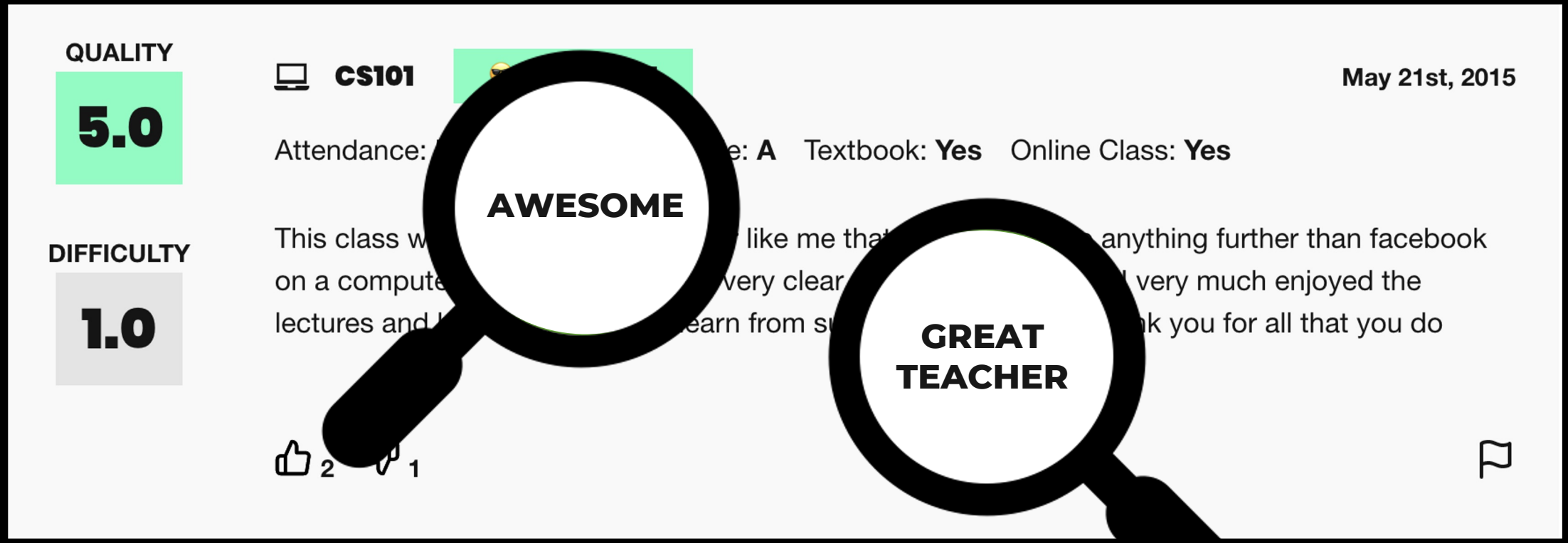

QUALITY
5.0
DIFFICULTY
1.0
WESOME
May 21st, 2015
atory Grade: A
Class: Yes
me. A beginner like
ne anything further than facebook
ofessor Nick was very clear
very much enjoyed the
ectures and how easy it was to learn from such a great teache
you for all that you do
2
1
Attendance: Not Mandatory
Textbook Required

Descriptive vs. Normative Language

Normative language:

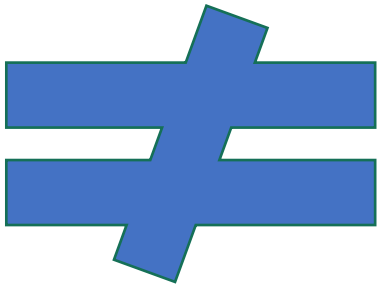
- Evaluative statements
- Express the speaker's opinions/reactions
- How they think things should be

- "right"
- "wrong"
- "good"
- "bad"
- "should"
- "should not"


QUALITY
5.0
DIFFICULTY
1.0
CS101
May 21st, 2015
Attendance:
e: A
Textbook: Yes
Online Class: Yes
This class w
like me that
anything further than facebook
on a compute
very clear
very much enjoyed the
lectures and
earn from su
k you for all that you do
2
1
AWESOME
GREAT TEACHER

Normative language:

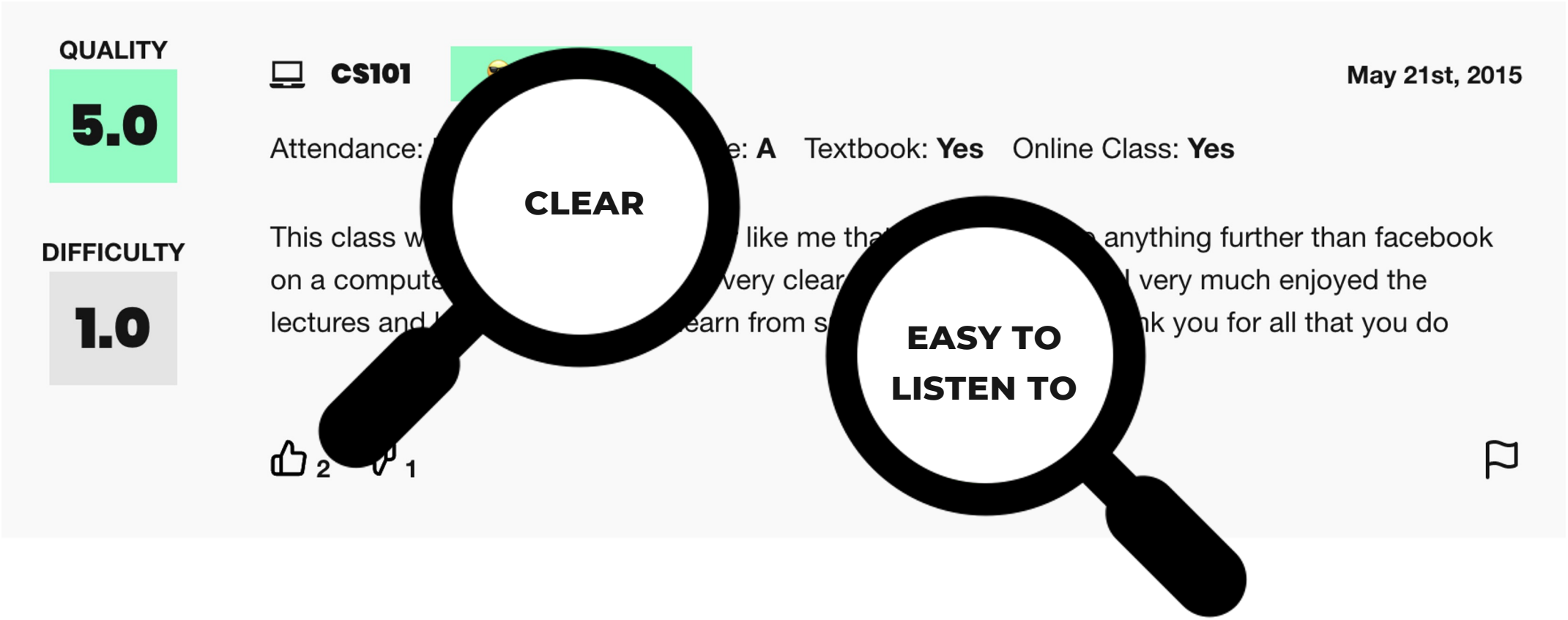
- Evaluative statements
- Express the speaker's opinions/reactions
- **How things should be**



Descriptive language

- Statements of fact
- What people did
- What happened
- **How things "are"**

Distinction between normative & descriptive language is not always clean-cut



24

Thick Normative Terms

Descriptive AND normative:

- Thick normative terms express morally or aesthetically "loaded" descriptions

- Cowardly
- Cautious
- Polite
- Rude
- Chill
- Kind
- Caring
- Smart
- Knowledgeable
- **Professional**

Example: toxic speech classification & context

We should not get rid of normative terms altogether

I want to collect & analyze data to solve a problem. What should I do?

Work with multiple stakeholders/people to identify interesting problems

Look for embedding of values in data during analysis

Check for social biases in your experiment/research setup