Basics of (R)esearch: Measurement

Research Methods for Human Inquiry

Andrew Perfors

I know! I'll get Shadow a dataset. I can ask all sorts of questions to our friends about what they really think and how they think in general.

Before we analyse our data in R, let's take a step back

🔥 Kareem Carr 🔥
@kareem_carr

Something I learned from working with Rafa is studying coding and math can only get you so far. After that you need to ground your analysis in reasonable beliefs about your data which requires really looking at your data and understanding it. This is hard.

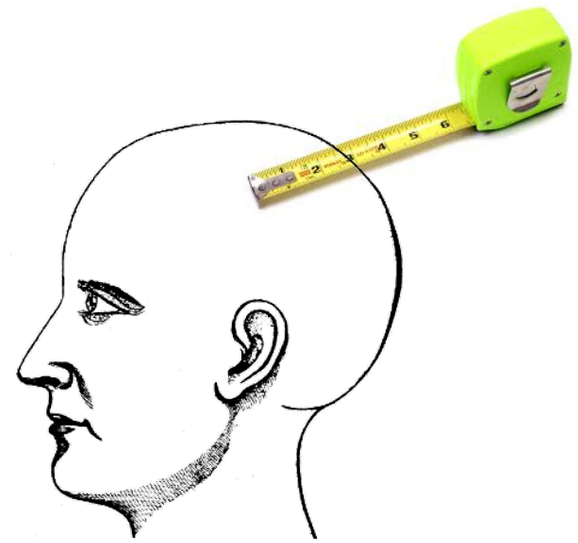
> **Rafael Irizarry** @rafalab · 8h
> The three most important things in data analysis are exploratory data analysis, exploratory data analysis, and exploratory data analysis.

6:48 AM · Mar 2, 2021 · Twitter for iPhone

14 Retweets **1** Quote Tweet **112** Likes

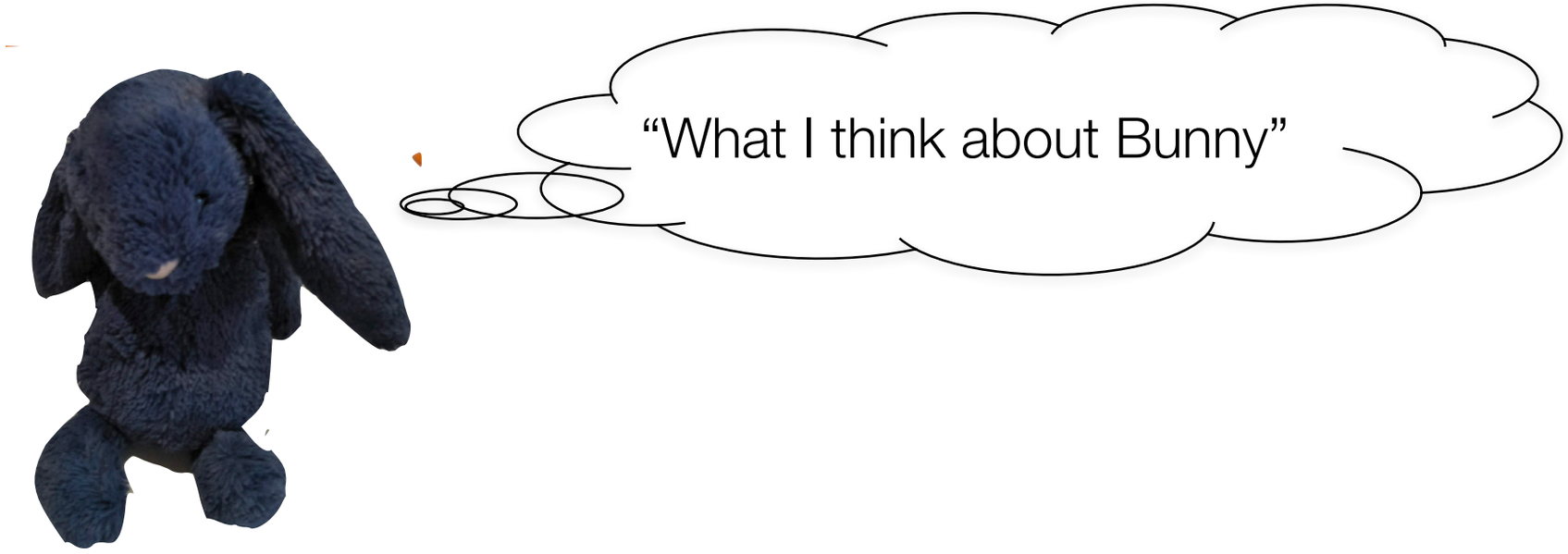
What is psychological measurement?

- Entities we need to distinguish between
  - Theoretical constructs
  - Measures
  - Data (observations)



- Processes involved
  - Operationalisation
  - Measurement

Entity #1: A theoretical construct



- Theoretical constructs are <u>unobservable</u> psychological entities
  - Attitudes, beliefs, hopes, information processing speeds
  - They probably don't "exist" in any literal sense
  - They're handy, but they're not real like "neurons" are real
  - They're tools to help us make sense of ourselves (or others)

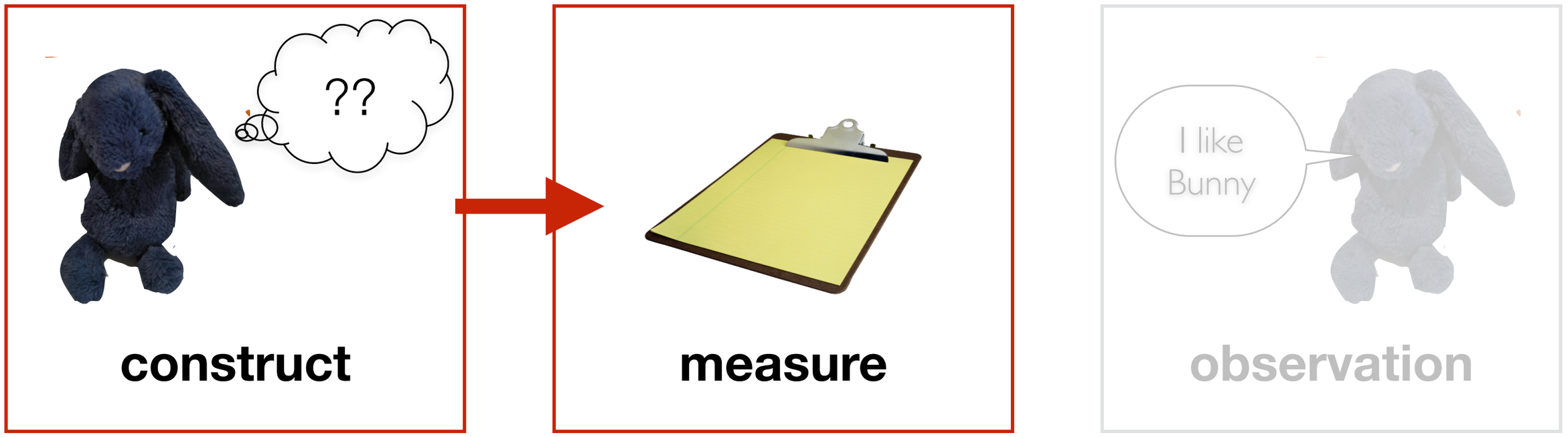
Entity #2: A measure

- A measure is a tool for getting people to produce data
  - Survey items, reaction times, blood oxygenation level
  - Ideally, it elicits data that are informative about the construct
  - (Measures don’t always measure what we think they do)

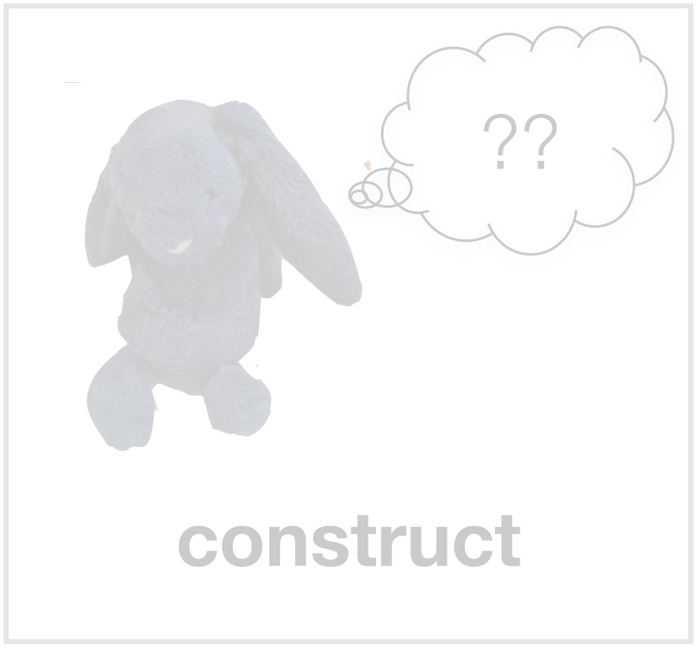
Entity #3: Data / observations

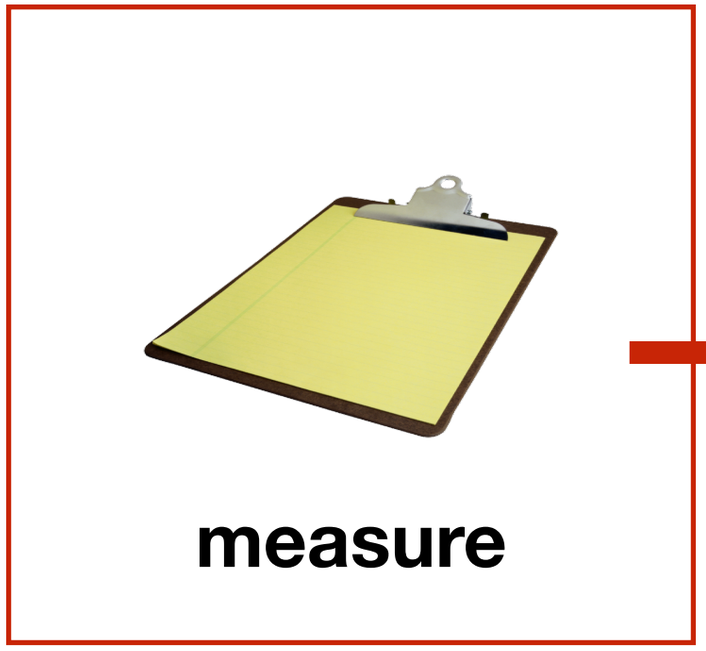
"Yes, I like Bunny"

- Data are what you get when you use the measure
  - Answers to questions, times taken to answer, etc
  - These are the things you can actually observe



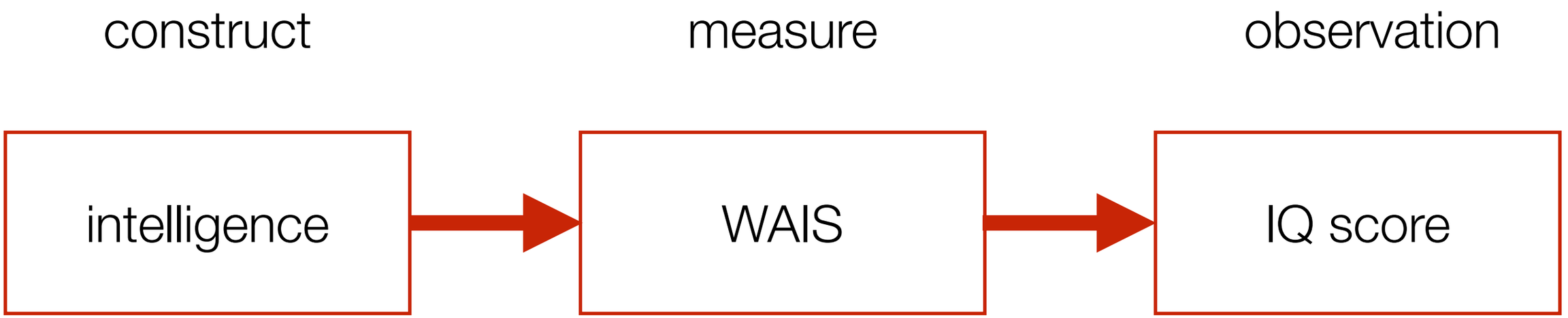
- **Operationalisation** relates “constructs” to “measures”
  - “Beck depression inventory” tries to measure “depression”
  - “Inspection time” tries to measure “processing speed”
  - Operationalisation is the process of finding a good measure
  - It’s hard to do well… measures are wayward beasts



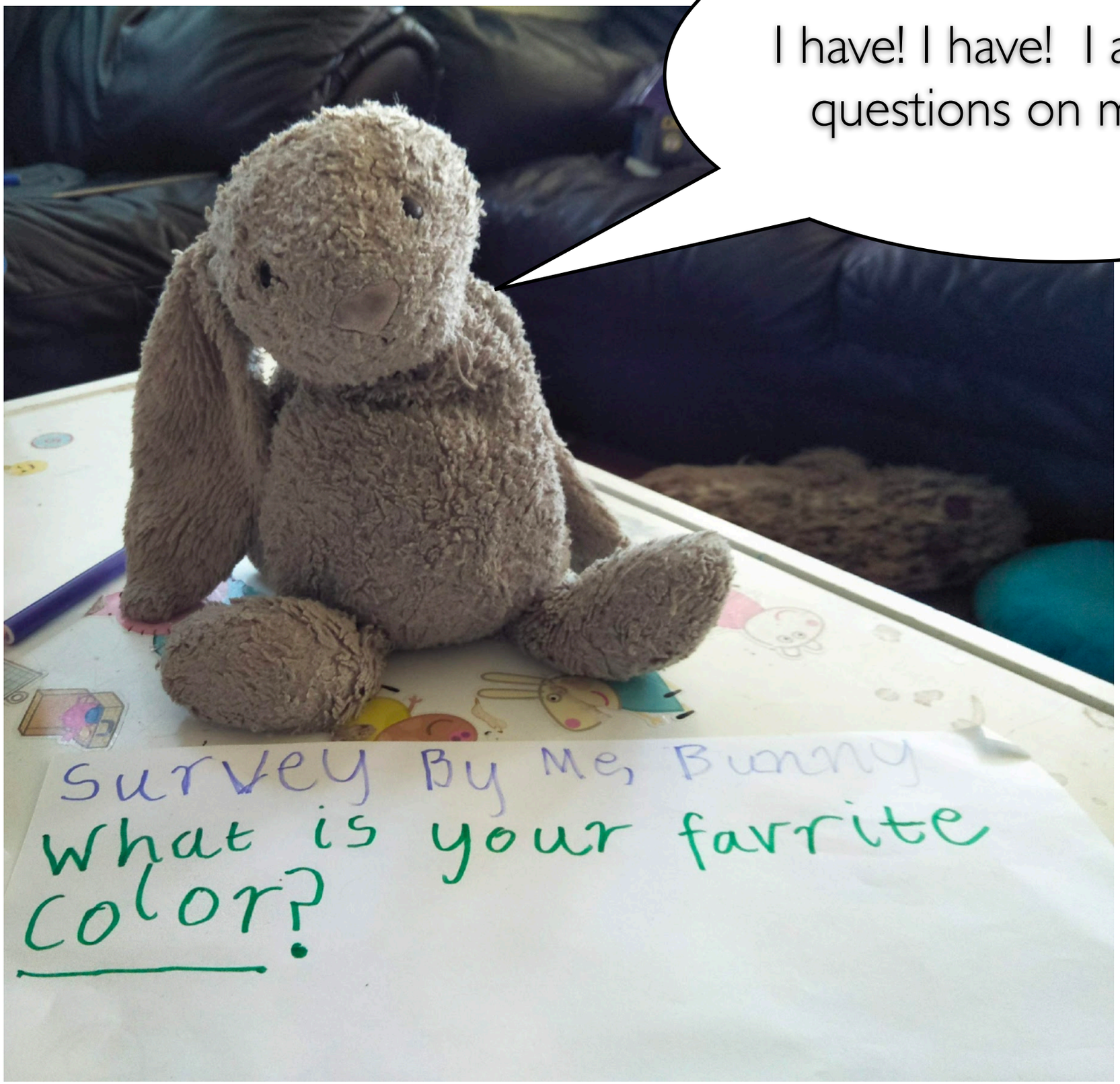


- **Measurement** relates “measures” to “observations”
  - Running the experiment, sending out the survey
  - It is the process of applying a measure to get the data
  - Not as murky as operationalisation, but still tricky

Examples in psychology



Okay, let’s say we’ve done our operationalisation, and applied our measures to get data…



1. What is your favourite colour?

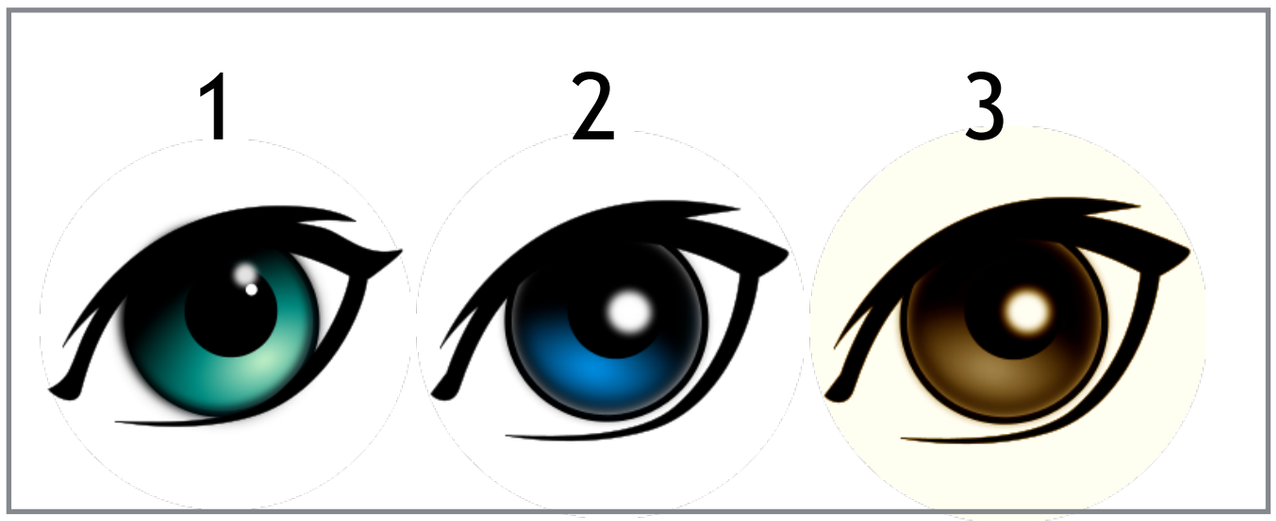
2. How tall are you in cm?

3. Rank the following from best to worst: (a) bunnies; (b) bears; (c) doggies

Scales of measurement

Nominal/categorical scale

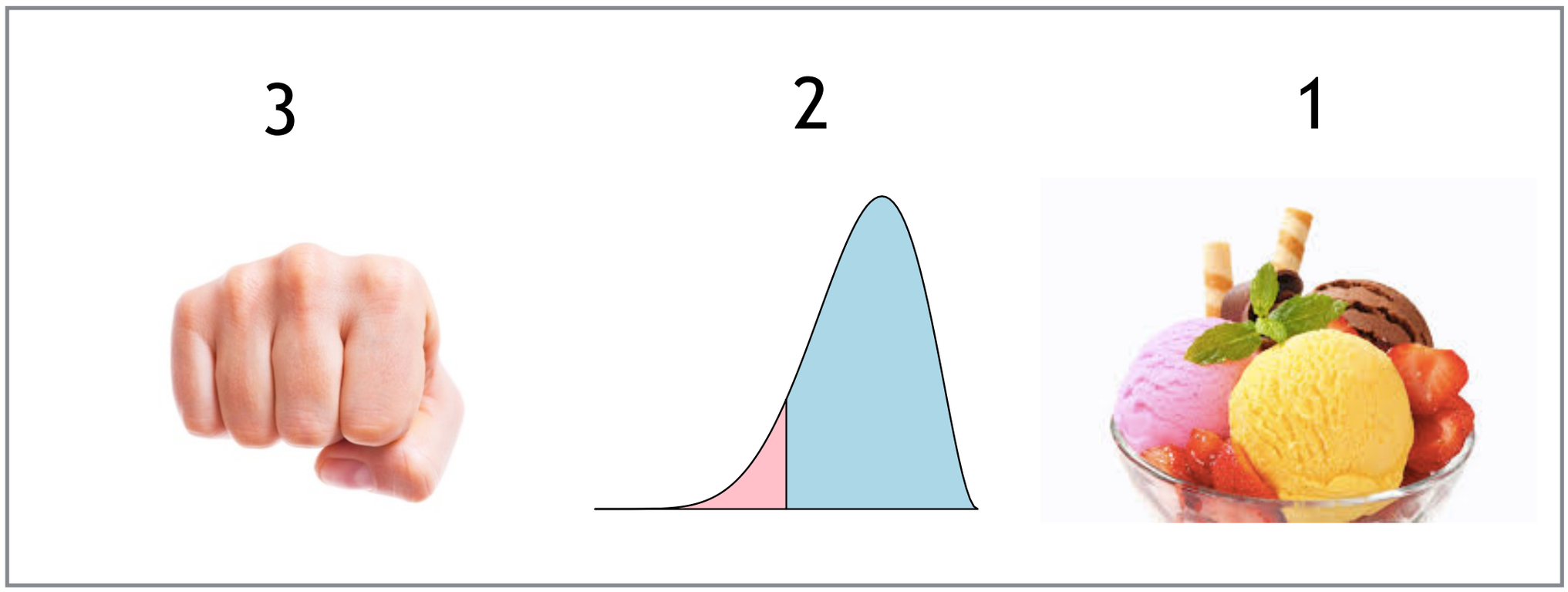
- Possible values have no particular relationship with each other.
- There's no meaningful numbering scheme we can use



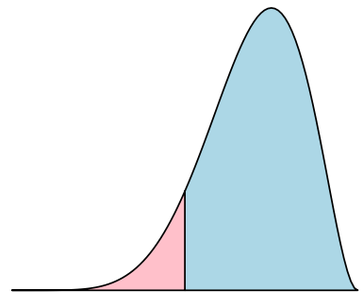
Neither scheme is any more meaningful than the other

Ordinal scale

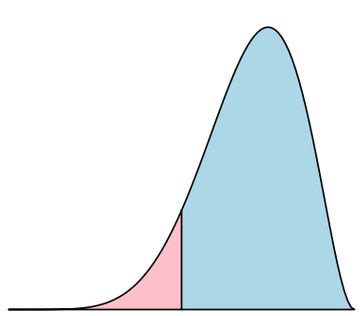
- There's a natural ordering, but differences aren't meaningful
- Ranking data: being punched < doing stats < eating ice cream



The numbers are informative here



Terrible	Meh	Awesome

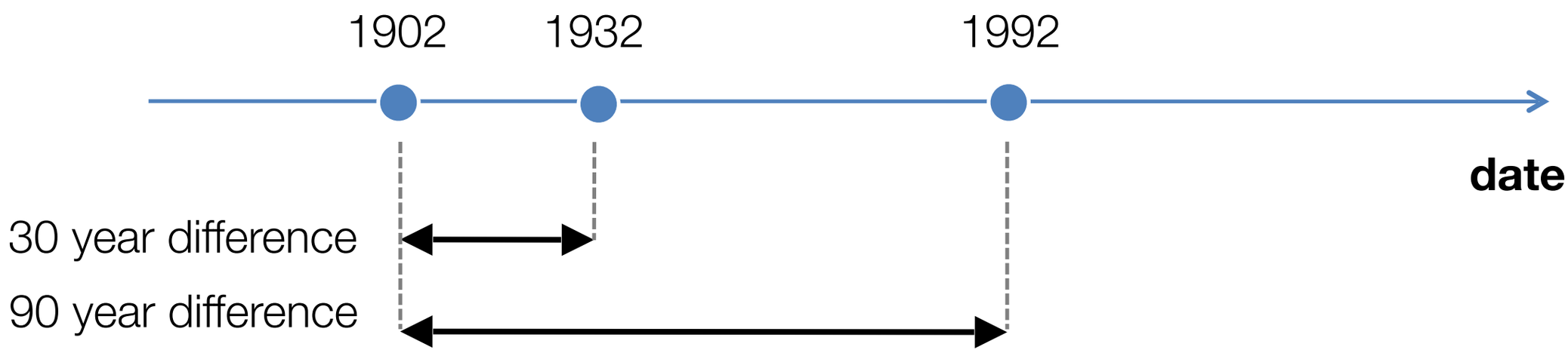


Terrible	Meh	Awesome

But we can’t tell the difference between these two situations…

Interval scale

- Numbers have natural ordering
- Differences between the numbers are meaningful
- Ratios between them are not



the big gap really is 3x the
size of the small one

Interval scale

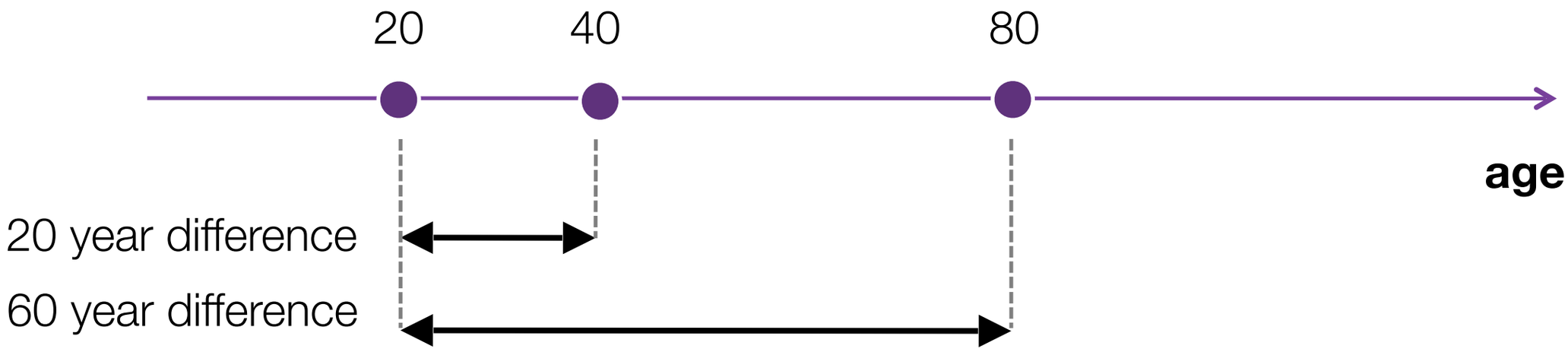
- Numbers have natural ordering
- Differences between the numbers are meaningful
- Ratios between them are not

1992 / 1902 = 1.047

We can’t say that the year of 1992
was “4.7% older” than year 1902

Ratio scale

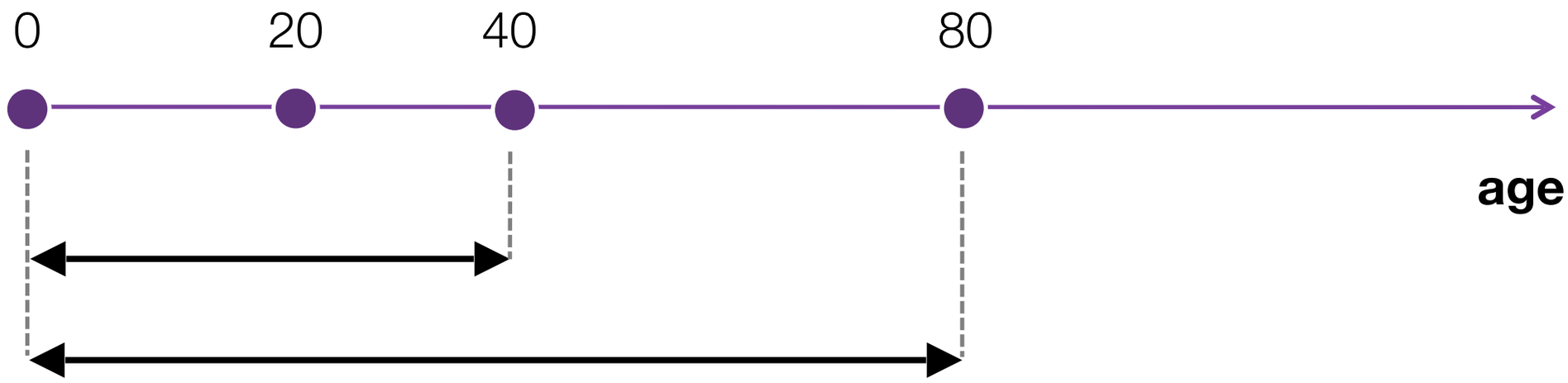
- Numbers have natural ordering, and “zero means zero”
- Differences between the numbers are meaningful
- Ratios between them are also meaningful



the big gap really is 3x the
size of the small one

Ratio scale

- Numbers have natural ordering, and “zero means zero”
- Differences between the numbers are meaningful
- Ratios between them are also meaningful

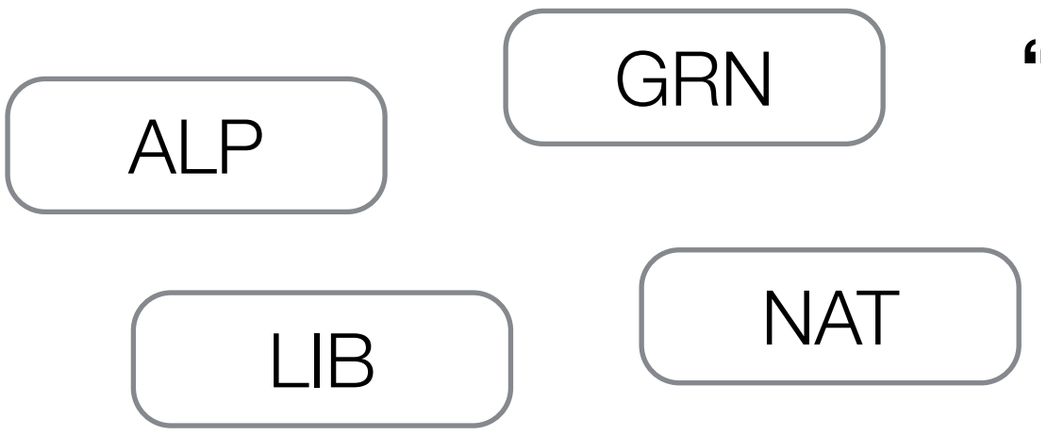


an 80 year old really is twice as
old as a 40 year old

Discrete variables

- Values come in specific categories, with no values “in between”
- e.g., “Year of birth”: 1982 and 1983 existed. 1983.32 didn't

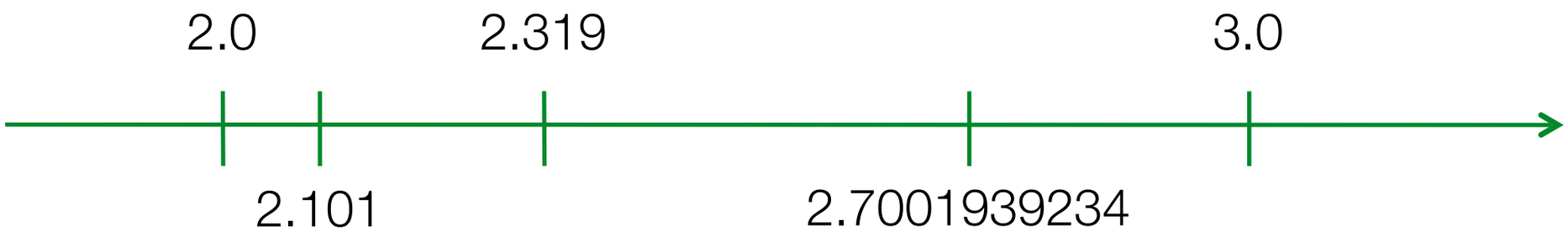
“Year of birth” is interval scale, and discrete



“Party voted for” is nominal scale, and discrete

Continuous variables

- Values vary smoothly; there's always something "in between"
- Response time is continuous: between 2.0s and 3.0s there's…
  - 2.101 seconds,
  - 2.319 seconds,
  - 2.7001939234 seconds, etc



Some are possible, others aren't

	continuous	discrete
nominal		✓
ordinal		✓
interval	✓	✓
ratio	✓	✓

The most important division

		continuous	discrete
categorical	nominal		✓
non-categorical	ordinal		✓
	interval	✓	✓
	ratio	✓	✓

Hey… categorical is like character variables in R, and non-categorical is like numeric!

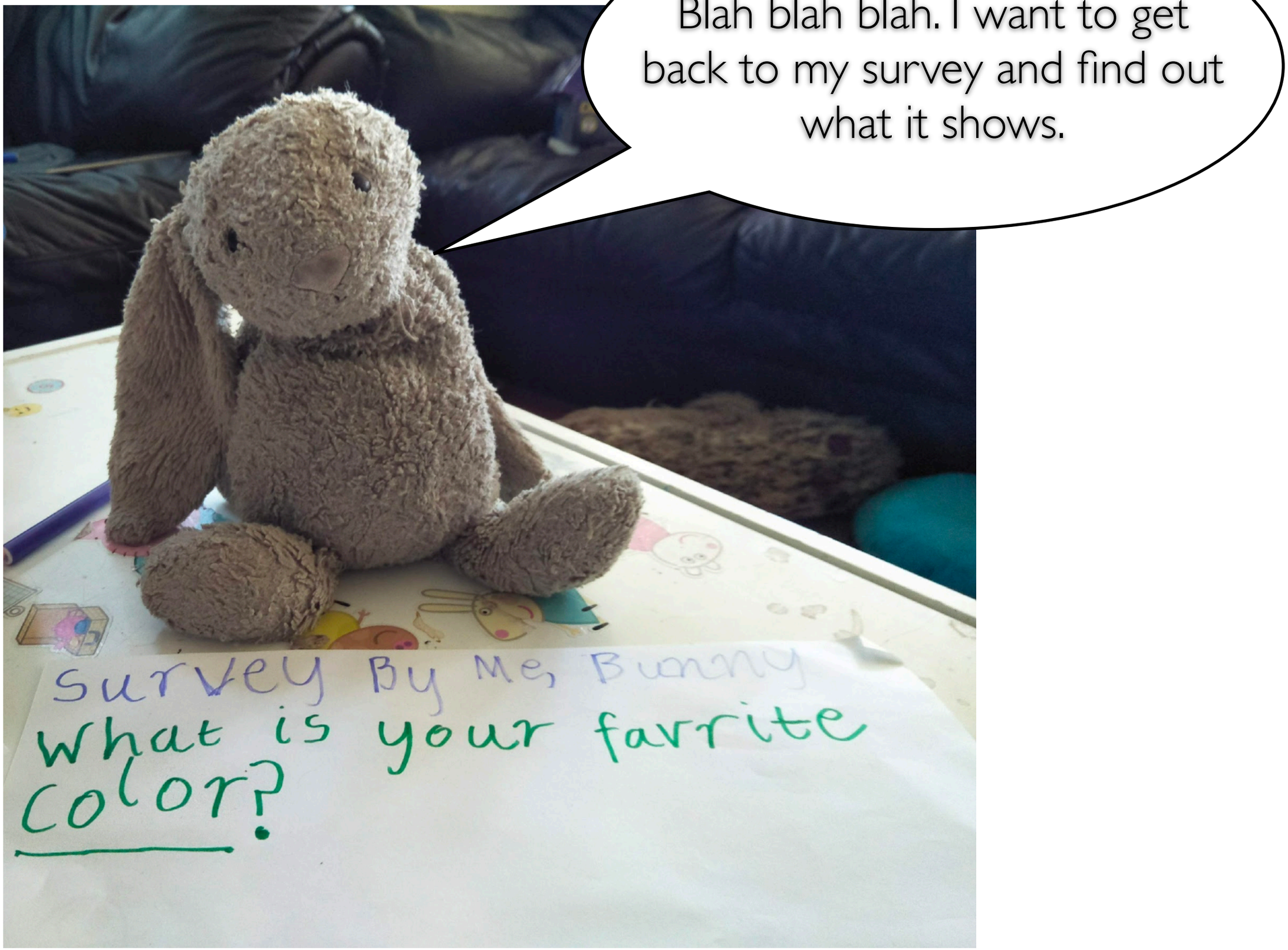
Variables can play different roles when analysing data

“A variable used to explain other variables”	**“A variable to be explained in terms of other variables”**
predictor	outcome
independent variable	dependent variable
treatment	response
etc…	etc…

Predictors and outcomes

- The “explanatory” relationship doesn’t have to be causal
  - smoking causes cancer
  - religiosity is associated with happiness
  - blondes have more fun

- The same variable can play multiple roles
  - religiosity is associated with happiness
  - happiness makes you live longer


Blah blah blah. I want to get back to my survey and find out what it shows.
Survey By Me, Bunny
What is your farrite color?

What is **reliability**?

- Measure the “same” thing twice: do you get the same answer?
  - If yes, your measurement is reliable
  - If no, then it’s not

- What do we mean by “the same thing”?
  - Test-retest reliability
  - Inter-rater reliability
  - Internal consistency reliability

Test-retest reliability

- Try the measurement **one time**, try it again **another time**
- Do you get the same answers?

“What is your favourite colour?”

Yesterday:

“Black”

“What is your favourite colour?”

Today:

“Pink”

Inter-rater reliability

- Try having different **people** make the measurements
- Do you get the same answers?

“What is your favourite colour?”

“Grey”

“What is your favourite colour?”

“Brown”

Internal consistency reliability

- Try two **theoretically equivalent** versions within the same measure: do "equivalent" subparts agree?
- Do you get the same answers?

"What's your favourite colour?"

"What colour do you like best?"

"What colour would you choose?"

"Pink"

"Black"

Exercises

1. For Bunny's survey question about colour, what is the construct, the measure, and the observation?

2. For each of Bunny's survey questions, say what kind of variable it yields (e.g., nominal, etc).

3. How do you think reliability of these survey questions might differ between themselves? Consider each of the kinds of reliability, and say how Bunny might go about measuring them.