Cost and Quality in Crowdsourcing Workflows*

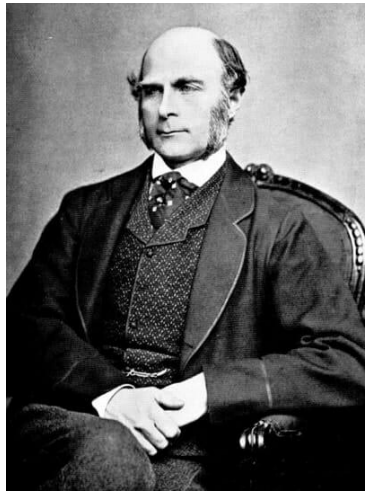
Rituraj Singh[1] Loïc Hélouët[2] Zoltan Miklos[1]

1 UMR IRISA

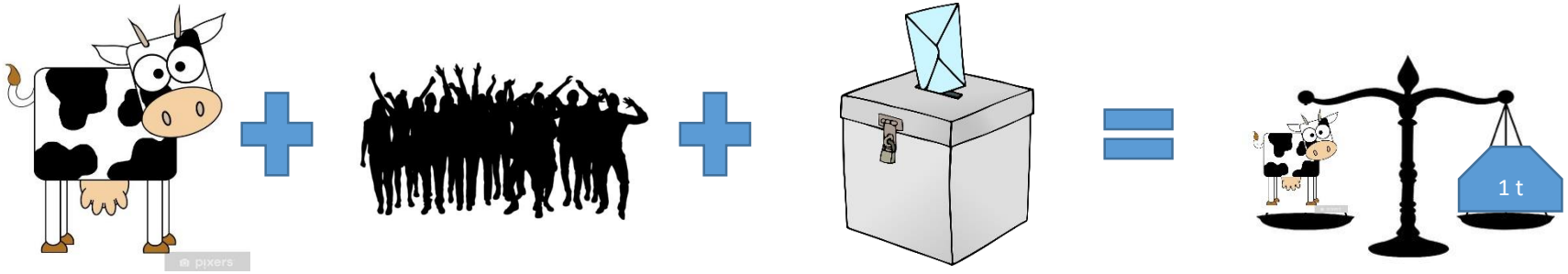
2 Inria

* Work supported by the headwork ANR project

Crowdsourcing : from 1906 to 2021



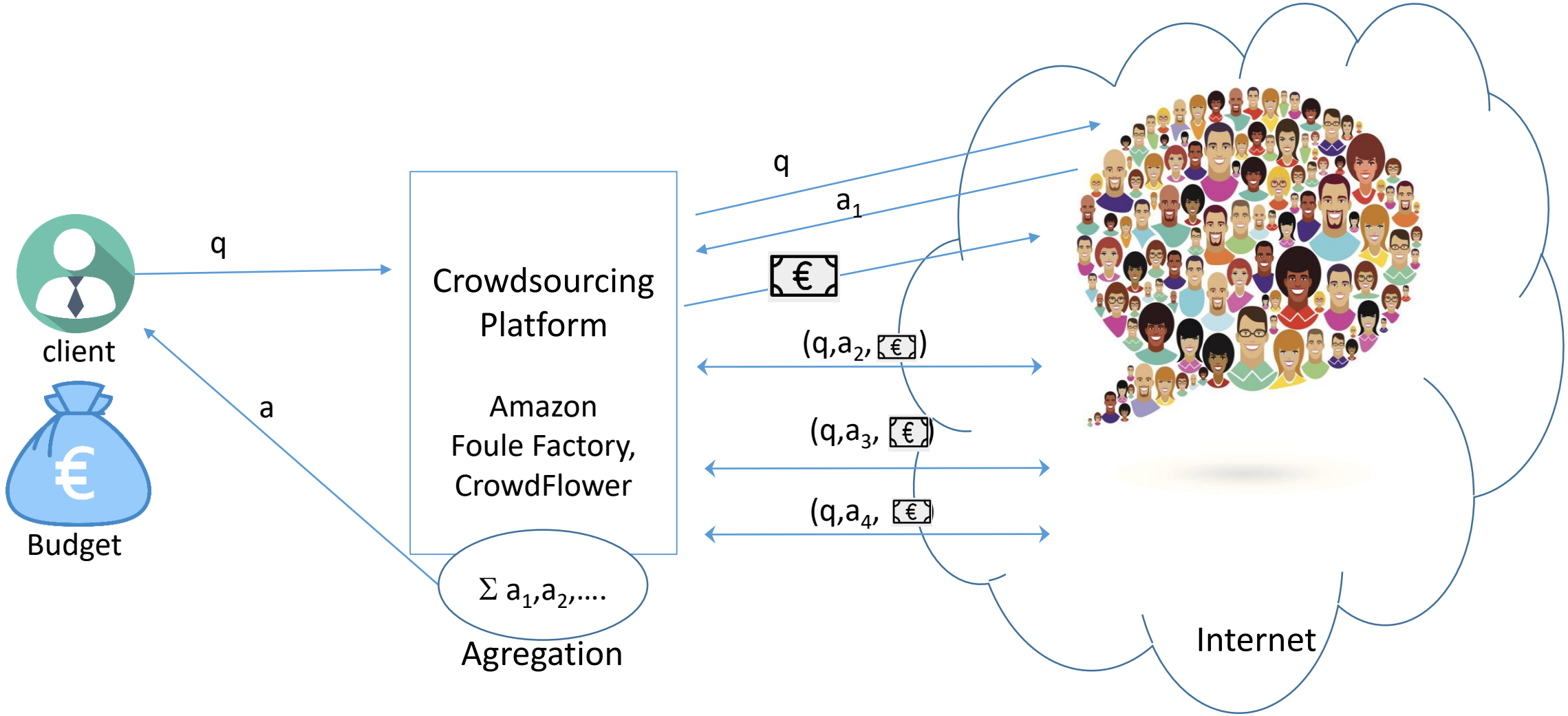
Francis Galton



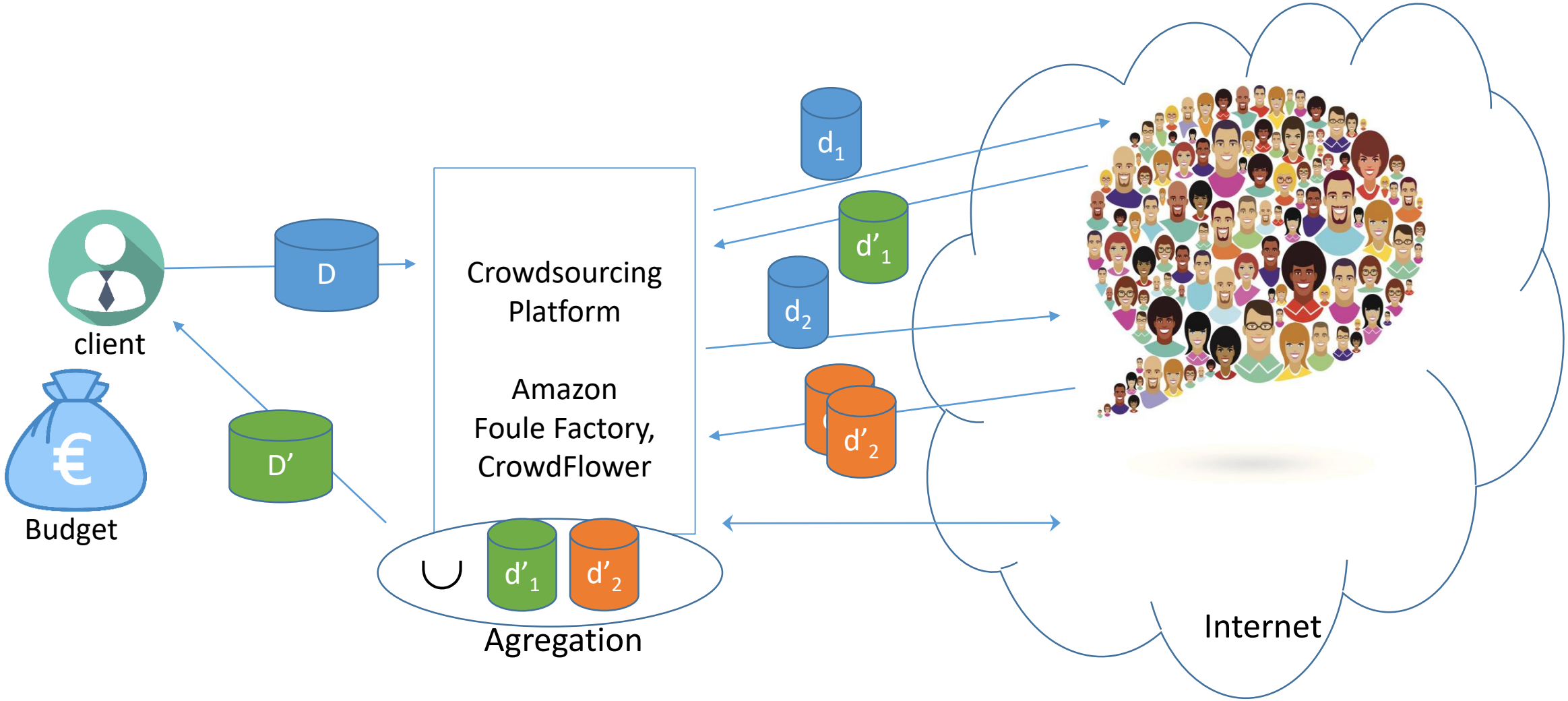
<u>Wisdom of Crowds (1906):</u>

if you ask enough people the same question,
they might come up with better answers than even the experts.

Crowdsourcing : since 2005



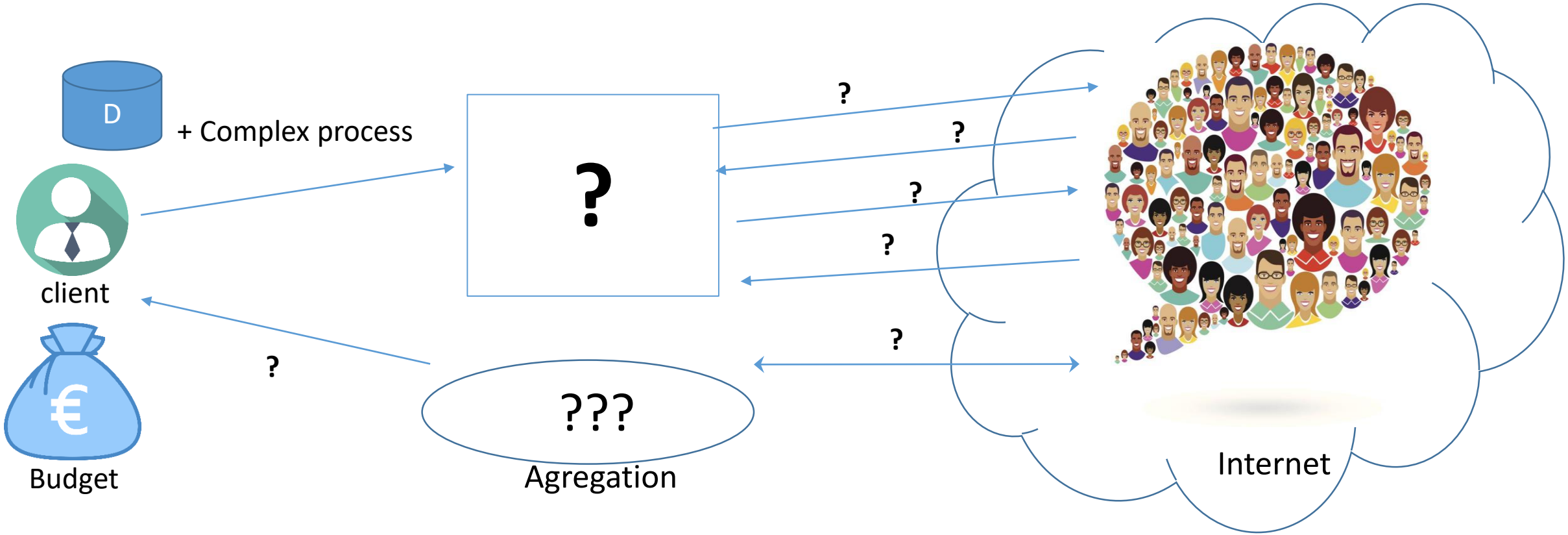
Crowdsourcing : since 2005



Crowdsourcing : 2020-2021

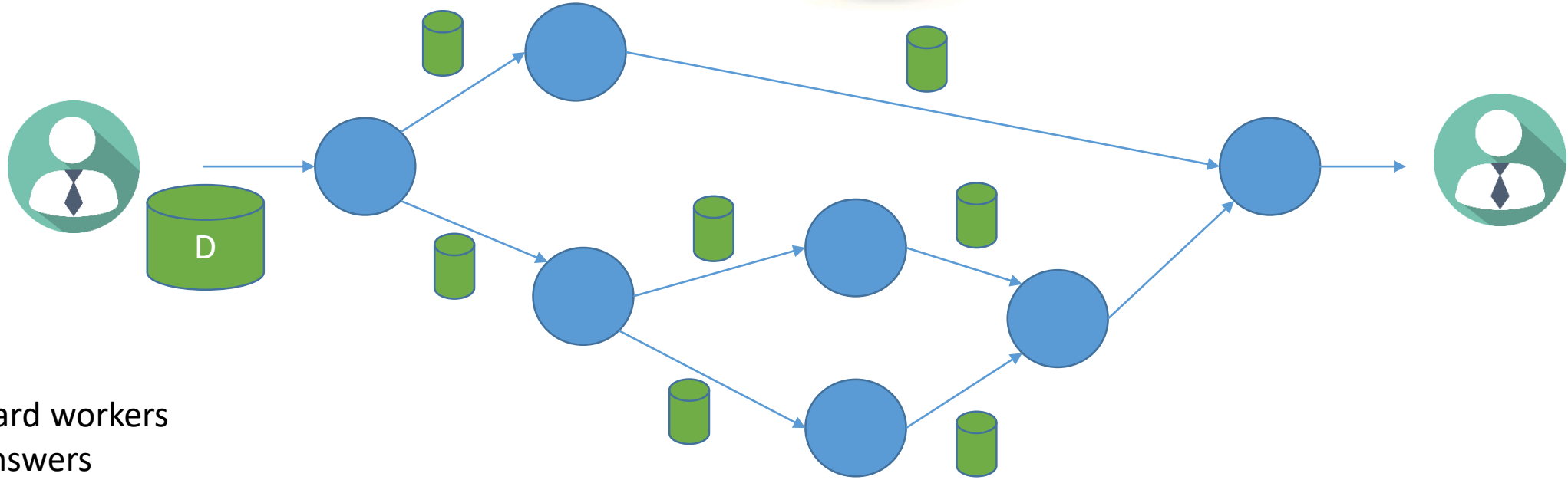
Contributive Science / IA Training / text-speech processing

- Data
- Complex processes
- Human + Data centric



Complex workflows

Close to Business processes/workflow nets ?



<u>But</u>:

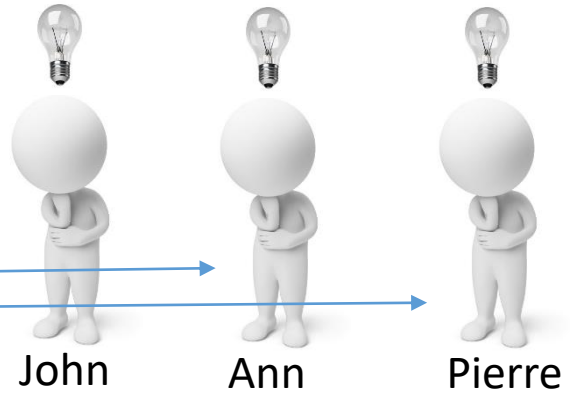
- Hire & reward workers
- agregate answers
- Split & Distribute datasets

Micro-tasks

Image	Quality	Date	Bee(Y/N)
Img001.jpg	Good	21/06/2021	-
Img002.jpg	Good	21/06/2021	-
Img003.jpg	Poor	21/06/2021	-
Img004.jpg	Good	22/06/2021	-
Img005.jpg	poor	22/06/2021	-

fill this column

Platform



Task = simple work unit

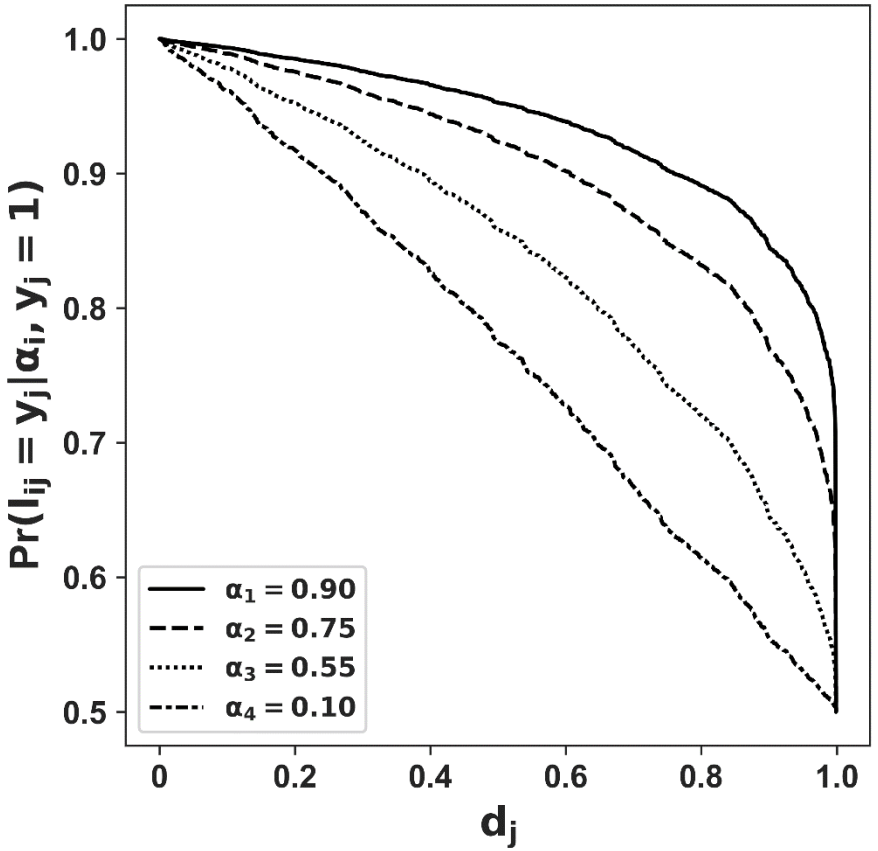
- Tag an image
- Simple text/language processing
- …

in this talk :

simple boolean answers

Users Behaviors

Image	Quality	Date	Bee(Y/N)
Img001.jpg	Good	21/06/2021	-
Img002.jpg	Good	21/06/2021	-
Img003.jpg	Poor	21/06/2021	-
Img004.jpg	Good	22/06/2021	-
Img005.jpg	poor	22/06/2021	-



Each worker wi has

a recall α_i: probability to answer 0 when ground truth is 0

a specificity β_i : probability to answer 1 when ground truth is 1

….that depends on difficulty d_j of a task t_j

α_i , β_i define the probability of correct answer

unknown variables
(but can be estimated)

A clever agregation

John

Ann

Pierre

Image	Quality	Date	Bee(Y/N)
Img001.jpg	Good	21/06/2021	0 0 0
Img002.jpg	Good	21/06/2021	0 1 0
Img003.jpg	Poor	21/06/2021	1 1 0
Img004.jpg	Good	22/06/2021	1 1 1
Img005.jpg	poor	22/06/2021	1 0 1

Platform

A clever agregation

John

Ann

Pierre

Image	Quality	Date	Bee(Y/N)
Img001.jpg	Good	21/06/2021	0 0 0
Img002.jpg	Good	21/06/2021	0 1 0
Img003.jpg	Poor	21/06/2021	1 1 0
Img004.jpg	Good	22/06/2021	1 1 1
Img005.jpg	poor	22/06/2021	1 0 1

Platform

Agregation

(Expectation Maximization)

Image	Quality	Date	Bee(Y/N)
Img001.jpg	Good	21/06/2021	0
Img002.jpg	Good	21/06/2021	1
Img003.jpg	Poor	21/06/2021	1
Img004.jpg	Good	22/06/2021	1
Img005.jpg	poor	22/06/2021	1

Confidence
0.98
0.75
0.82
0.99
0.82

	α	β
John	0.8	0.82
Ann	0.95	0.95
Pierre	0.71	0.73

Static worker allocation, synchronous policy

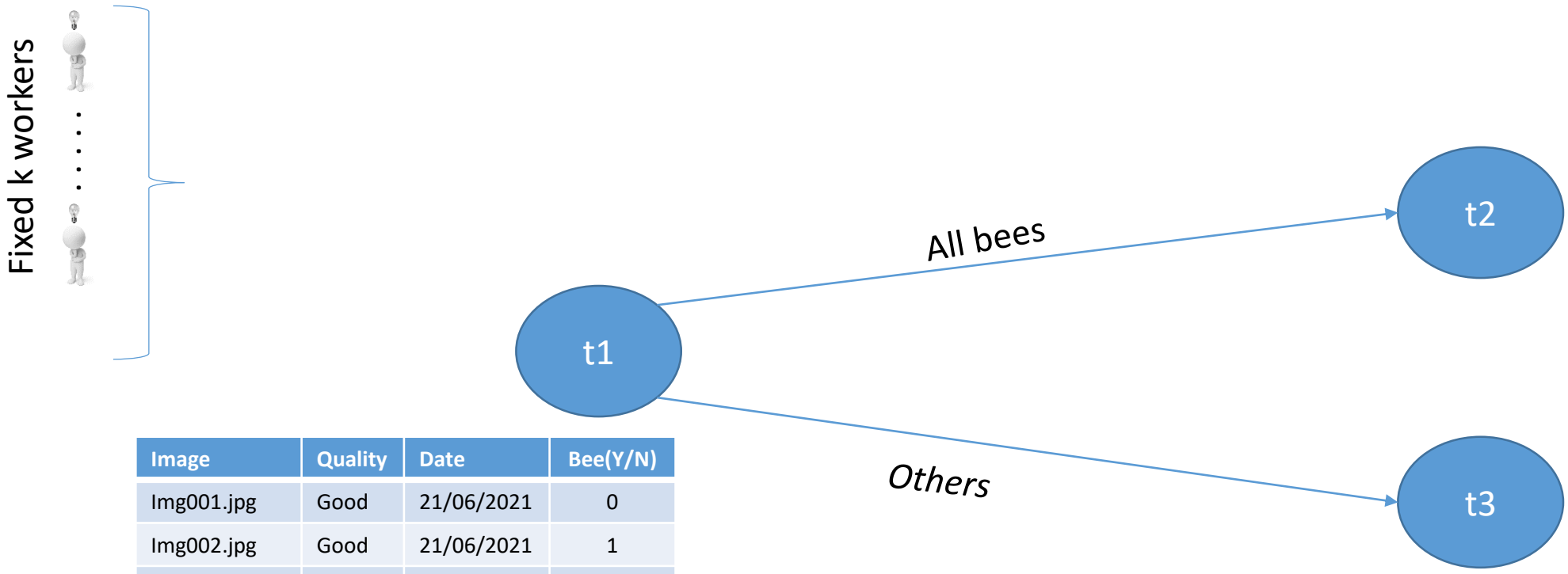


Image	Quality	Date	Bee(Y/N)
Img001.jpg	Good	21/06/2021	0
Img002.jpg	Good	21/06/2021	1
Img003.jpg	Poor	21/06/2021	1
Img004.jpg	Good	22/06/2021	1
Img005.jpg	poor	22/06/2021	1

Static worker allocation, synchronous policy

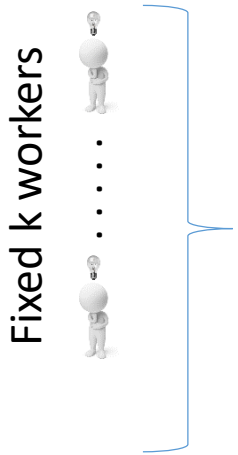


Image	Quality	Date	Bee(Y/N)
Img001.jpg	Good	21/06/2021	0
Img002.jpg	Good	21/06/2021	1
Img003.jpg	Poor	21/06/2021	1
Img004.jpg	Good	22/06/2021	1
Img005.jpg	poor	22/06/2021	1

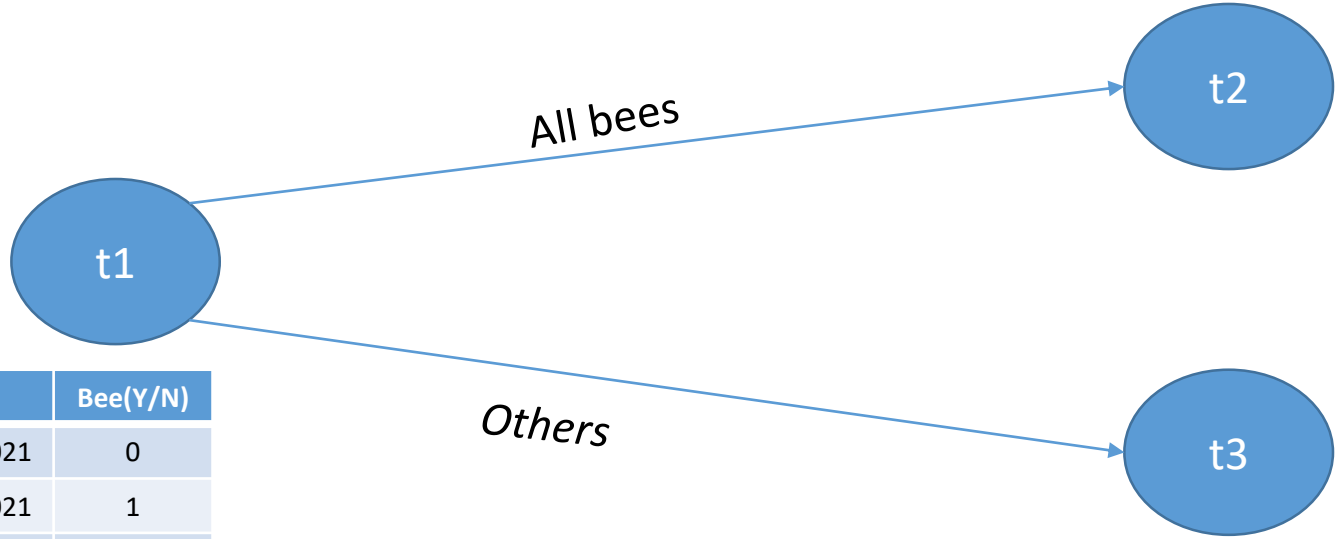


Image	Quality	Date	Bee(Y/N)
Img002.jpg	Good	21/06/2021	1
Img003.jpg	Poor	21/06/2021	1
Img004.jpg	Good	22/06/2021	1
Img005.jpg	poor	22/06/2021	1

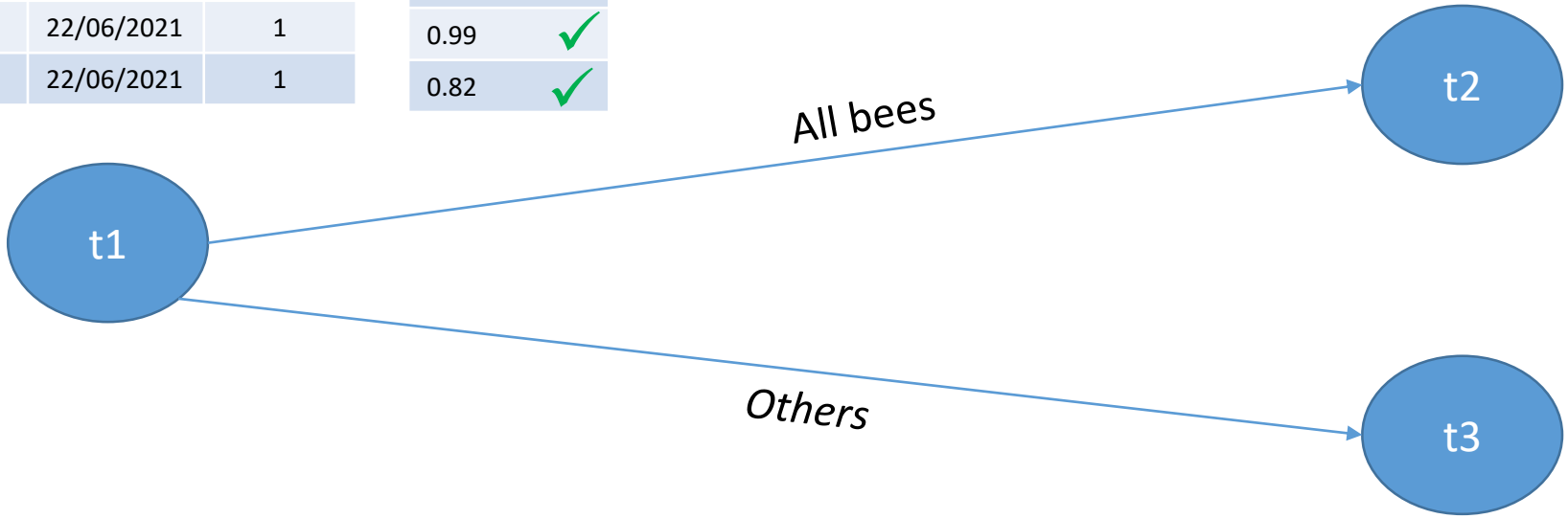
Image	Quality	Date	Bee(Y/N)
Img001.jpg	Good	21/06/2021	0

Dynamic worker allocation,Synchronous policy

Threshold = 0.7

Image	Quality	Date	Bee(Y/N)
Img001.jpg	Good	21/06/2021	0
Img002.jpg	Good	21/06/2021	1
Img003.jpg	Poor	21/06/2021	1
Img004.jpg	Good	22/06/2021	1
Img005.jpg	poor	22/06/2021	1

Confidence
0.98 ✓
0.75 ✓
0.82 ✓
0.99 ✓
0.82 ✓



Dynamic worker allocation,Synchronous policy

Image	Quality	Date	Bee(Y/N)
Img001.jpg	Good	21/06/2021	0
Img002.jpg	Good	21/06/2021	1
Img003.jpg	Poor	21/06/2021	1
Img004.jpg	Good	22/06/2021	1
Img005.jpg	poor	22/06/2021	1

Threshold = 0.7

Confidence
0.98 ✓
0.75 ✓
0.82 ✓
0.99 ✓
0.82 ✓

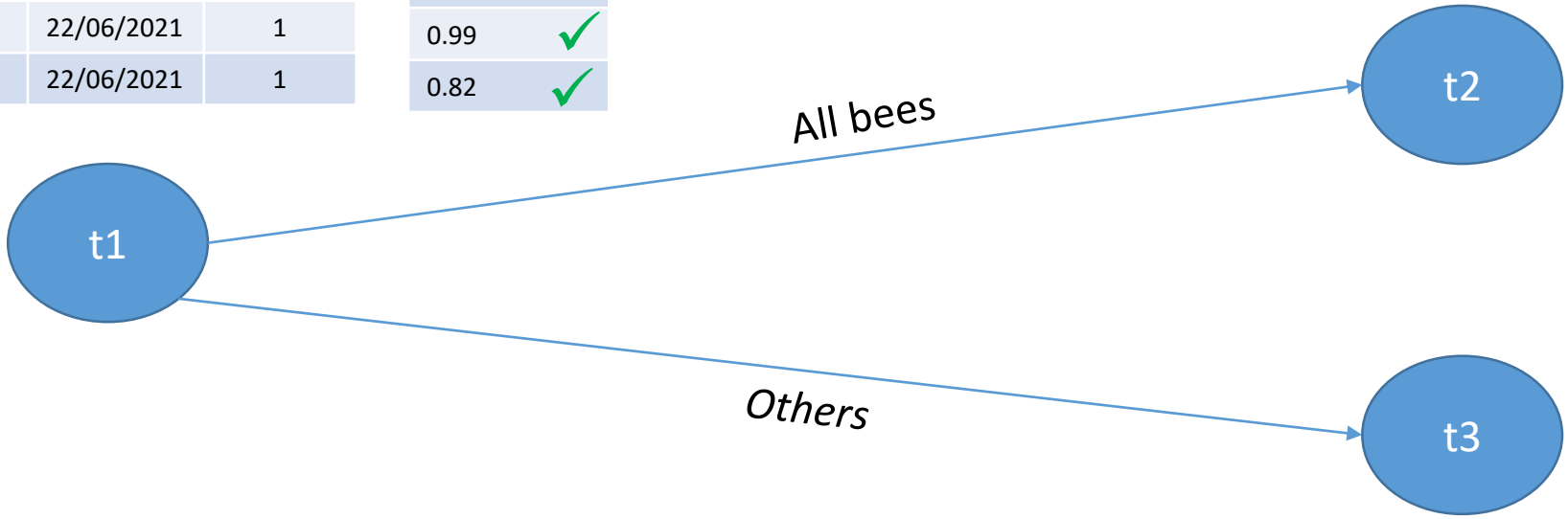


Image	Quality	Date	Bee(Y/N)
Img002.jpg	Good	21/06/2021	1
Img003.jpg	Poor	21/06/2021	1
Img004.jpg	Good	22/06/2021	1
Img005.jpg	poor	22/06/2021	1

Image	Quality	Date	Bee(Y/N)
Img001.jpg	Good	21/06/2021	0

Dynamic worker allocation,Synchronous policy

Image	Quality	Date	Bee(Y/N)
Img001.jpg	Good	21/06/2021	0
Img002.jpg	Good	21/06/2021	1
Img003.jpg	Poor	21/06/2021	1
Img004.jpg	Good	22/06/2021	1
Img005.jpg	poor	22/06/2021	1

Confidence	
0.98	✓
0.75	✗
0.82	✓
0.99	✓
0.82	✓

Threshold = 0.8

Platform

Mary

t1

All bees

t2

Others

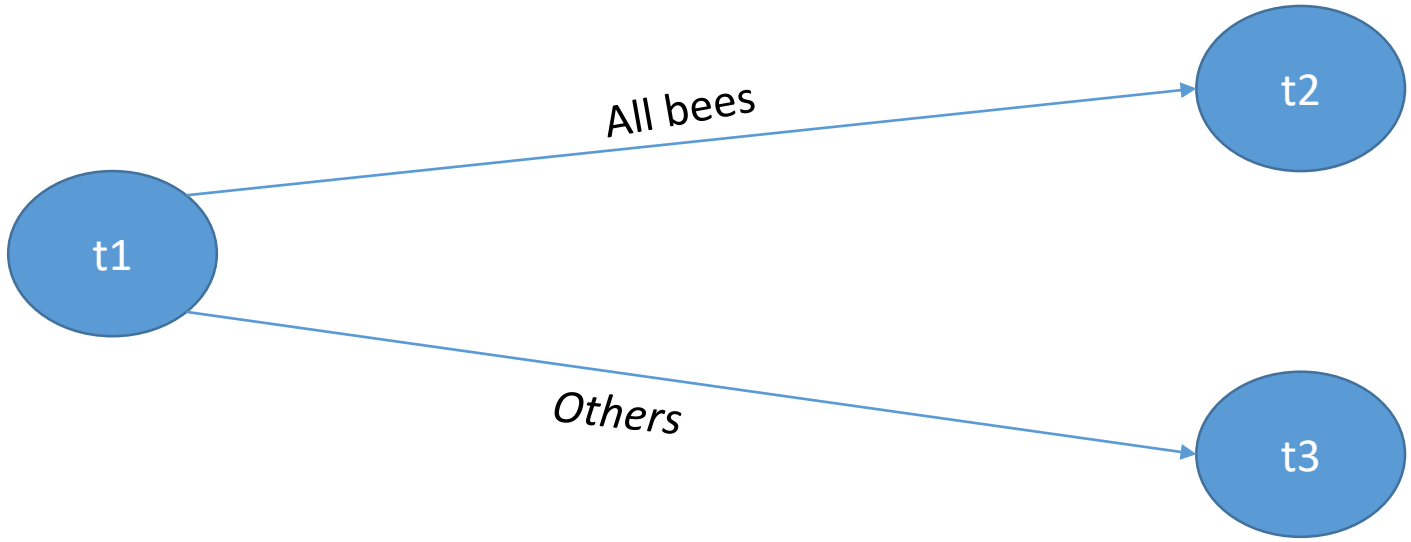
t3

Dynamic worker allocation, Asynchronous policy

Image	Quality	Date	Bee(Y/N)
Img001.jpg	Good	21/06/2021	0
Img002.jpg	Good	21/06/2021	1
Img003.jpg	Poor	21/06/2021	1
Img004.jpg	Good	22/06/2021	1
Img005.jpg	poor	22/06/2021	1

Confidence	
0.98	✓
0.75	✗
0.82	✓
0.99	✓
0.82	✓

Threshold = 0.8



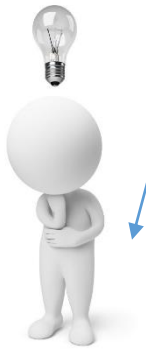
Dynamic worker allocation, Asynchronous policy

Image	Quality	Date	Bee(Y/N)
Img002.jpg	Good	21/06/2021	1

Threshold = 0.8

Image	Quality	Date	Bee(Y/N)
Img003.jpg	Poor	21/06/2021	1
Img004.jpg	Good	22/06/2021	1
Img005.jpg	poor	22/06/2021	1

Platform



Mary

t1

All bees

t2

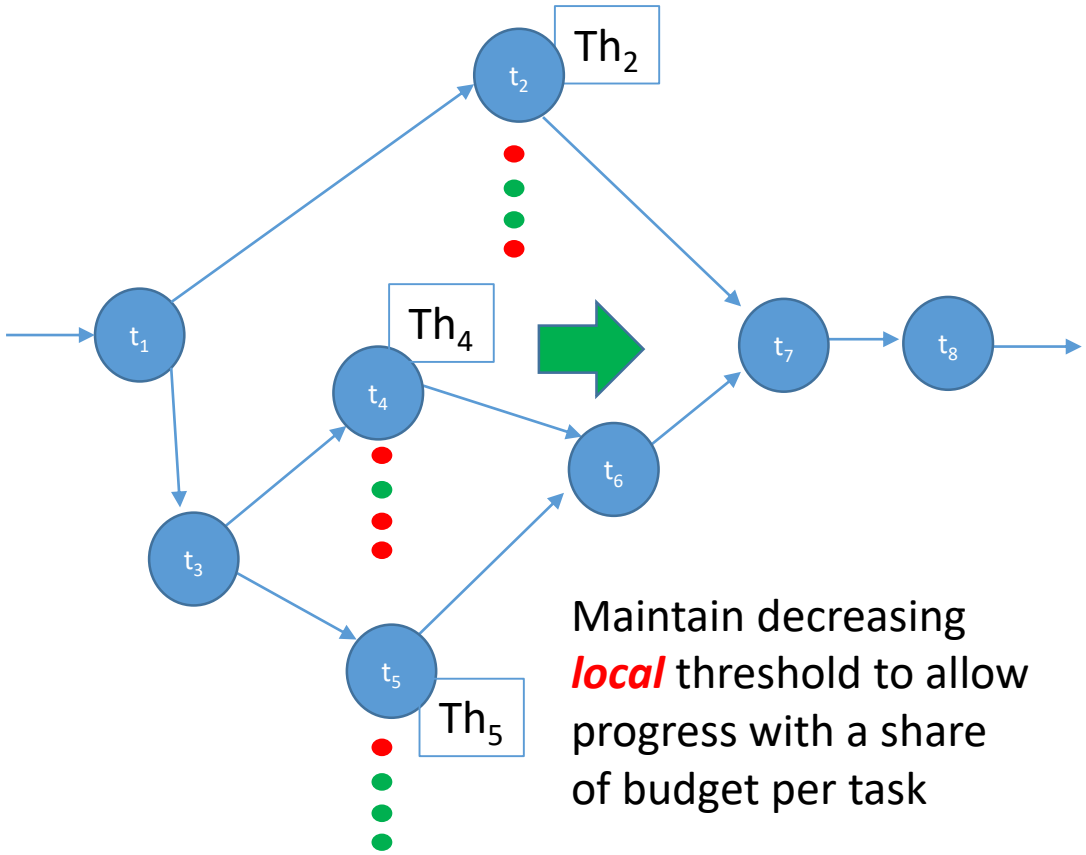
Others

t3

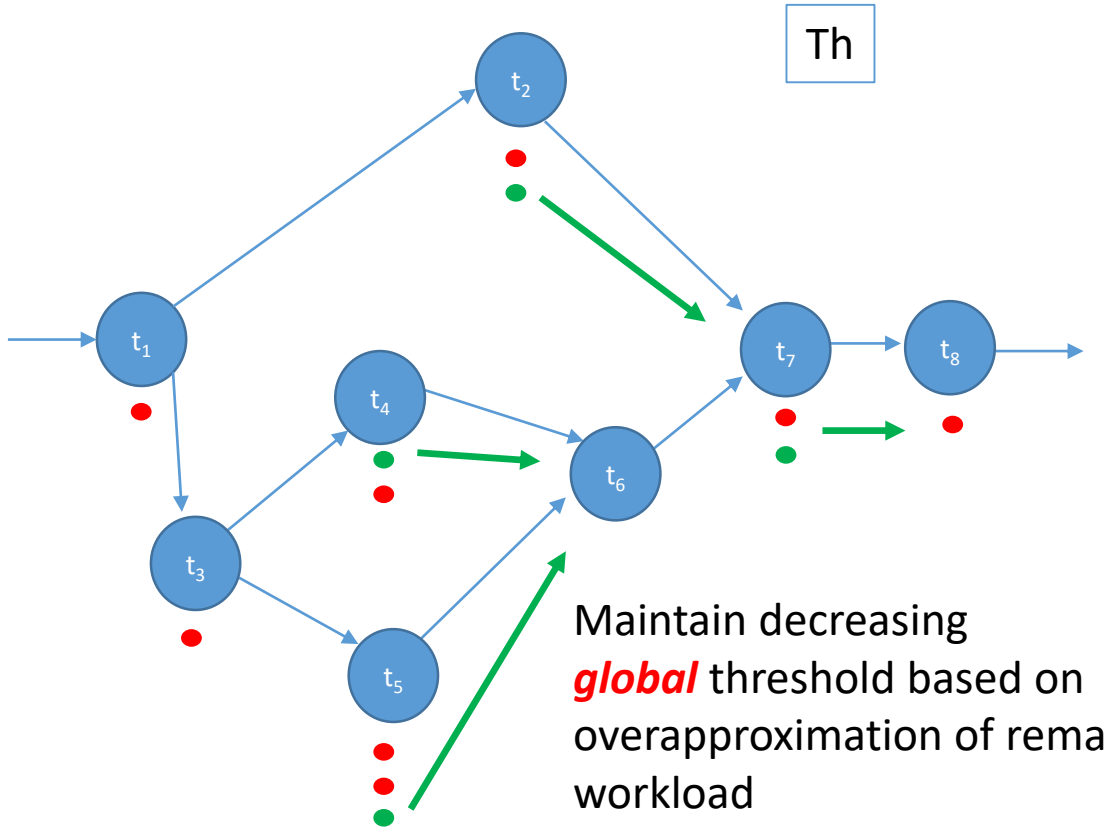
Image	Quality	Date	Bee(Y/N)
Img001.jpg	Good	21/06/2021	0

Threshold vs Budget

Reminder : Clients have limited budget workers are paid => a bounded number of calls to the crowd



Maintain decreasing ***local*** threshold to allow progress with a share of budget per task



Maintain decreasing ***global*** threshold based on overapproximation of remaining workload

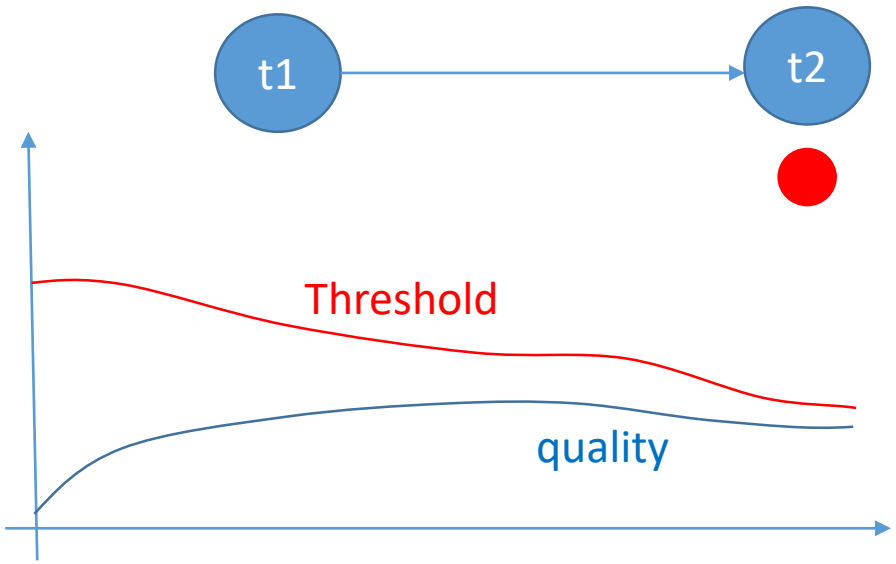
Questions

Given a fixed population of workers U
a workflow W
an input Dataset D to process
a maximal budget B_{max}

- Does W always terminate with dynamic allocation ?
- Is dynamic worker allocation performing better than static allocation on W?
- Is asynchronous allocation performing better than synchronous allocation on W?
- What is the best allocation strategy for U, W, D, B_{max} ?

Questions

- Does W always terminate with dynamic allocation ? **NO**

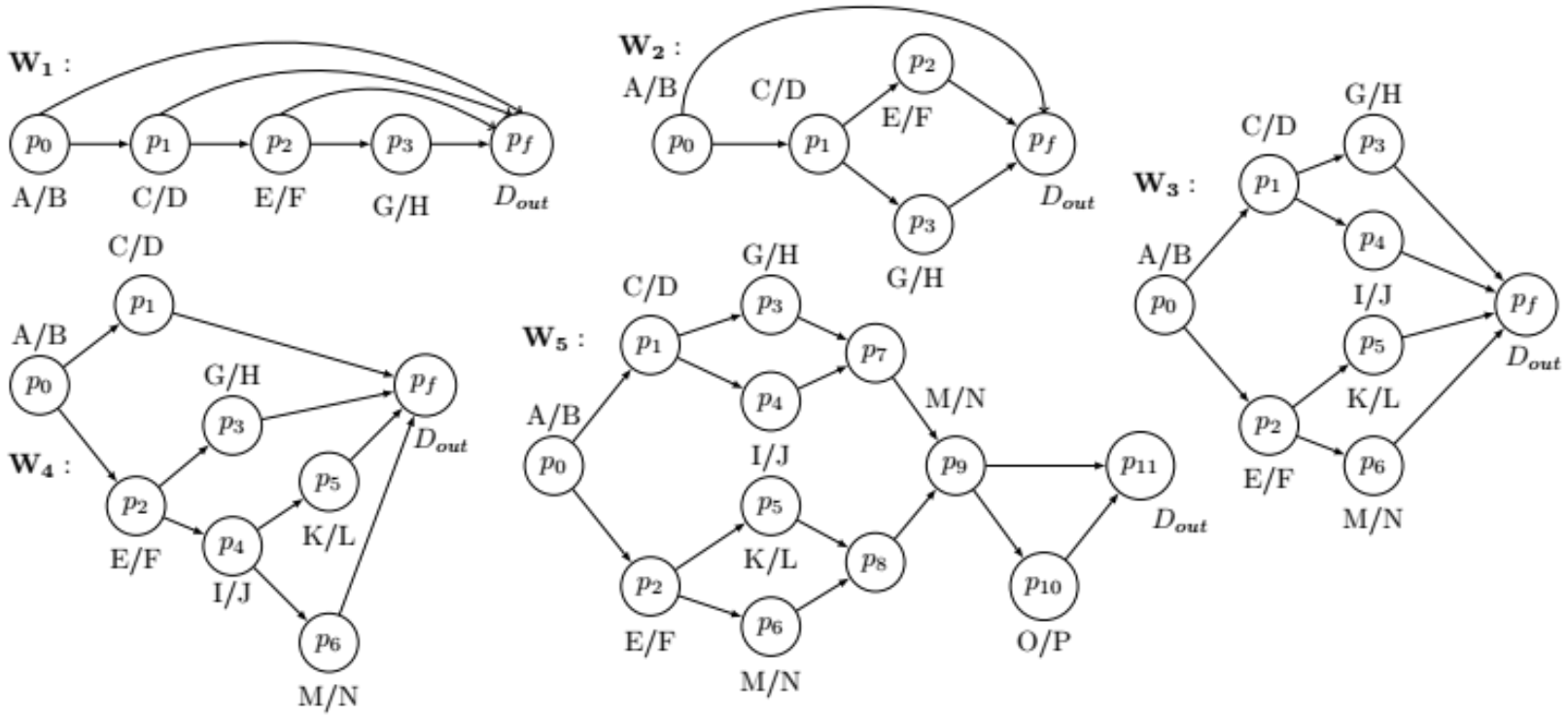


- What is the best allocation strategy for U, W, D, B_{max} ?

a HUGE Markov Decision process (contains Datasets)
Experimental results only.

Experiment (1)

No available benchmark for complex processes



We built a set of 5 small typical workflows : linear, fork-join, short vs long branches, ….

Experiment (2)

F1	F2	F3	F4	F5	F6	F7	F8	F9
A	B	C	0	1	0	1	1	0
A	D	E	1	1	0	1	0	0
D	F	G	0	1	1	0	1	0
A	D	B	1	0	1	0	0	1
L	M	N	0	0	0	1	1	1

Existing but to be discovered by the crowd

Step 2 : We generated synthetic data
Ground truth is known
Datasets are balanced or unbalanced

Step 3 : we generated 4 sets of 50 crowdworkers with various expertize levels
Low | mid |average | high
=> ground truth + expertize of a worker allows to sample a plausible answer

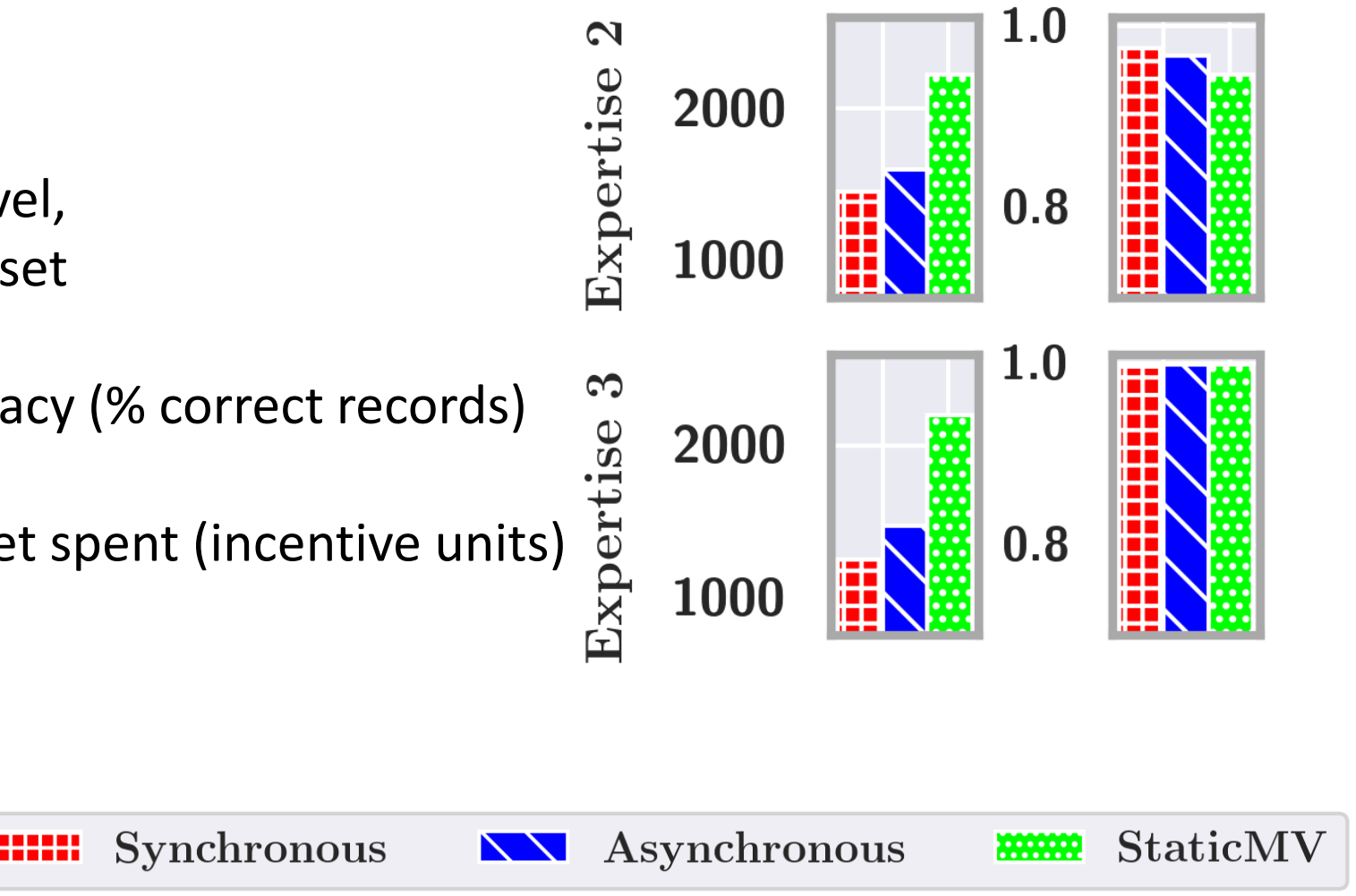
Step 4 : We realized each complex workflow with a static allocation policy
and 10, 20, 30 answers per records in a task
achieved costs then used as initial budget B_{10}, B_{20}, B_{30} for dynamic allocation experiments

Step 5 : We realized all complex workflows with dynamic policies (several experiments to avoid bias)
for each workflow | Dataset | experience level
with initial budget B_{10}, B_{20}, B_{30}

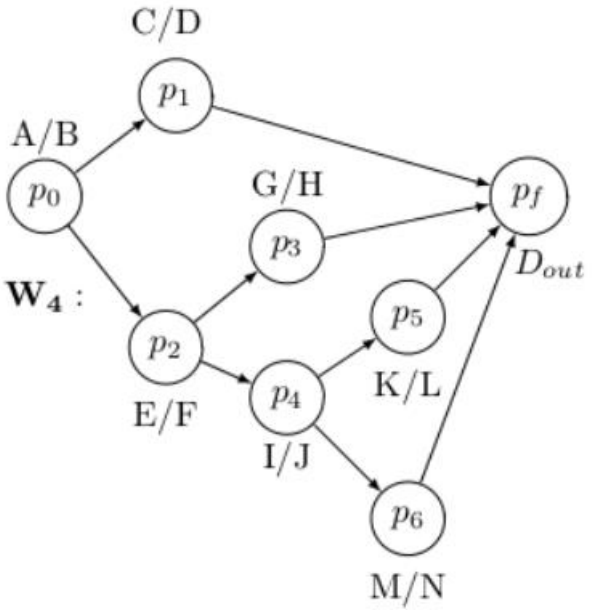
Measures obtained

For each workflow,
For each expertize level,
For each type of dataset

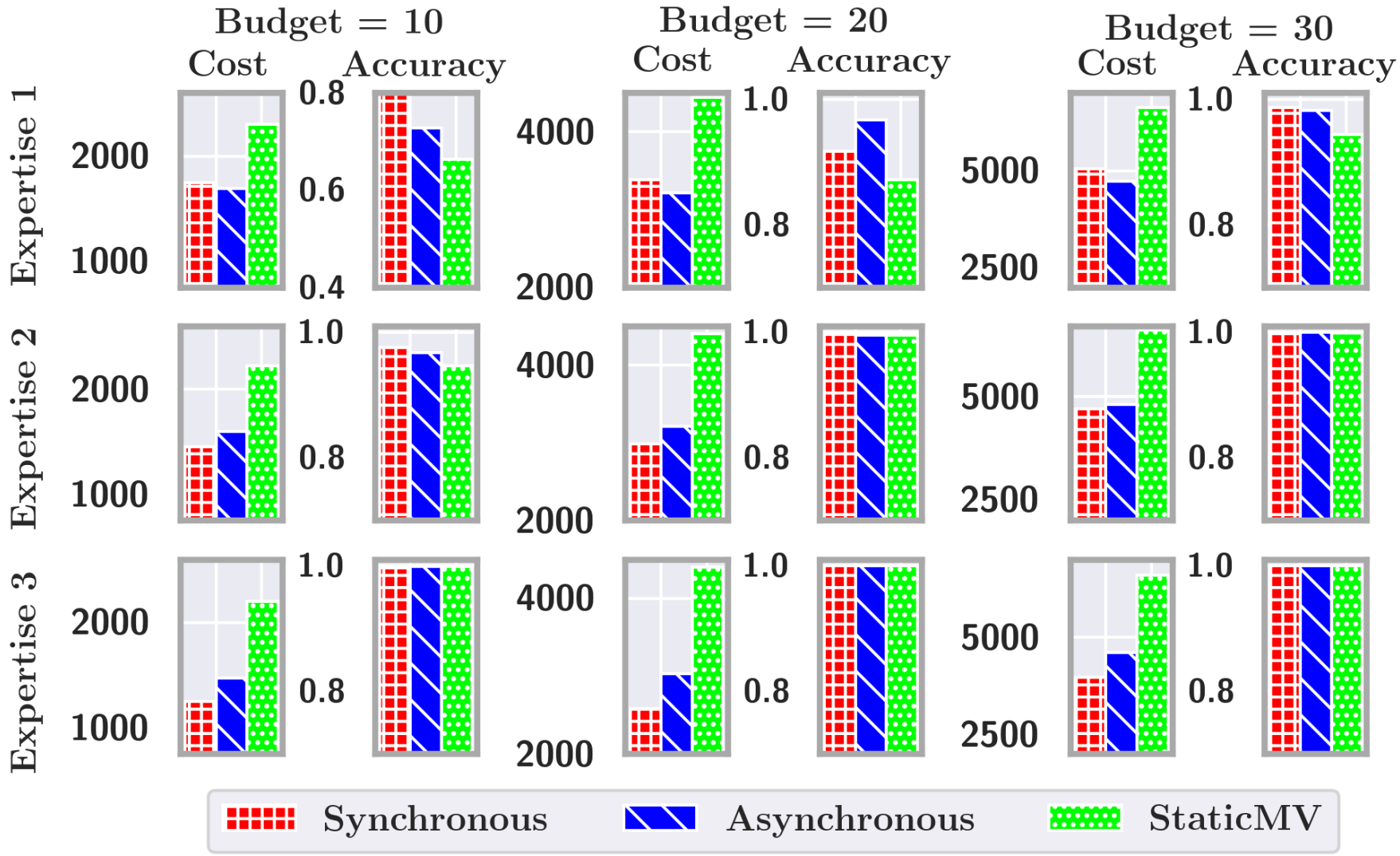
- average accuracy (% correct records)
- average budget spent (incentive units)



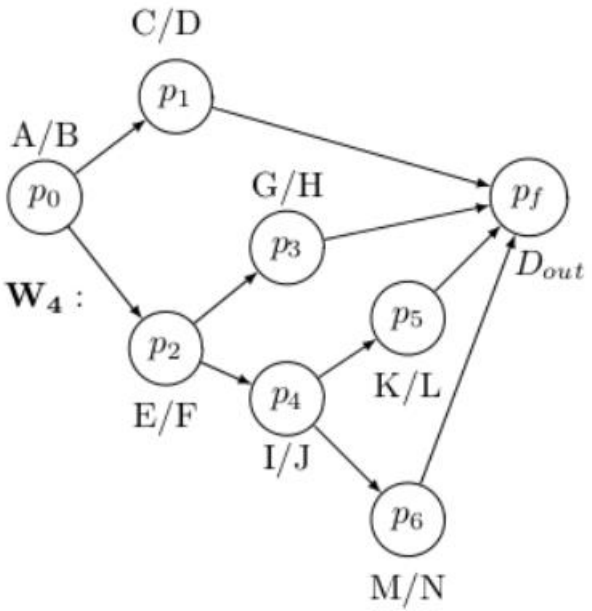
Results



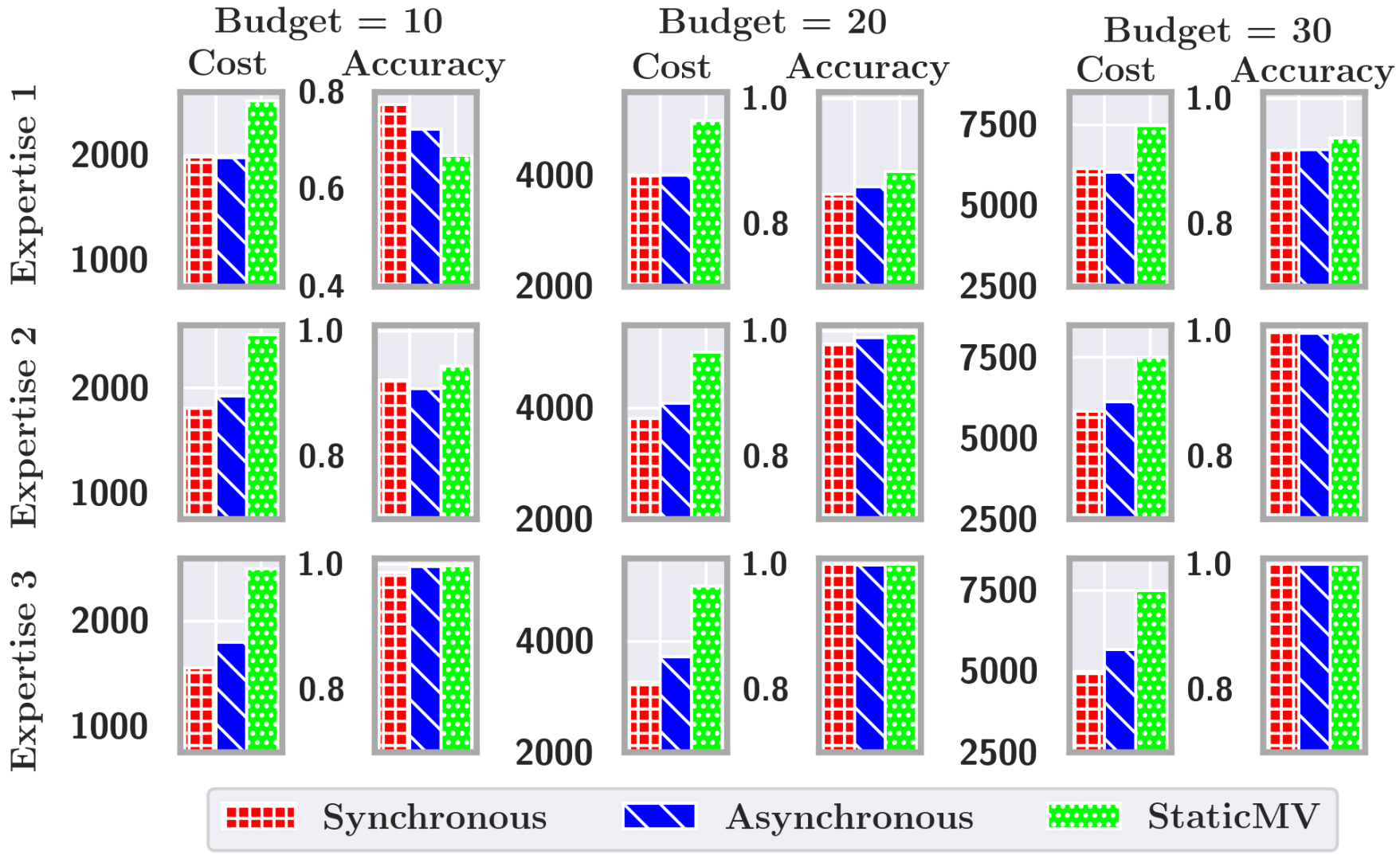
(balanced data)



Results



(unbalanced data)



Lessons learned

- Low expertize leads to poor accuracy
- Dynamic schemes outperform static ones (cost & accuracy).
- When worker expertise increase, accuracy increases too, & the consumed budget decreases.

Expected

- Synchronous policy achieves better accuracy than asynchronous policy
- Unbalanced data consumes more budget than balanced data.

Unexpected