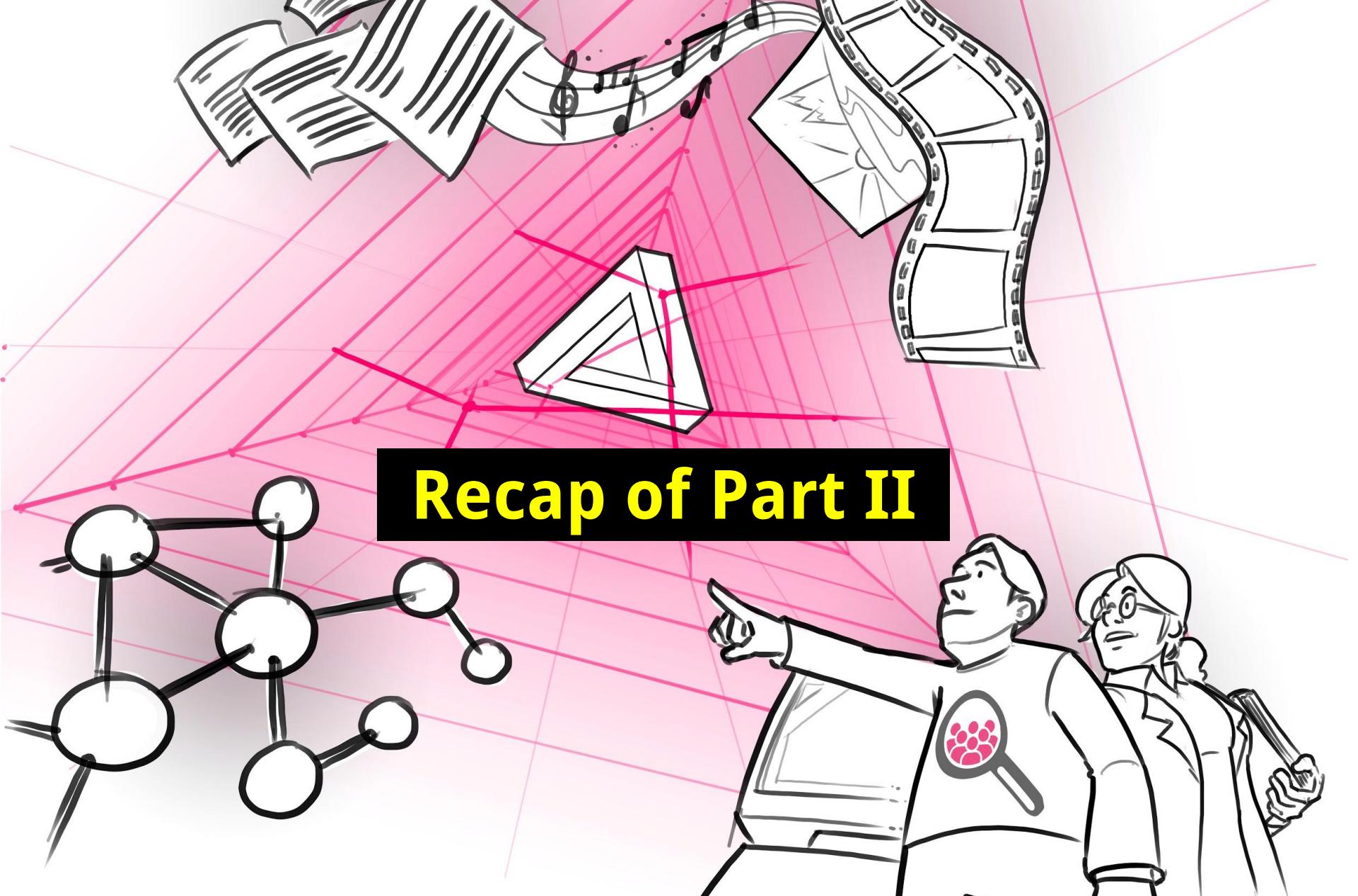


CROWDTRUTH TUTORIAL

Part III: Quality Assessment, Data Processing & Metrics

Lora Aroyo, Anca Dumitrache, Oana Inel, Chris Welty



Recap of Part II

CROWD TASKS

Describe The People In The Video

- Watch the video in full at least once.



OPEN ENDED

- Enter tags relevant to the people that appear in the video:
 - Click ADD+ or press Enter to add your tag to the list

Your tags: (click to remove)

Woman Repair Ship

Identify People In Video By Highlighting Text

- Please look at this video in full.



SELECT WORDS FROM TEXT

- Highlight the WORDS or PHRASES in the text that identify or refer to PEOPLE in the video.

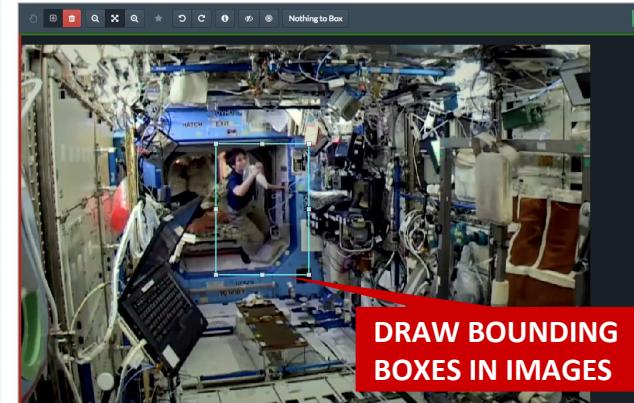
Italian astronaut samantha cristoforetti uploads a strange photograph she's taken

[x] Italian astronaut
 [x] astronaut
 [x] samantha cristoforetti
 [x] strange photograph

Draw a bounding box around the persons in the image.

Keep in mind that:

- It is possible to draw more than one bounding box if necessary
- It is also possible to choose "Nothing to box" when no person appears in the image.



DRAW BOUNDING BOXES IN IMAGES

Are There People In The Video?

- Look at the video carefully.



BINARY CHOICE

- Are there any people in the video?

Select your answer:

Yes No
- Submit your answer.

Are There People In The Video?

- Look at the video carefully.



TERNARY CHOICE

- Are there any people in the video?

Select your answer:

Yes No Not Clear
- Submit your answer.

MULTIPLE CHOICE

Describe The Video By Selecting Tags Relevant To People

- Look at the video carefully.



MULTIPLE CHOICE

- Select all tags that identify or refer to people explicitly shown in the video.

Selected by me:

Archeologist	Architect	Artist
<input checked="" type="checkbox"/> Astronaut	<input type="checkbox"/> Athlete	<input type="checkbox"/> BusinessPerson
Celebrity	Chef	Criminal
<input checked="" type="checkbox"/> Engineer	<input type="checkbox"/> Farmer	<input type="checkbox"/> FictionalCharacter
Journalist	Judge	Lawyer
MilitaryPerson	Model	Monarch
Philosopher	Politician	Presenter
Producer	Psychologist	<input checked="" type="checkbox"/> Scientist
SportsManager	Writer	
<input type="checkbox"/> Other	<input type="checkbox"/> None	

REDUCED MULTIPLE CHOICE

Describe The Video By Selecting Tags Relevant To People

- Look at the video carefully.



REDUCED MULTIPLE CHOICE + OPEN INPUT

- Select all tags that identify or refer to people explicitly shown in the video.

Selected by me:

Italian	<input checked="" type="checkbox"/> Astronaut	<input checked="" type="checkbox"/> Samantha
Strange	<input checked="" type="checkbox"/> Photograph	Industry
Grinder	Production	<input checked="" type="checkbox"/> Technology
Equipment	Central Processing Unit	
Hardware	Technology	Business
<input checked="" type="checkbox"/> Computer		

Added by me: (click to remove)

Type your tag here

ADD+

Describe The Video By Selecting Tags Relevant To People

- Look at the video carefully.



REDUCED MULTIPLE CHOICE + OPEN INPUT

- Select all tags that identify or refer to people explicitly shown in the video.

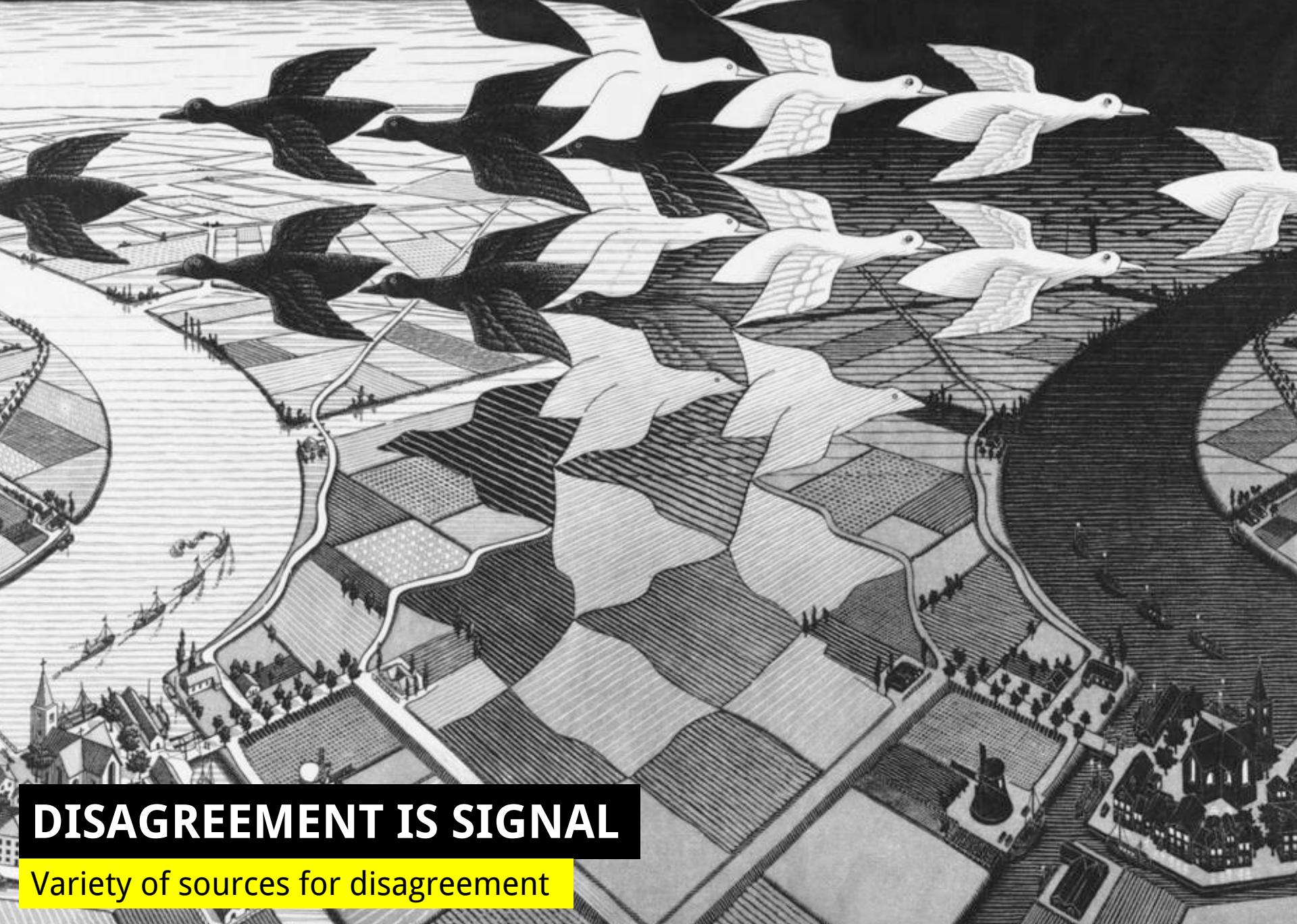
Selected by me:

Italian	<input checked="" type="checkbox"/> Astronaut	<input checked="" type="checkbox"/> Samantha
Strange	<input checked="" type="checkbox"/> Photograph	Industry
Grinder	Production	<input checked="" type="checkbox"/> Technology
Technology	Central Processing Unit	
Technology	Business	
<input checked="" type="checkbox"/> Computer		

Added by me: (click to remove)

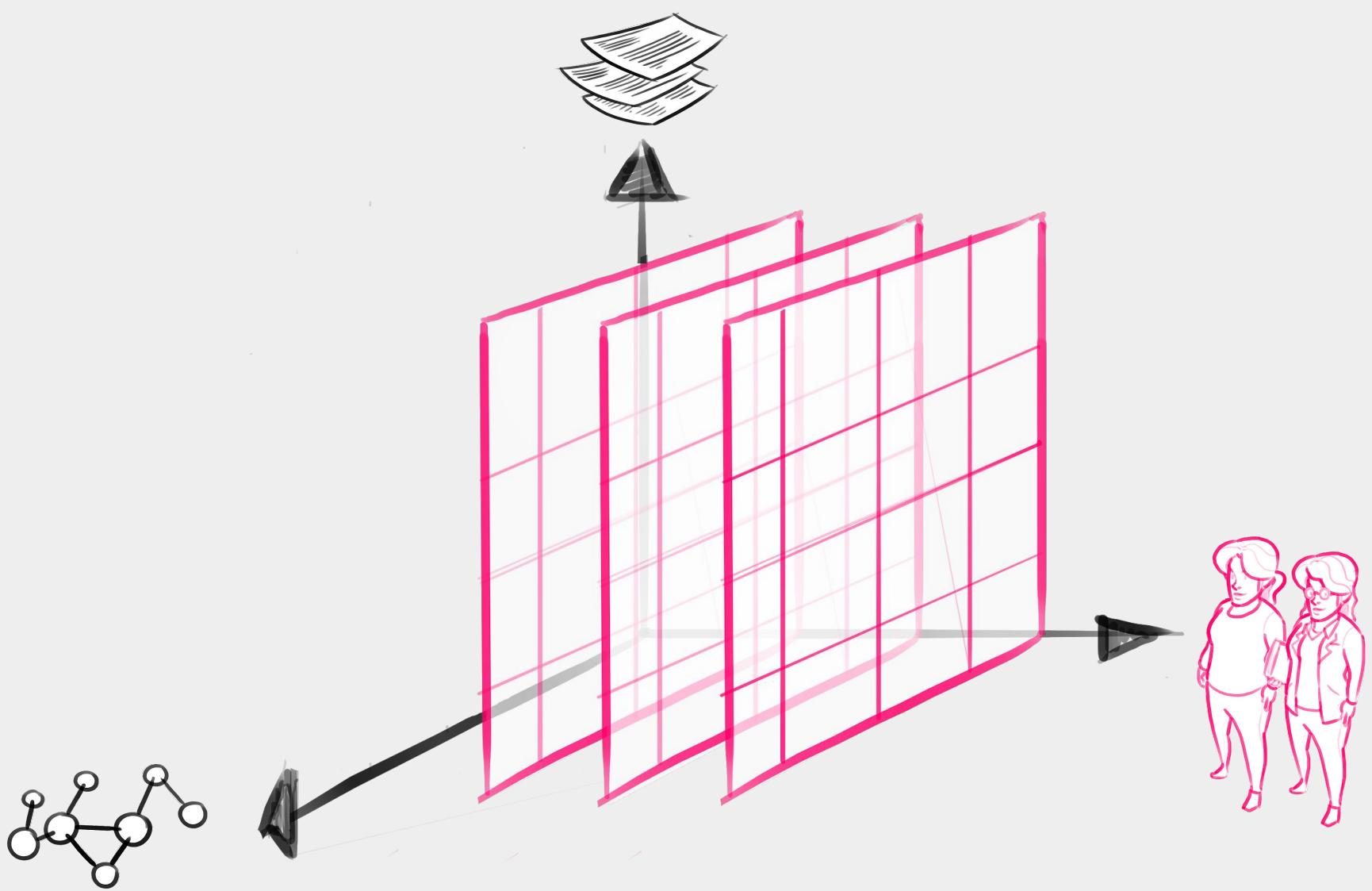
Type your tag here

ADD+



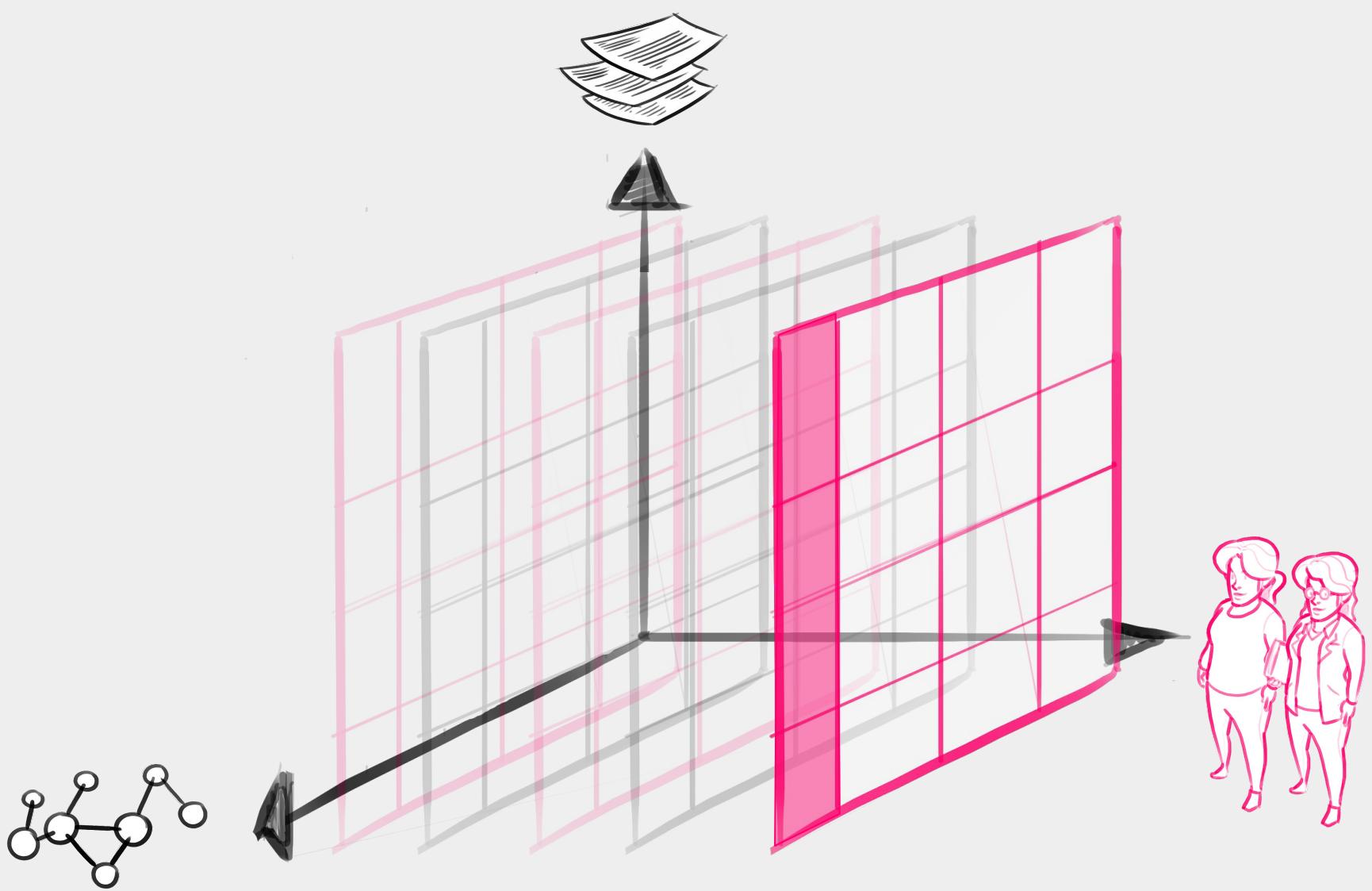
DISAGREEMENT IS SIGNAL

Variety of sources for disagreement



DISAGREEMENT IS SIGNAL

Source 1: People's bias & perspective



DISAGREEMENT IS SIGNAL

Source 1: Worker systematically give same answer

DO THE SENTENCES EXPRESS A TREAT RELATION?

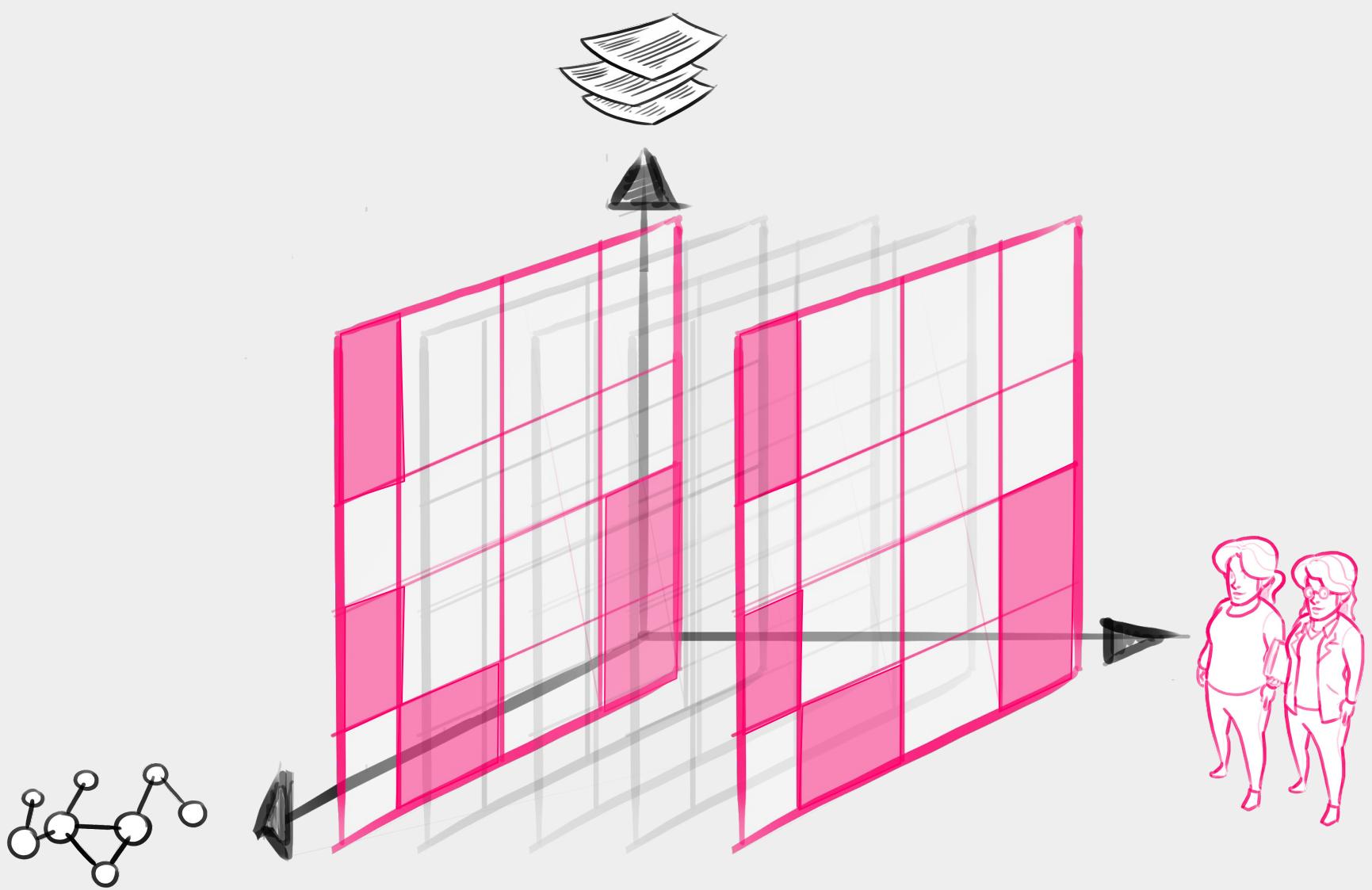
ANTIBIOTICS are the first line treatment for vectors of **TYPHUS**.

QUININE is not a reliable cure for **MALARIA**.

Worker	Sentence (1)	Sentence (2)
Worker 1	yes	no
Worker 2	yes	no
Worker 3	yes	
Worker 4		no
Worker 5	no	yes

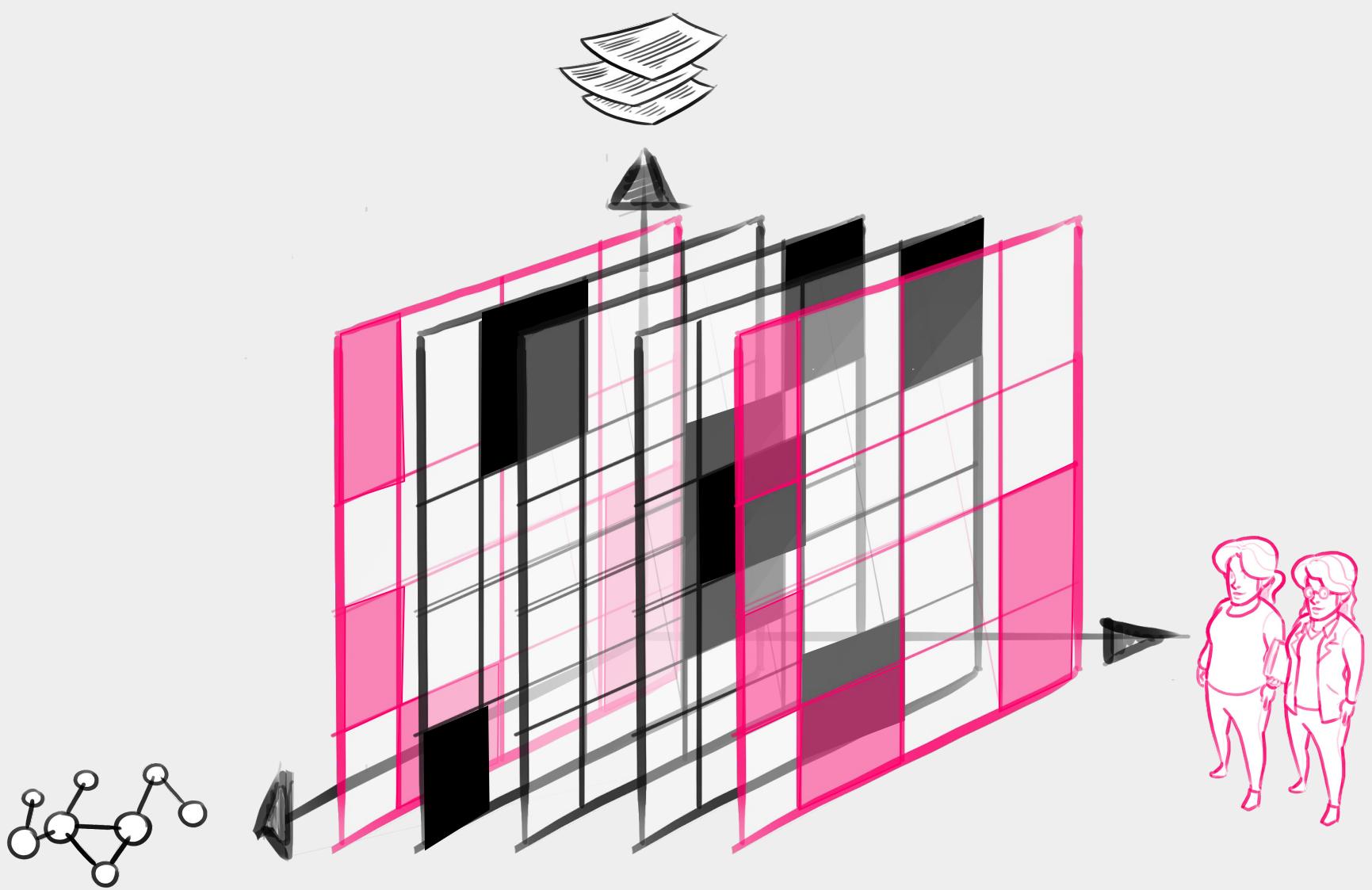
DISAGREEMENT CAUSE

Source 1: Low Quality Workers



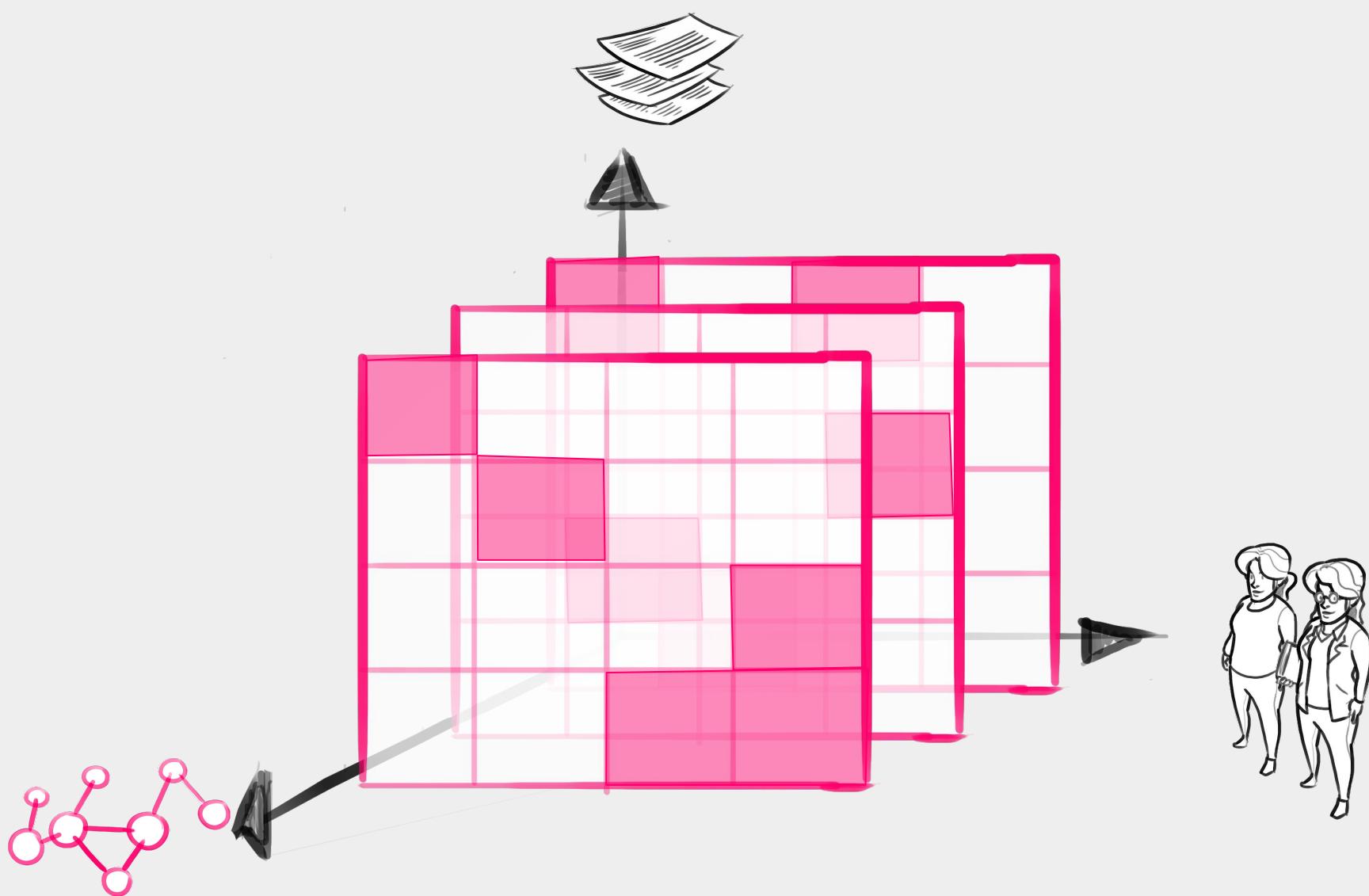
DISAGREEMENT IS SIGNAL

Source 1: Worker communities of expertise



DISAGREEMENT IS SIGNAL

Source 1: Worker communities of expertise



DISAGREEMENT IS SIGNAL

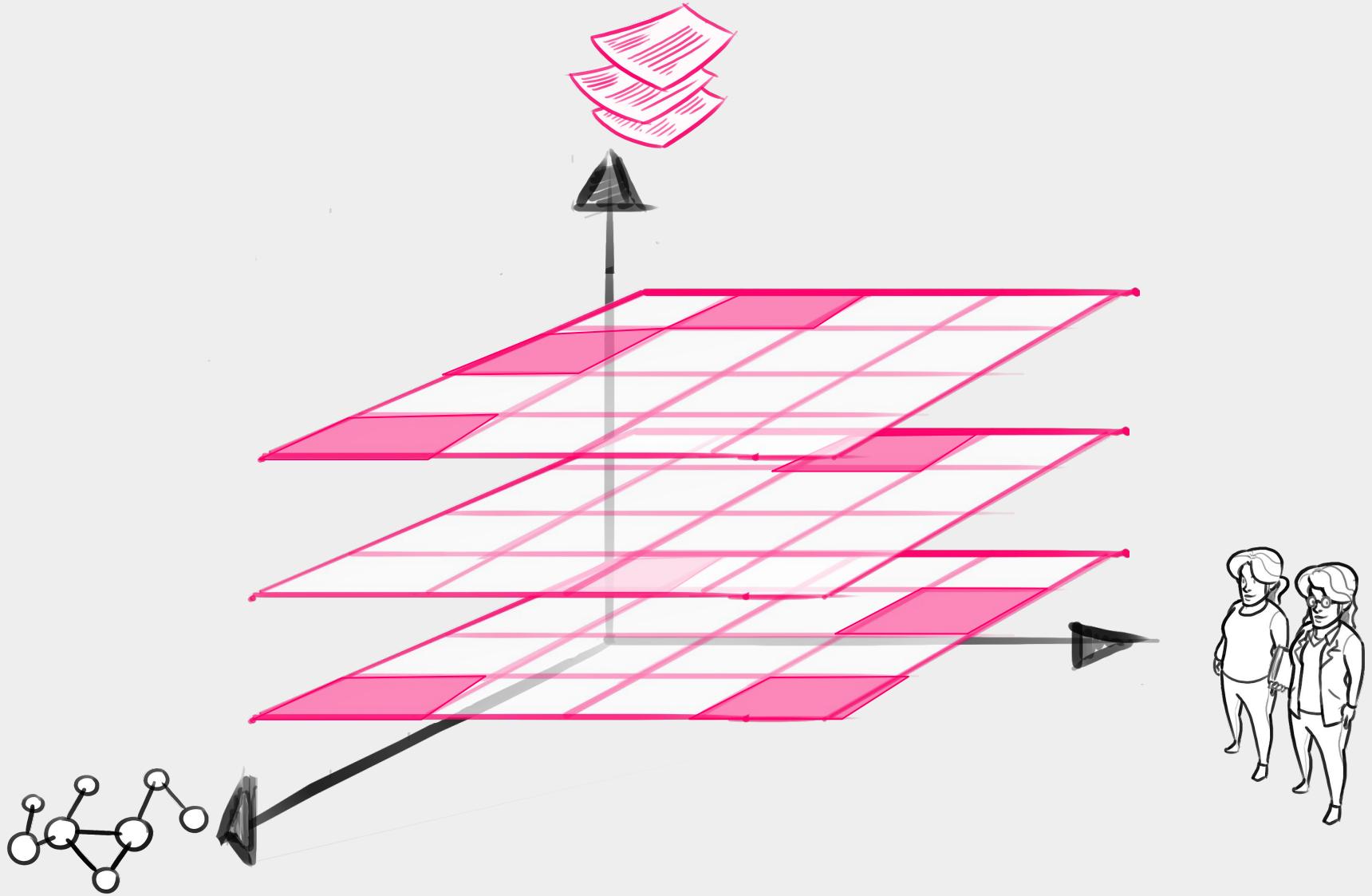
Source 2: Target semantics

WHAT IS THE RELATION EXPRESSED? TREAT OR PREVENT?

METHYLERGOMETRINE is a blood vessel constrictor most commonly used to control excessive **BLEEDING**.

DISAGREEMENT CAUSE

Source 2: Target semantics



DISAGREEMENT IS SIGNAL

Source 3: Sentence Clarity

DO THE SENTENCES EXPRESS A TREAT RELATION? BETWEEN CHLOROQUINE AND MALARIA ?

Rheumatoid arthritis and **MALARIA** have been treated with **CHLOROQUINE** for decades.

95%

For prevention of malaria, use only in individuals traveling to malarious areas where **CHLOROQUINE** resistant P. falciparum **MALARIA** has not been reported.

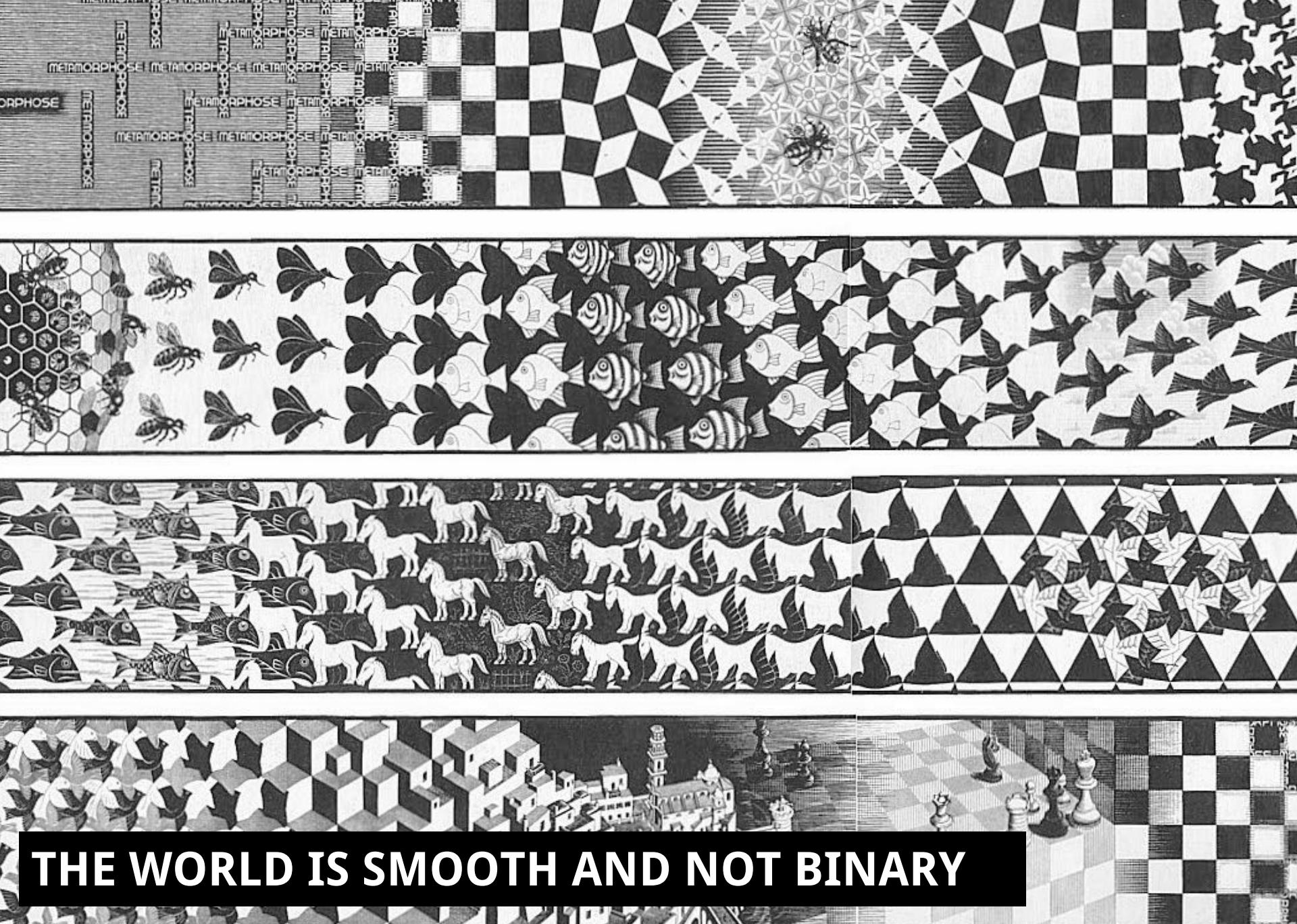
75%

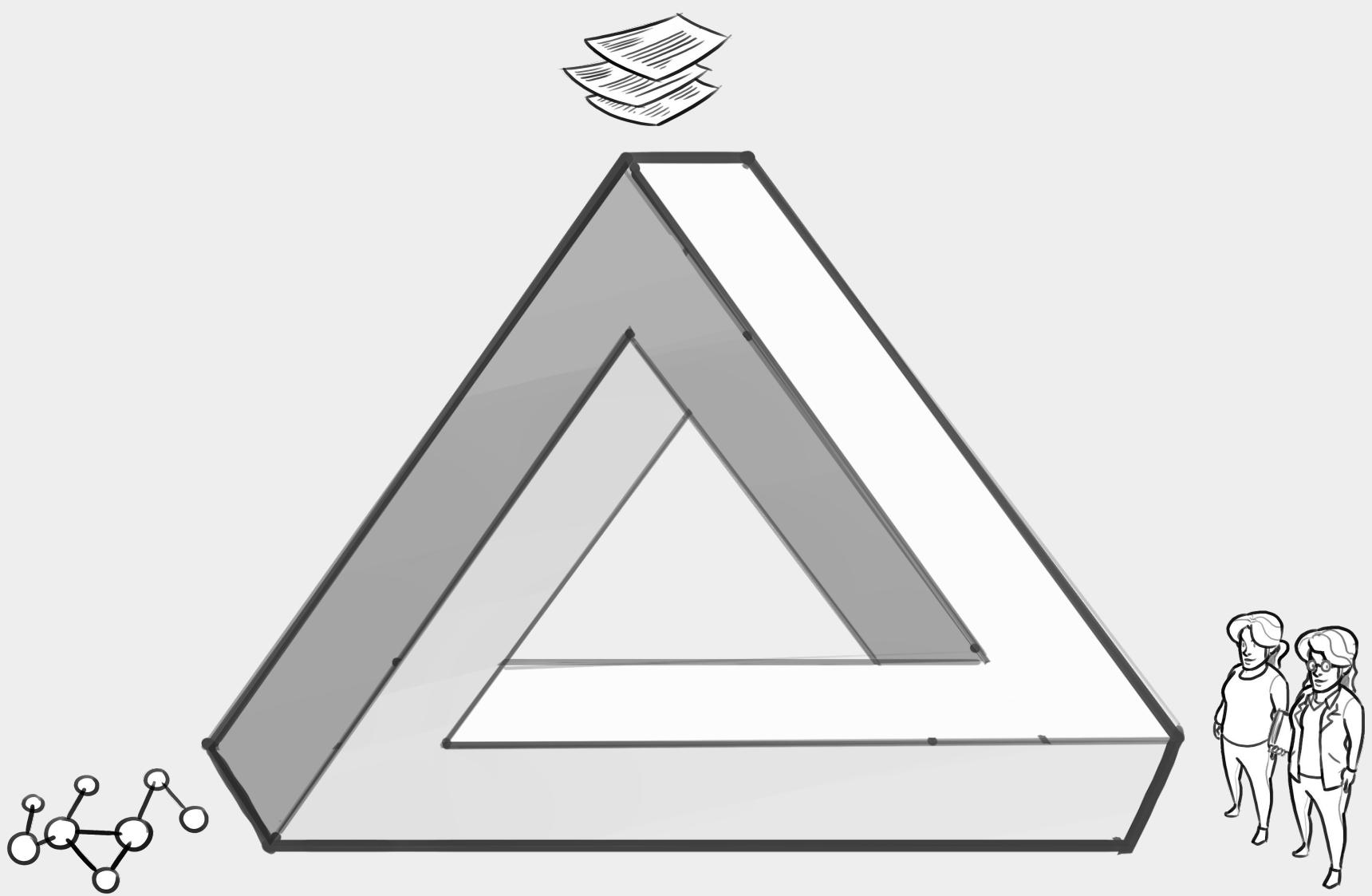
Among 56 subjects reporting to a clinic with symptoms of **MALARIA** 53 (95%) had ordinarily effective levels of **CHLOROQUINE** in blood.

50%

DISAGREEMENT CAUSE

Source 3: Sentence Clarity

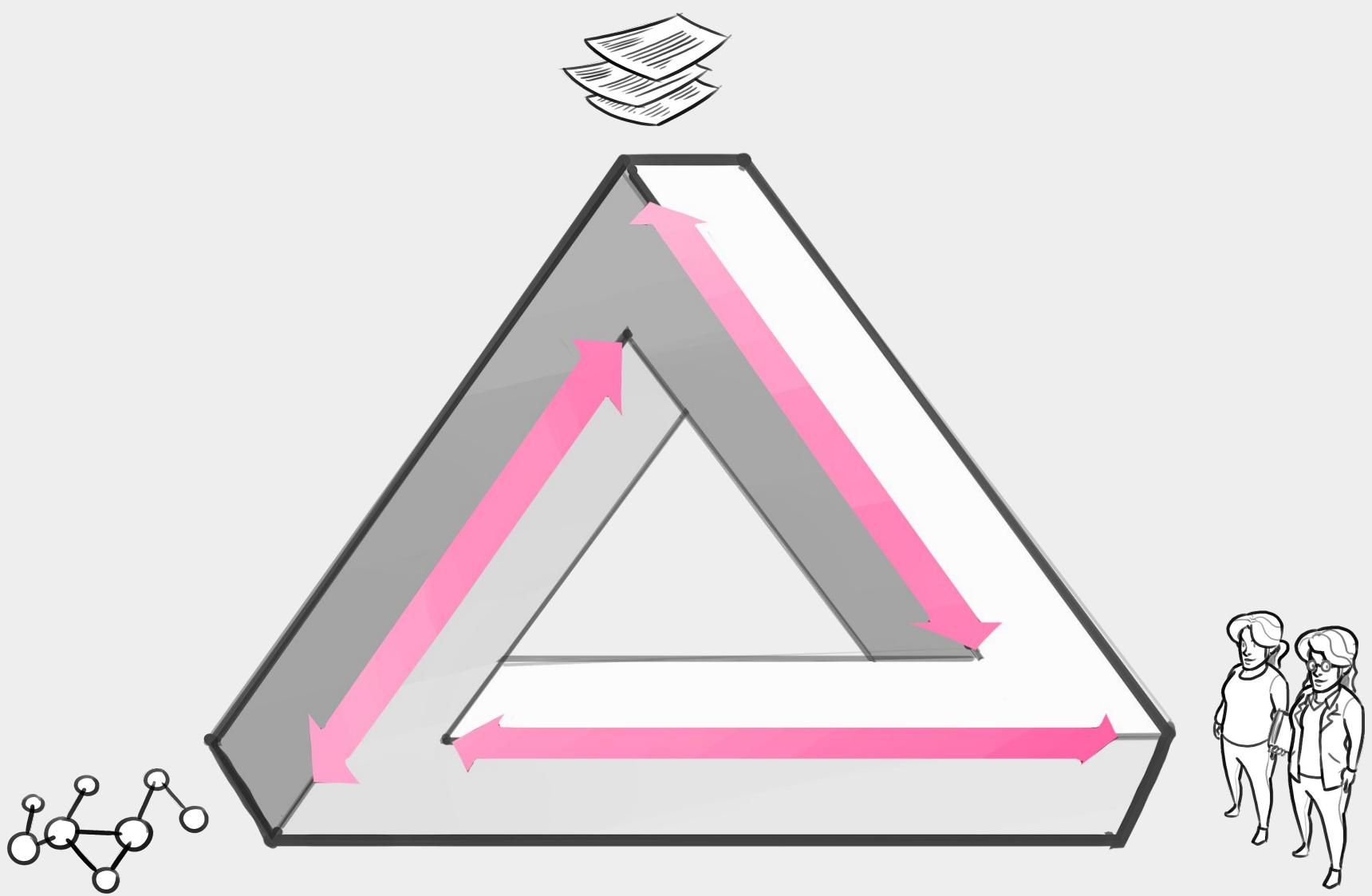




TRIANGLE OF MEANING

Model of semantic interpretation

*(Ogden & Richards, 1936)
The Meaning of Meaning*



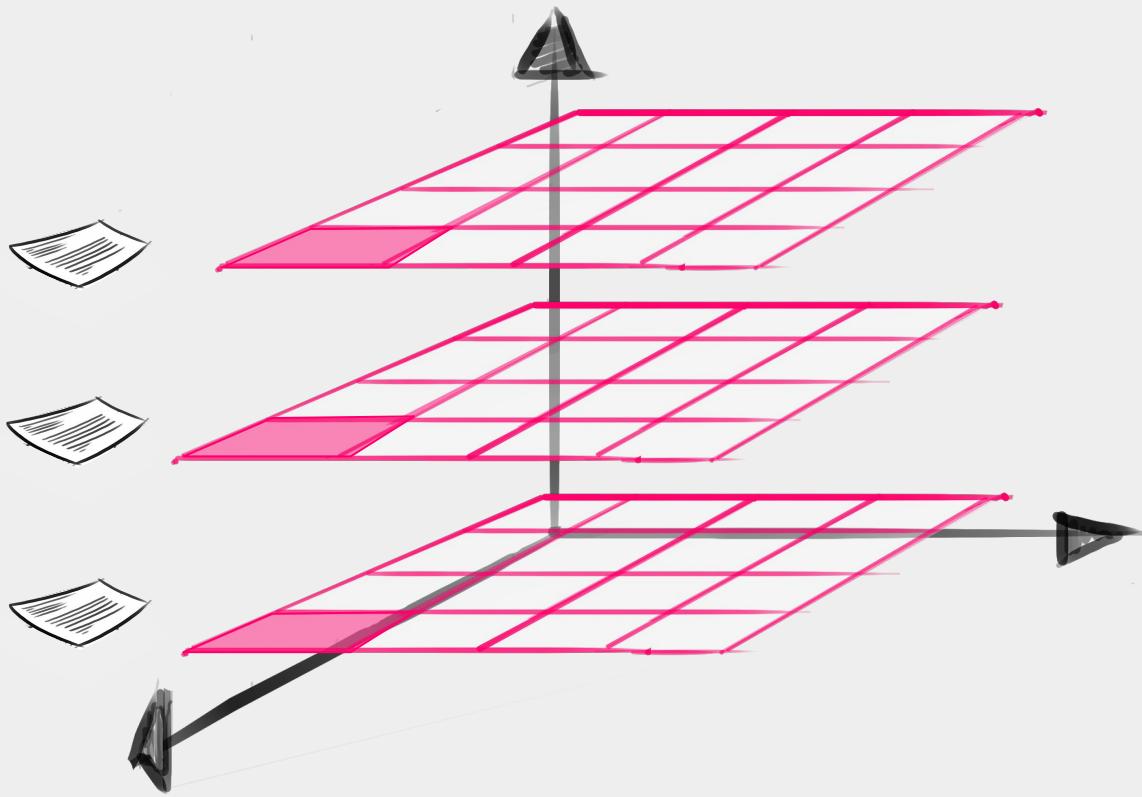
TRIANGLE OF MEANING

Model of semantic interpretation



TRIANGLE OF MEANING

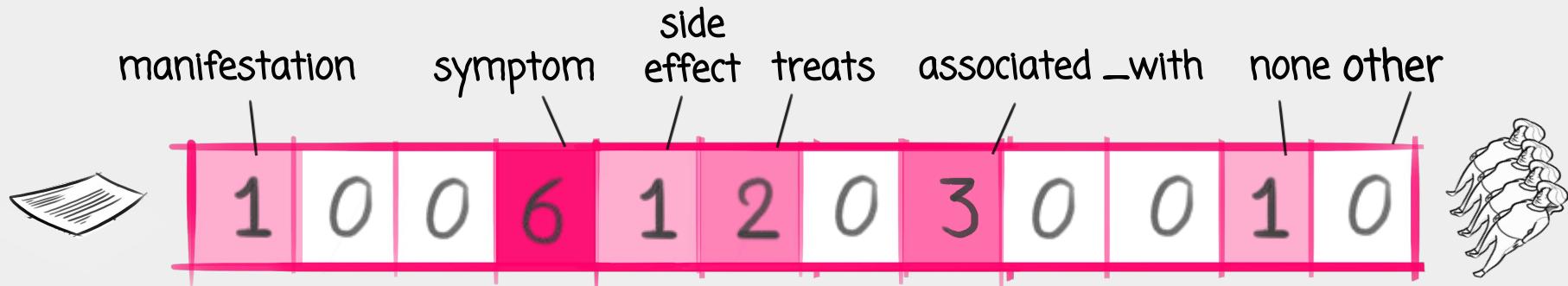
CrowdTruth metrics for quality assessment



QUALITY ASSESSMENT

Spam

Among 56 subjects reporting to a clinic with symptoms of **MALARIA** 53 (95%) had ordinarily effective levels of **CHLOROQUINE** in blood.



QUALITY ASSESSMENT

Sentence ambiguity

manifestation contraindicates treats associated_with none other



1	0	3	0	0	6	0	0	2	0	1	4
---	---	---	---	---	---	---	---	---	---	---	---



manifestation contraindicates treats associated_with none other



2	0	2	0	0	8	0	0	0	0	1	0
---	---	---	---	---	---	---	---	---	---	---	---



treats

other



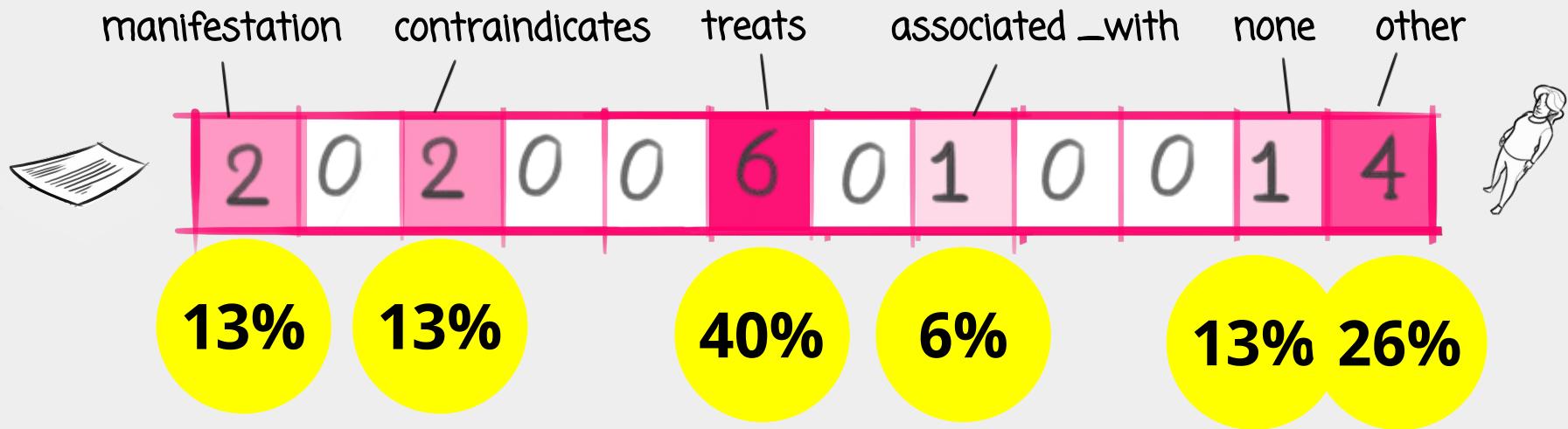
0	0	0	0	0	14	0	0	0	0	0	1
---	---	---	---	---	----	---	---	---	---	---	---



TREATS RELATION

Yes or No?

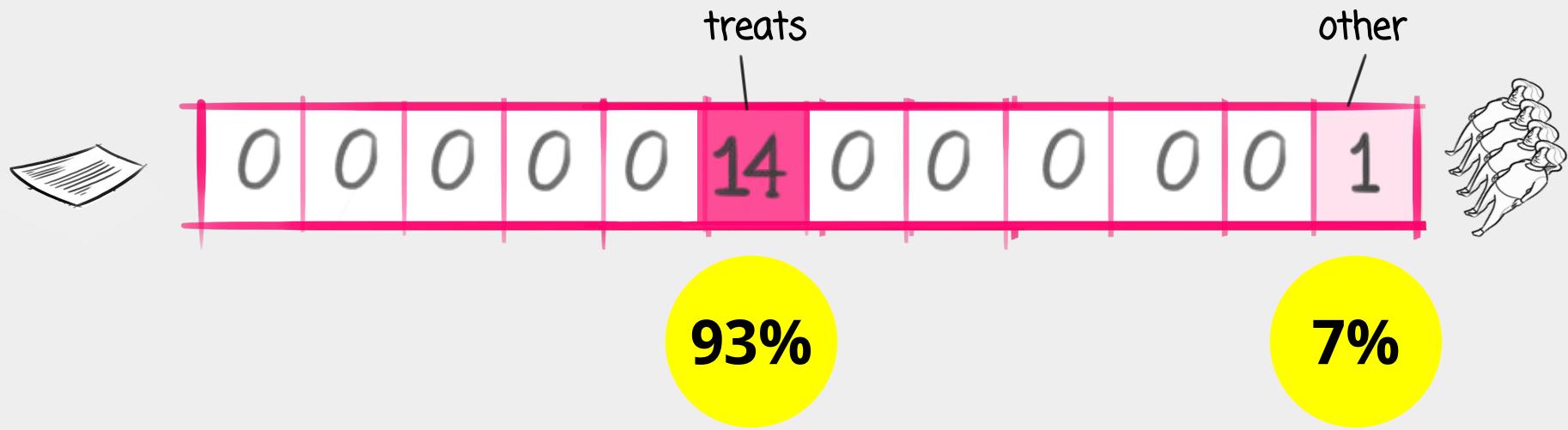
For prevention of malaria, use only in individuals traveling to
malarious areas where **CHLOROQUINE** resistant P.
falciparum **MALARIA** has not been reported.



CROWDTRUTH METRICS

Agreement as percentage

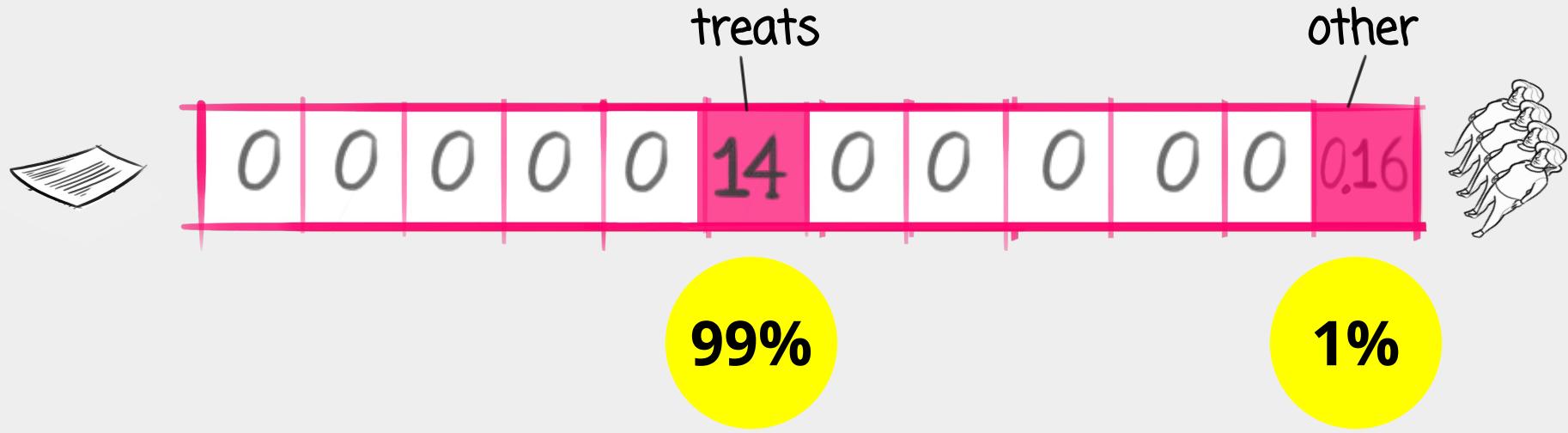
For prevention of malaria, use only in individuals traveling to
malarious areas where **CHLOROQUINE** resistant P.
falciparum **MALARIA** has not been reported.



CROWDTRUTH METRICS

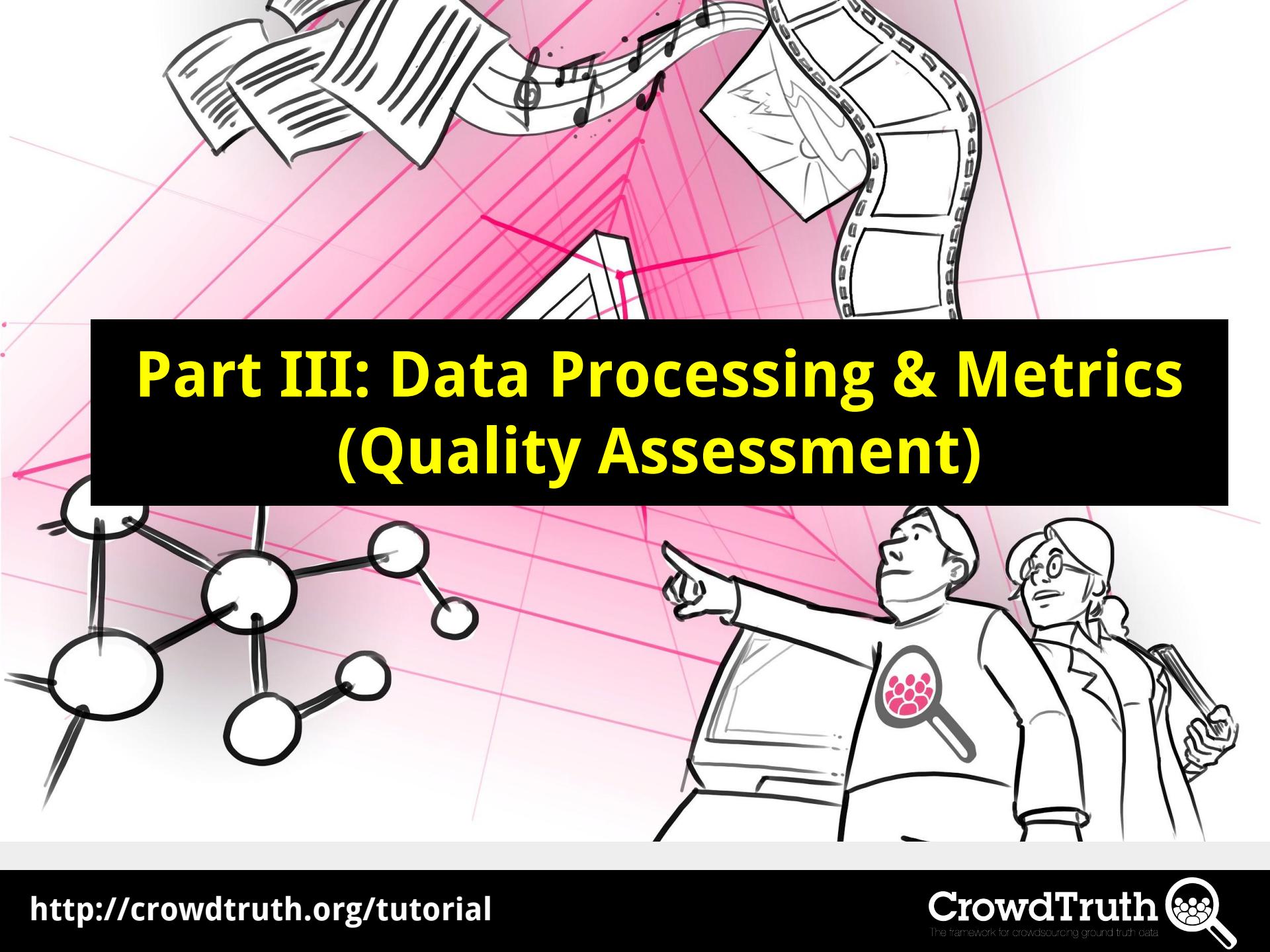
Applying all sides of the triangle

For prevention of malaria, use only in individuals traveling to
malarious areas where **CHLOROQUINE** resistant P.
falciparum **MALARIA** has not been reported.

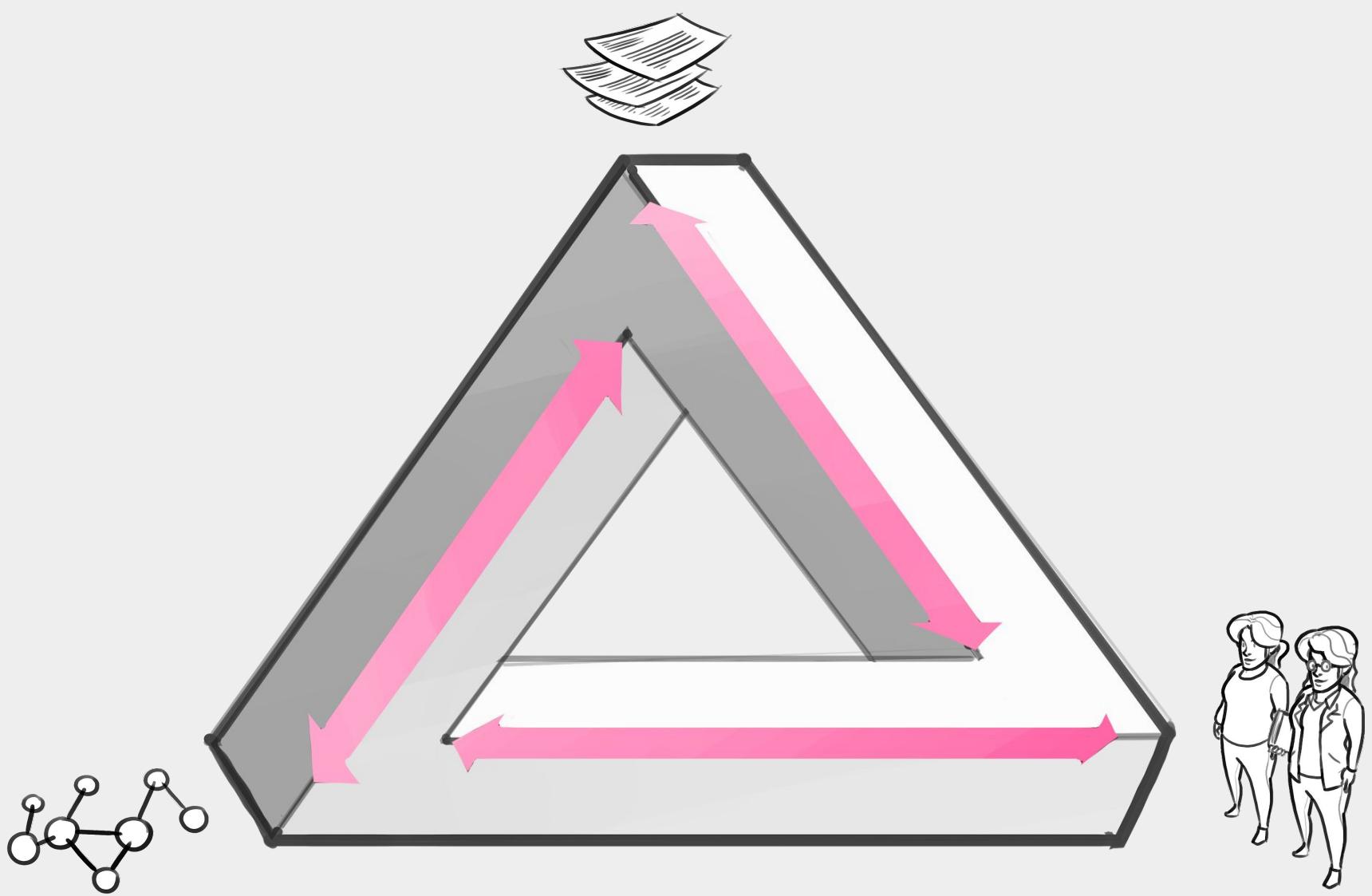


CROWDTRUTH METRICS

Applying all sides of the triangle



Part III: Data Processing & Metrics (Quality Assessment)



TRIANGLE OF MEANING

Model of semantic interpretation

APPLYING ALL SIDES OF THE TRIANGLE

Unit quality:

overall worker agreement over one input unit
(e.g. sentence, video)

Worker quality:

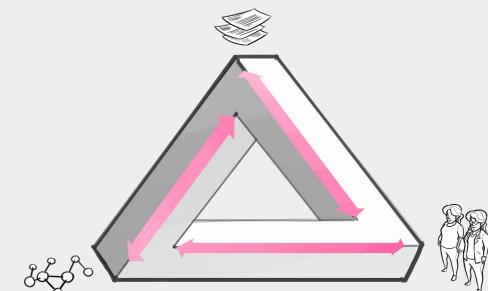
overall agreement between single worker and the other workers

Annotation quality:

overall worker agreement over one annotation

CROWDTRUTH QUALITY METRICS

Ambiguity at any corner disseminates in the other corners



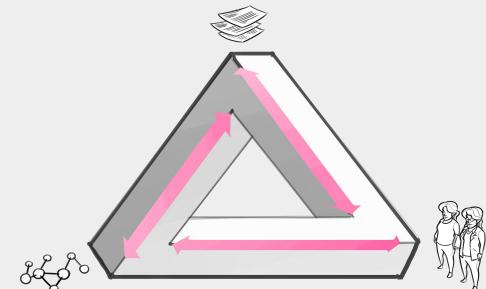
APPLYING ALL SIDES OF THE TRIANGLE

Calculated using **average cosine similarity** over **worker & unit vectors**.

Quality metrics are **inter-dependent & continuously updated**:
e.g. ambiguous units count less when determining worker quality, because disagreement is expected

CROWDTRUTH QUALITY METRICS

Ambiguity at any corner disseminates in the other corners



Closed Task

Merkel is serving as Chancellor of Germany.

In the sentence above, what are the relations expressed between the terms in bold? Pick from the list below:

- Merkel **employed by** Germany
- Merkel **born in** Germany
- Merkel **has origin** Germany
- none**

media units

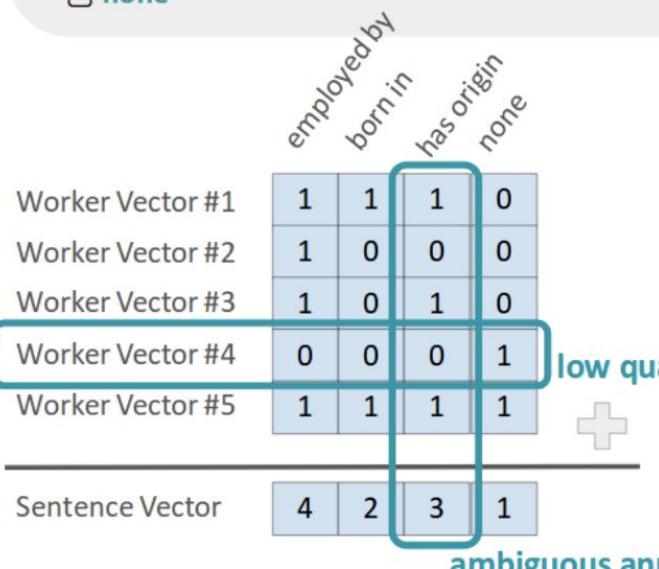
annotations

Open Task



In the textbox below, write comma-separated tags describing the image.

Tags:

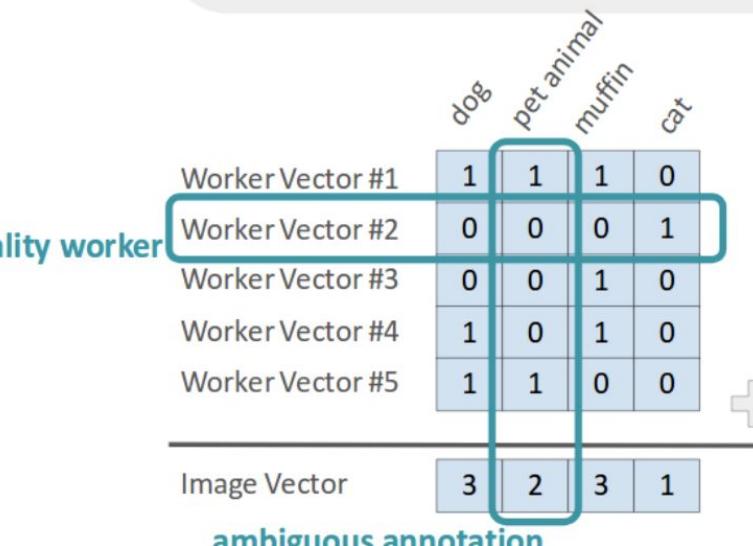


low quality worker

low quality worker



ambiguous annotation



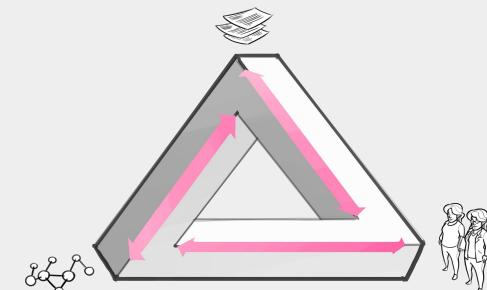
low quality worker

ambiguous annotation



CROWDTRUTH QUALITY METRICS

Worker vs. Unit Annotation Vectors



Unit Quality Score (UQS): measures the overall worker agreement over one input unit

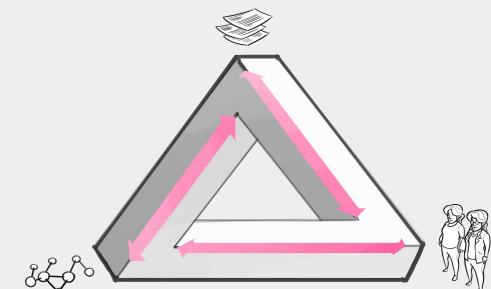
- $UQS(u)$ = average cosine similarity between all *worker vectors for unit u*, weighted by *worker quality (WQS)* & *annotation quality (AQS)*

$$SQS(s) = \frac{\sum_{w1, w2 \in \text{workers for } s} \text{weighted cos sim}(\vec{w1}, \vec{w2}, FQS) \cdot \text{worker quality}(w1, w2)}{\sum_{w1, w2 \in \text{workers for } s} \text{worker quality}(w1, w2)}$$

The weights mean that **workers & annotations with lower quality** have less impact on the **unit quality**

CROWDTRUTH QUALITY METRICS

Measuring the quality of the units



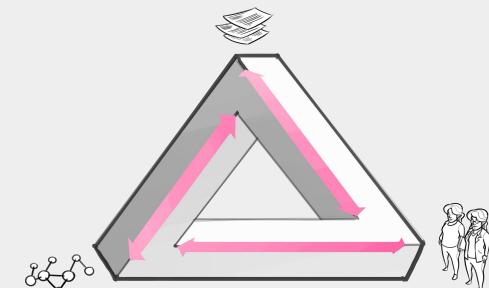
Worker Quality Score (WQS): measures the overall agreement of one crowd worker with the other workers:

- $WQS(i)$ = product of 2 metrics
 - worker-worker agreement $WWA(i)$
 - worker-media unit agreement $WUA(i)$

$$WQS(i) = WUA(i) \cdot WWA(i).$$

CROWDTRUTH QUALITY METRICS

Measuring the quality of workers



Worker-Worker Agreement (WWA): measures the average agreement between a worker and all other workers

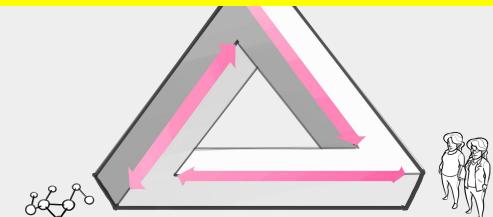
- indicates if there are *consistently like minded workers* - to identify *communities of expertise*
- $WWA(i)$ = average cosine similarity between annotation vectors of:
 - **worker i**
 - all **other workers** that worked on the same units as worker i, weighted by the **worker, unit** and **annotation qualities**.

$$WWA(i) = \frac{\sum_{j,u} WorkVecWcos(i, j, u) WQS(j) UQS(u)}{\sum_{j,u} WQS(j) UQS(u)},$$
$$\forall j \in workers(u \in units(i)), i \neq j.$$

The weights mean that **workers, units & annotations with lower quality** have less impact on the **worker quality**

CROWDTRUTH QUALITY METRICS

Measuring the quality of workers



Worker-Media Unit Agreement (WUA): measures the similarity between the annotations of a worker and the aggregated annotations of the rest of the workers.

- calculates the agreement with the consensus over all workers.
- $WUA(i)$ = average cosine similarity between:
 - annotation vectors of *worker i*
 - *unit vectors* for the media units worker i worked on, weighted by the *media unit (UQS)* and *annotation quality (AQS)*.

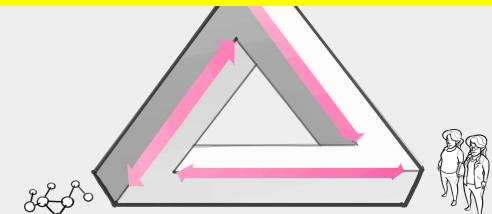
$$WUA(i) = \frac{\sum_{u \in units(i)} WorkUnitWcos(u, i) UQS(u)}{\sum_{u \in units(i)} UQS(u)},$$

$$WorkUnitWcos(u, i) = Wcos(WorkVec(i, u), \\ MediaUnitVec(u) - WorkVec(i, u))$$

The weights mean that **units & annotations with lower quality** have less impact on the **worker quality**.

CROWDTRUTH QUALITY METRICS

Measuring the quality of workers



Annotation Quality Score (AQS): measures the agreement over an annotation in all media units that it appears:

- only applicable to closed tasks, where the same annotation set is used for all input media units
- based on $P_a(i|j) = \text{probability that if } \text{worker } j \text{ annotates } a \text{ in a media unit, worker } i \text{ will also annotate } a$
- $AQS(a) = \text{weighted average of } P_a(i|j) \text{ for all possible pairs of workers } i \& j.$

$$AQS(a) = \frac{\sum_{i,j} WQS(i) WQS(j) P_a(i|j)}{\sum_{i,j} WQS(i) WQS(j)},$$

$\forall i, j \text{ workers, } i \neq j.$

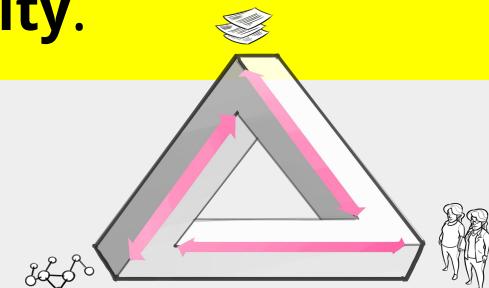
$$P_a(i|j) = \frac{\sum_u UQS(u) WorkVec(i, s)[a] WorkVec(j, s)[a]}{\sum_u UQS(u) WorkVec(j, u)(r)},$$

$\forall u \in \text{units}(i) \cap \text{units}(j).$

The weights mean that **units & workers** with **lower quality** have less impact on the **annotation quality**.

CROWDTRUTH QUALITY METRICS

Measuring quality of annotations



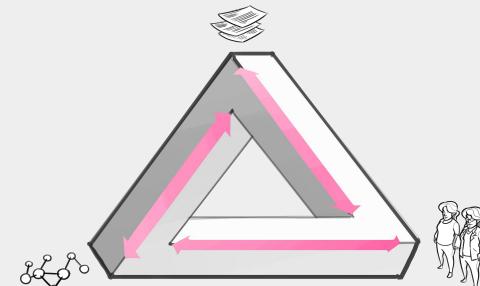
Unit-Annotation Score (UAS): measures the degree of clarity with which an annotation is expressed in a unit.

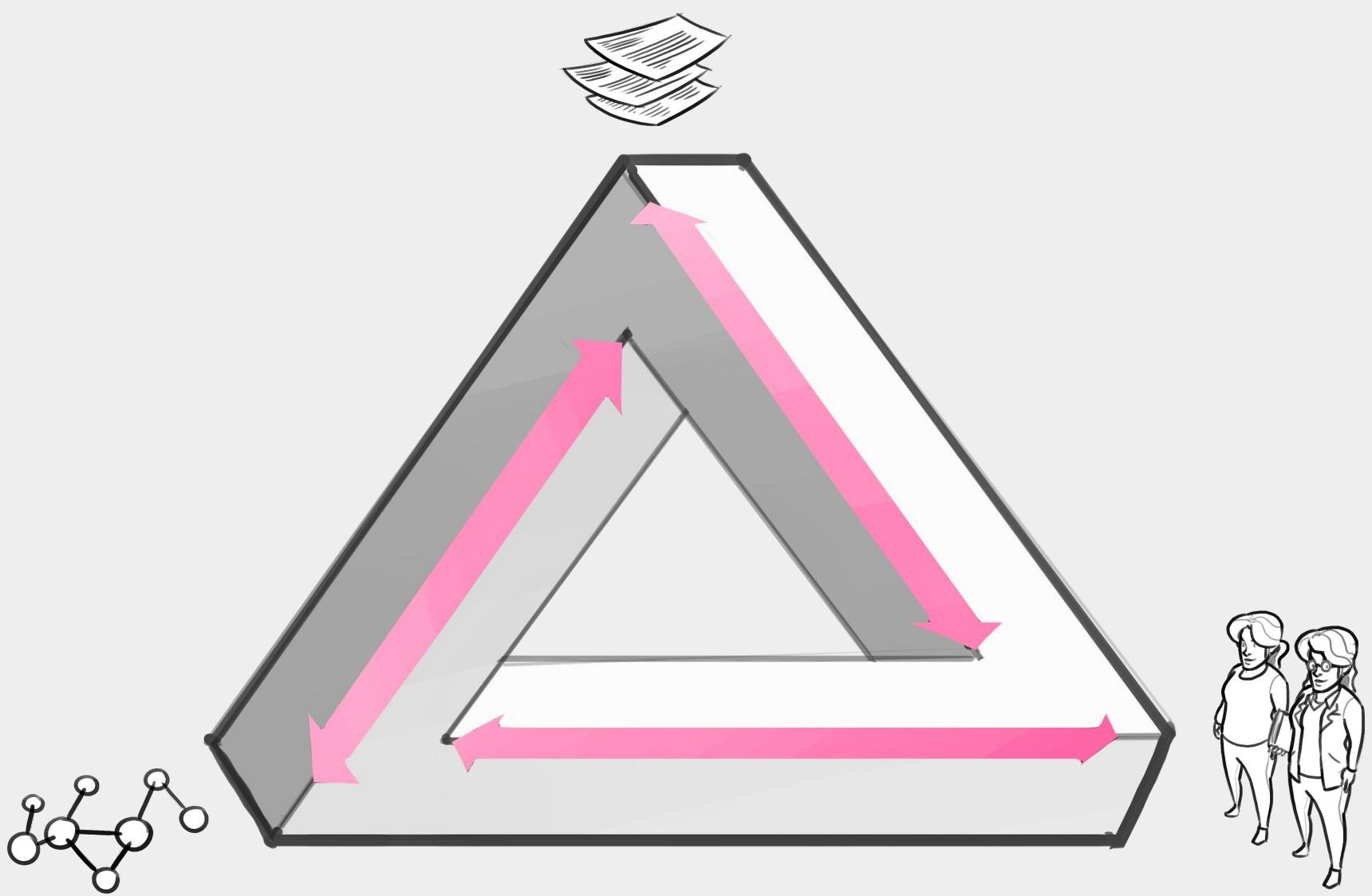
- $UAS(u, a)$ = ratio of workers that picked *annotation a*, over all workers that annotated the *unit u*, weighted by the *worker* quality.

$$UAS(u, a) = \frac{\sum_{i \in \text{workers}(u)} \text{WorkVec}(i, u)(a) WQS(i)}{\sum_{i \in \text{workers}(u)} WQS(i)}.$$

CROWDTRUTH QUALITY METRICS

Measuring the quality of annotations

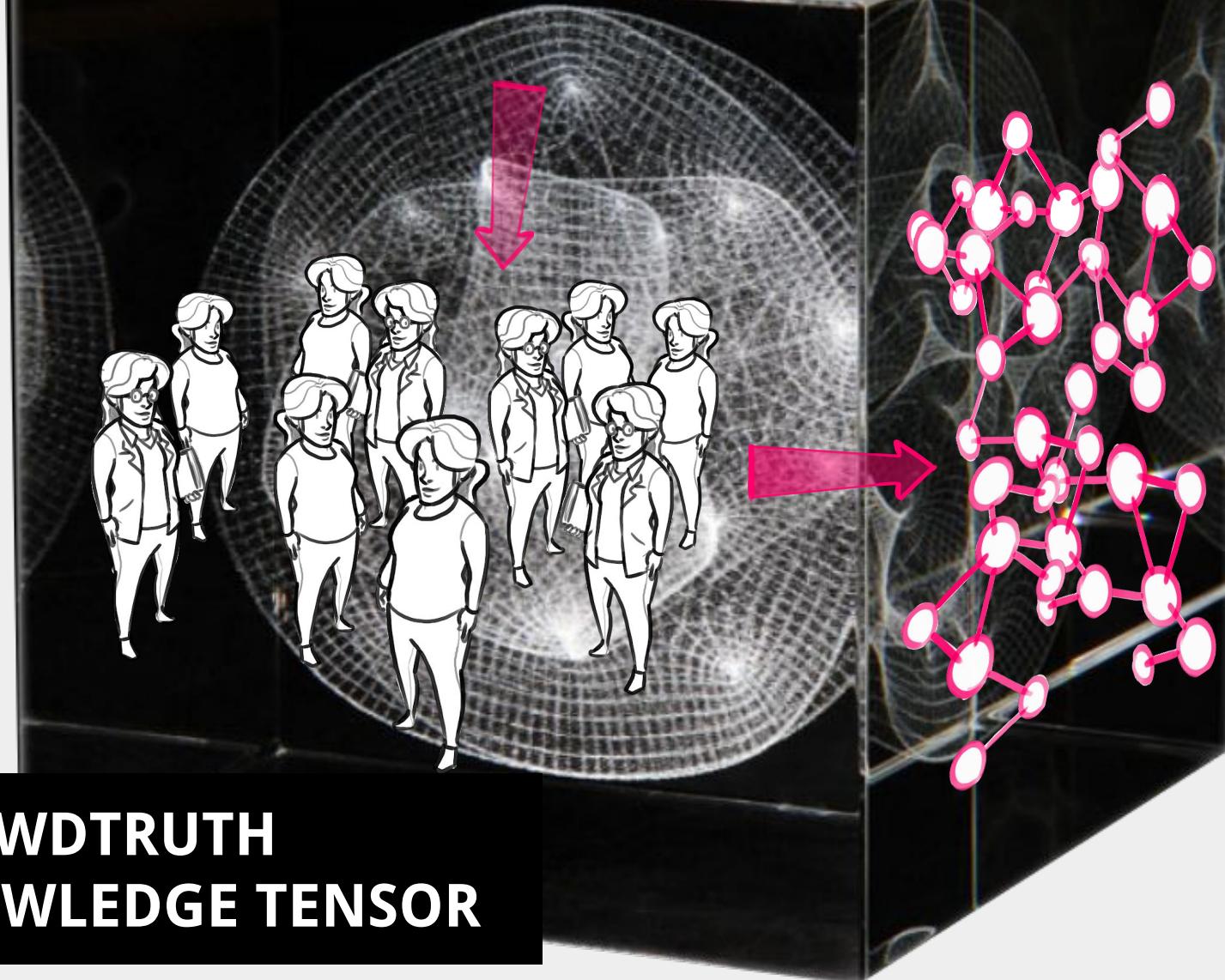




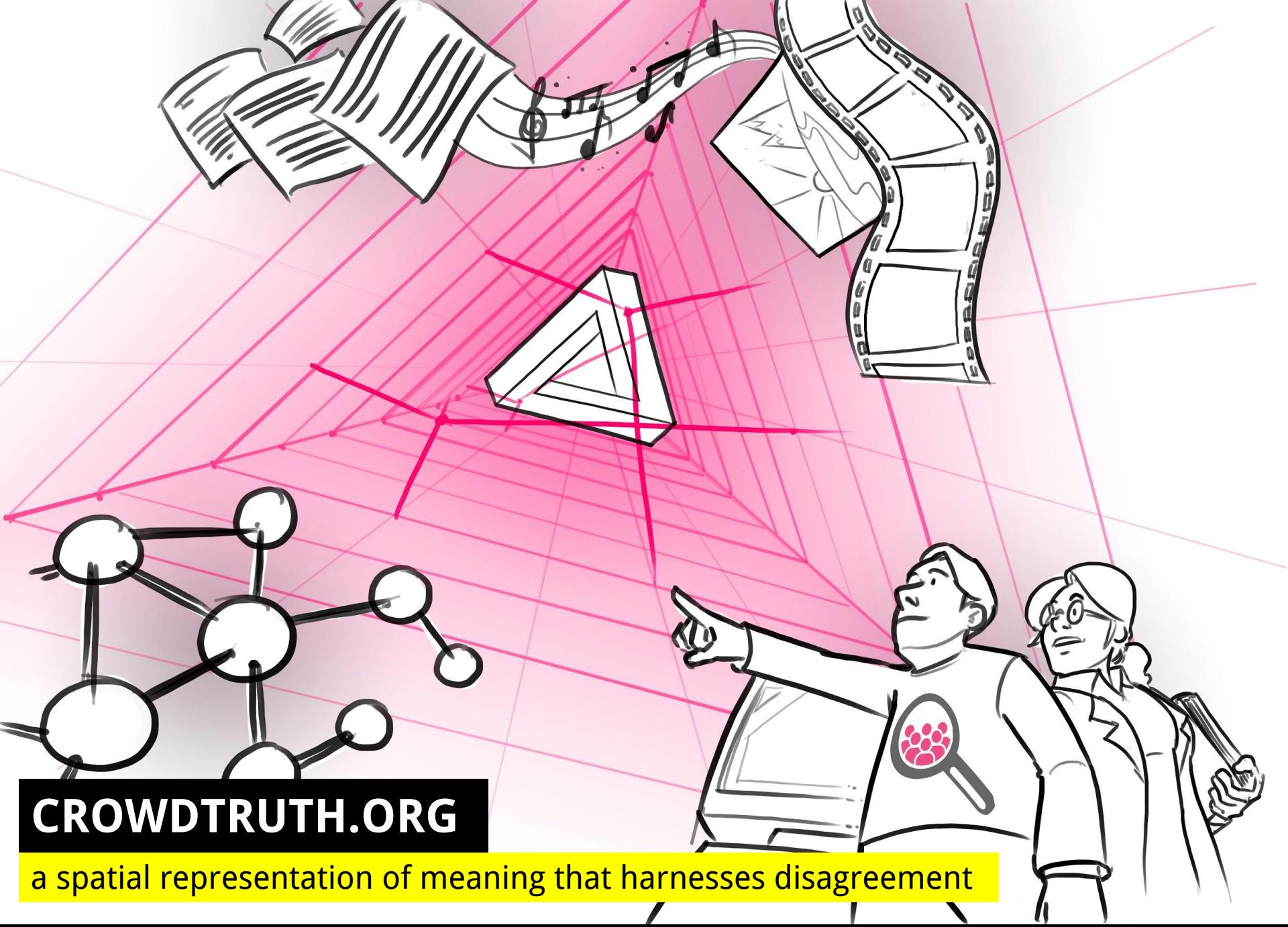
INTERDEPENDENT QUALITY SCORES

Updated in iterative fashion

For prevention of malaria, use only in individuals traveling to malarious areas where **CHLOROQUINE** resistant P. falciparum **MALARIA** has not been reported.



CROWDTRUTH KNOWLEDGE TENSOR



CROWDTRUTH.ORG

a spatial representation of meaning that harnesses disagreement



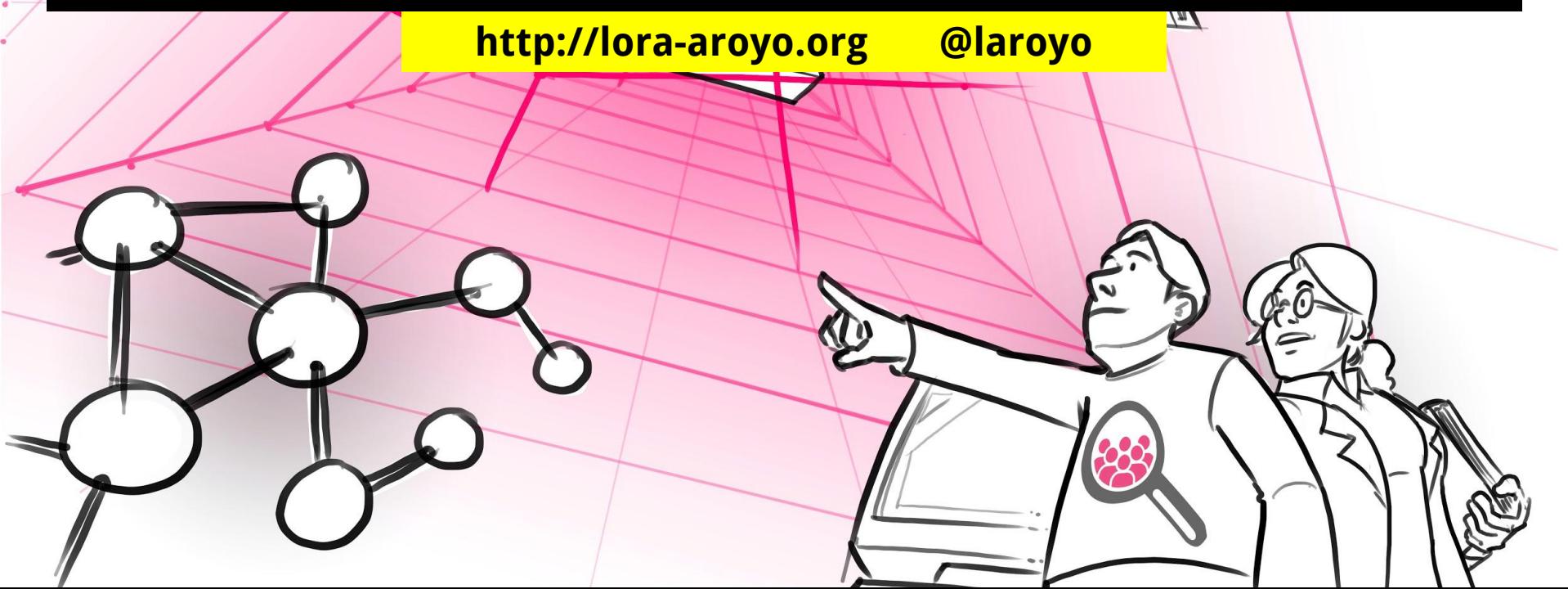
Hands-on Exercises: **Session III**

CROWDTRUTH.ORG

GITHUB.COM/CROWDTRUTH/CROWDTRUTH-CORE
PYPI.ORG/PROJECT/CROWDTRUTH

DATA.CROWDTRUTH.ORG

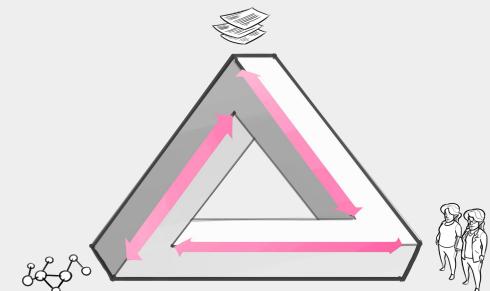
<http://lora-aroyo.org> @laroyo



- $\text{workers}(u)$: all workers that annotate media unit u ;
- $\text{units}(i)$: all input media units annotated by worker i ;
- $\text{WorkVec}(i, u)$: annotations of worker i on media unit u as a binary vector;
- $\text{MediaUnitVec}(s) = \sum_{i \in \text{workers}(s)} \text{WorkVec}(i, s)$, where s is an input media unit.

CROWDTRUTH QUALITY METRICS

Notation



Agreement between 2 workers on the same media unit:

- given 2 worker vectors vec_1 and vec_2 on the same media unit
- compute cosine similarity over the 2 worker vectors
- weight the cosine similarity by the Annotation Quality Score (AQS)
- for open tasks AQS = 1

$$\begin{aligned} Wcos(\text{vec}_1, \text{vec}_2) &= \\ &= \frac{\sum_a \text{vec}_1(a) \text{vec}_2(a) \text{AQS}(a)}{\sqrt{(\sum_a \text{vec}_1^2(a) \text{AQS}(a)) (\sum_a \text{vec}_2^2(a) \text{AQS}(a))}}, \\ &\quad \forall a - \text{annotation.} \end{aligned}$$

CROWDTRUTH QUALITY METRICS

Notation

