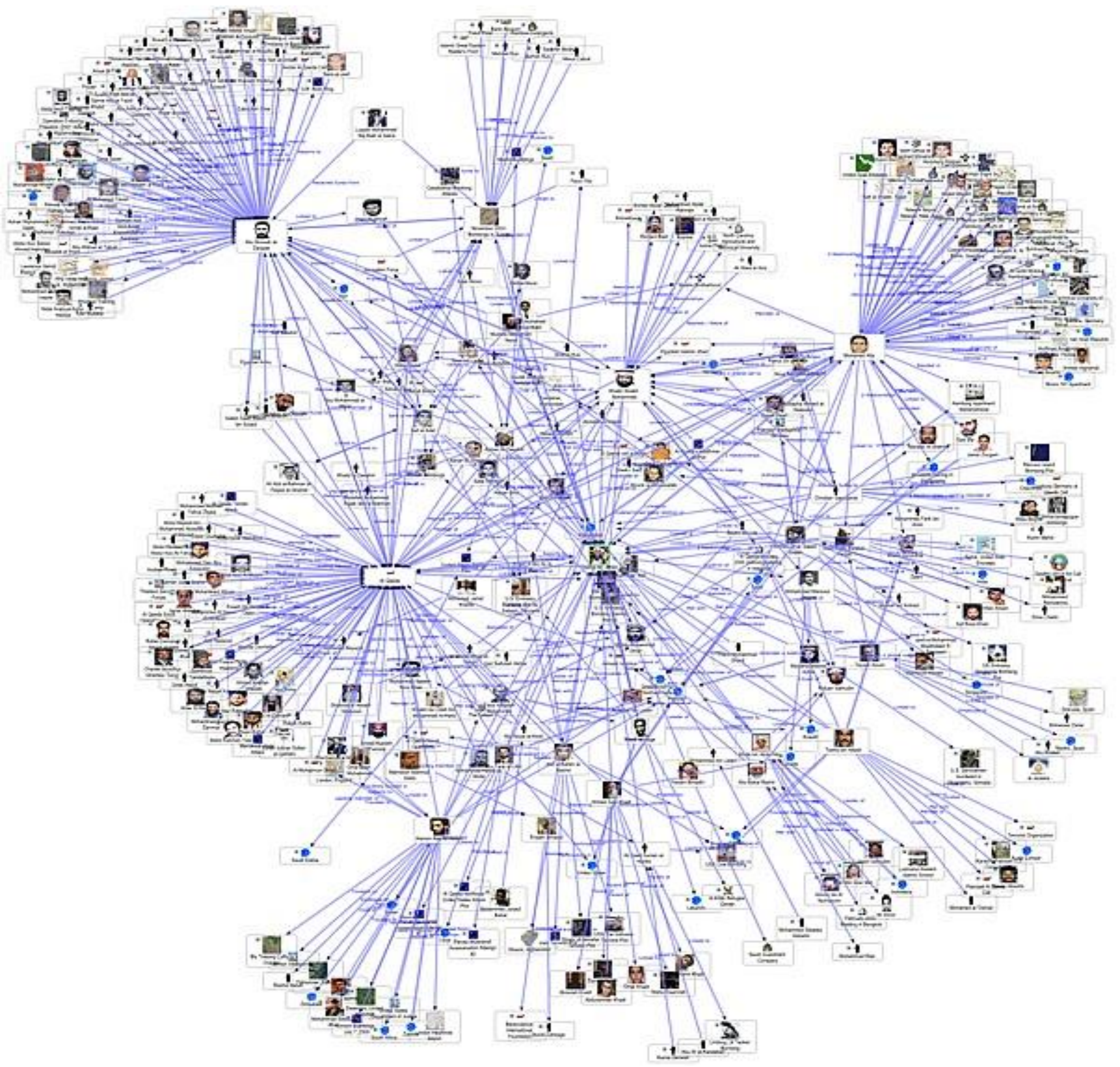
Agent-Based Modelling

$$\mathcal{H} = -\sum_{\langle j,k \rangle} J_{jk}\sigma_j\sigma_k - H\sum_j \sigma_j$$

Table of Contents

1. What ABMs are
2. Why we use ABMs
3. ABM construction
4. Verification and Validation
5. Concerns
6. Summary



What is Agent-Based Modelling?

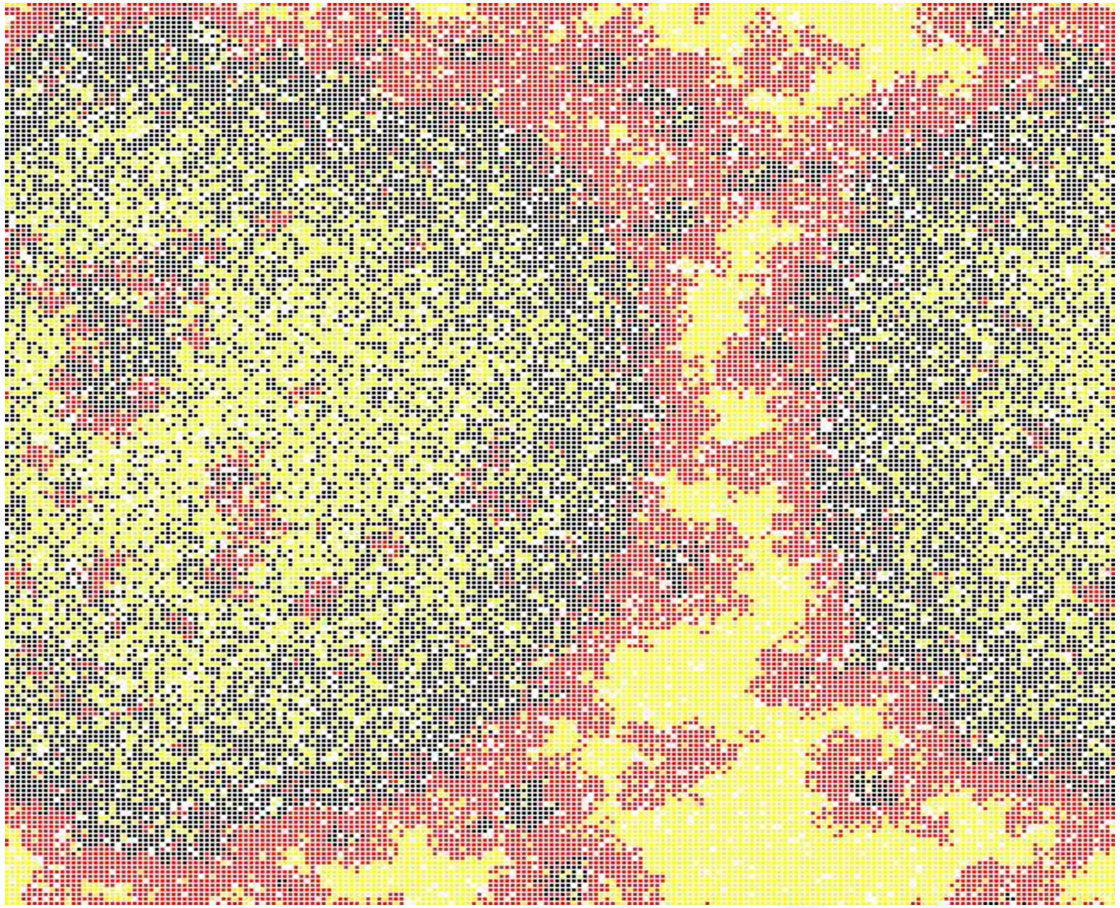
It’s a method of computational modelling consisting of three distinct elements:

- A set of agents
- A system of communication
- The environment

Together, these are referred to as *the System*.

Why ABMs?

- Self-organisation
- Emergent phenomena
- Spatial distribution
- Representation of real life
- Individual agency
- Pretty pictures

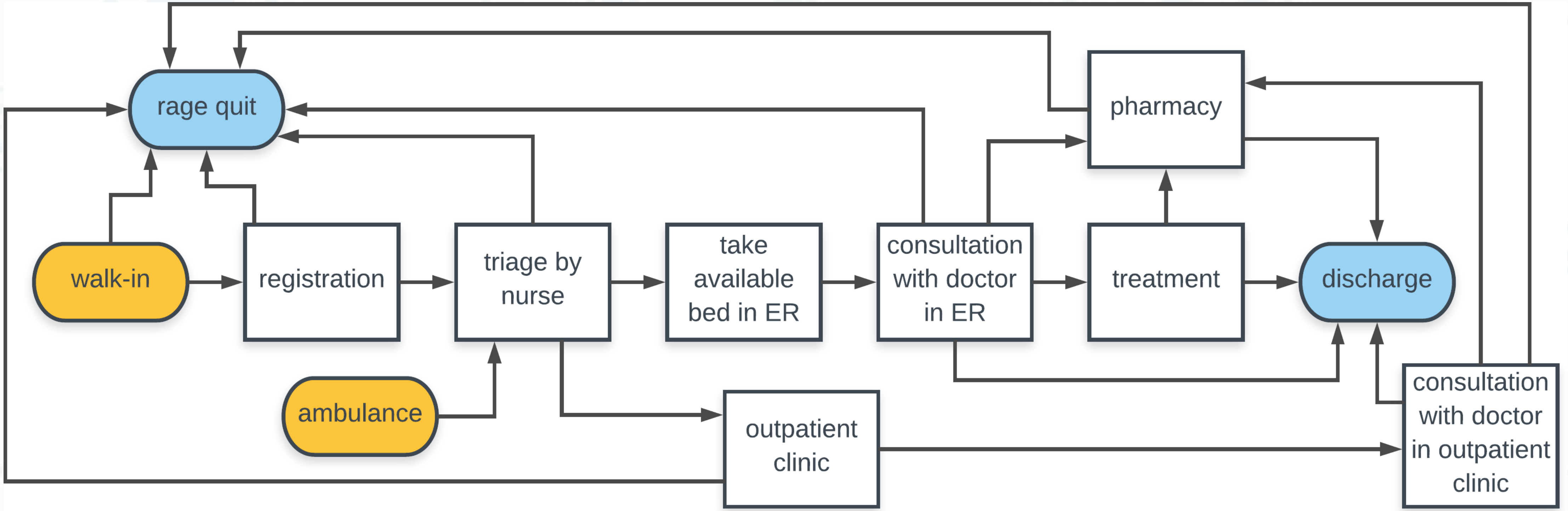


How do we make an ABM?

The System (I)

- aggregates all the elements of the ABM
- tailored towards the research questions to be answered
- constrains the action of agents and their influence on the environment
- any data fed to the simulation
- the group of other models available for comparison
- any data available for comparison of model outputs
- A visual representation of the specific model being used

Patient Path through the Riverdale General Hospital's Emergency Room



The Setup (I)

- This is a model of the Emergency Room flow at the Riverdale General Hospital (RGH), from arrival to discharge
- There will be increased patient load from the nearby Bedrock Teaching Hospital (BTH), closed for repairs after a recent fire
- Medical Director of the RGH, Monica Geller, is worried that the current model may perform inefficiently under the load of increased patients
- She has access to surplus doctors, nurses, other staff, medical equipment and beds from a nearby hospital that's closed for repairs

Question 1

a. Is the research question clear form the previous slides?

b. If not, what information do we need?

c. What are some of the solutions we can investigate?

The Setup (II)

- Ms. Geller has a specific concern: the time between triage and treatment for the worst patients.
- Your literature search has turned up three other methods used to solve similar problems: Discrete Event Simulation, System Dynamics and Queueing Theory
- All the others focus on events, rather than the individual patients.

Question 2

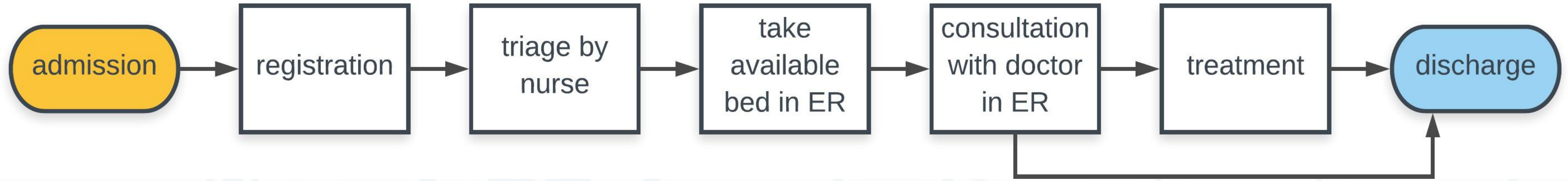
a. Why choose Agent-Based Modelling?

b. What are some pros and cons of the ABM approach?

The Setup (III)

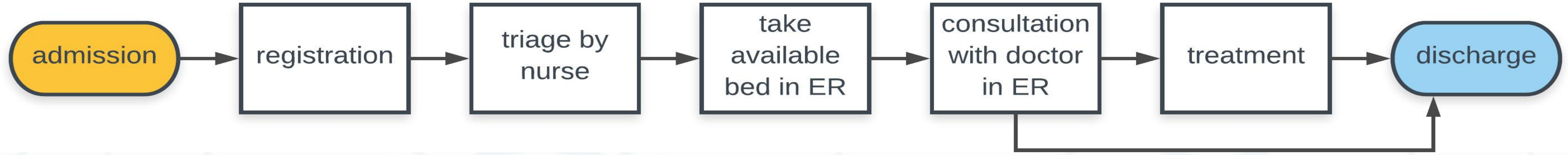
- Your PI says that the current model is too complex to start with
- She believes that the development time and computational requirements are too much
- She suggests a simplified model

Simplified Model of the RGH Emergency Room

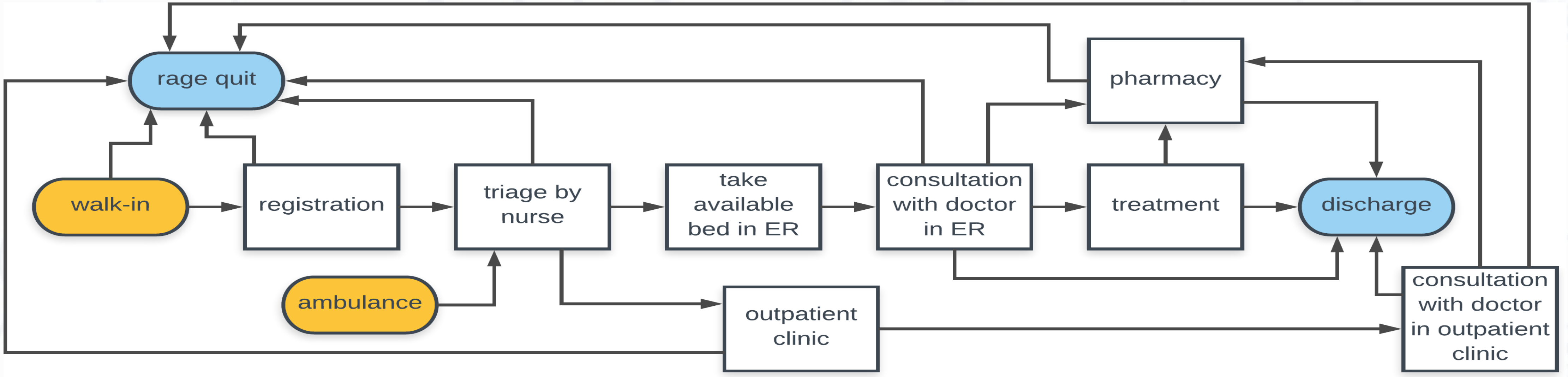


Question 3

Is this new model over-simplified?

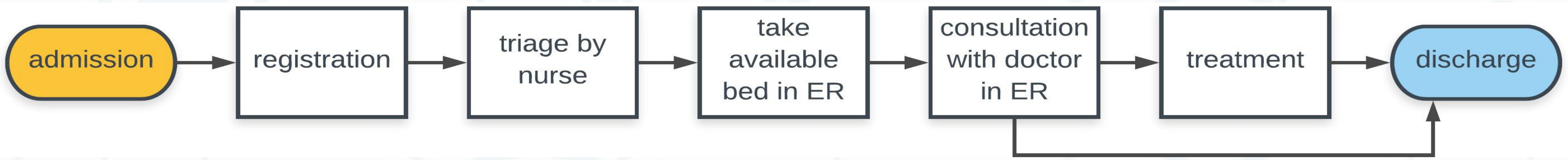


Old model:

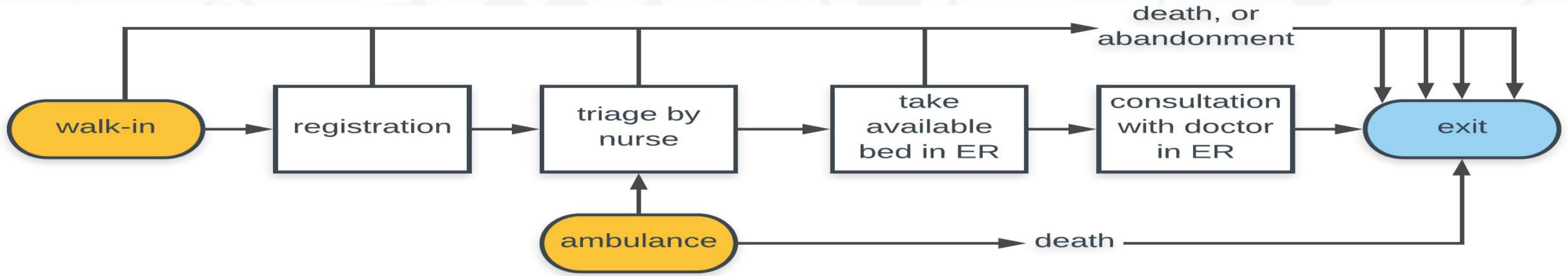


How about this?

Is this new model over-simplified?



My suggestion:



The Agent

- discrete and self-contained
- capable of autonomous action, and adaptation
- actions can be both passive and active
- has at least one characteristic dynamic “state”
- is able to communicate with other agents in the system
- is task- or goal- oriented
- locally heterogenous in the system (at least once)

The Agent

- Decide what the agents of the system represent
  - If they don't have dynamic properties necessary for data collection, the proposed "agents" may be better represented in other ways (data structures, for example)
  - They must be discrete and autonomous, and communicate with each other.
- What behaviours will the agents simulate?

Question 4

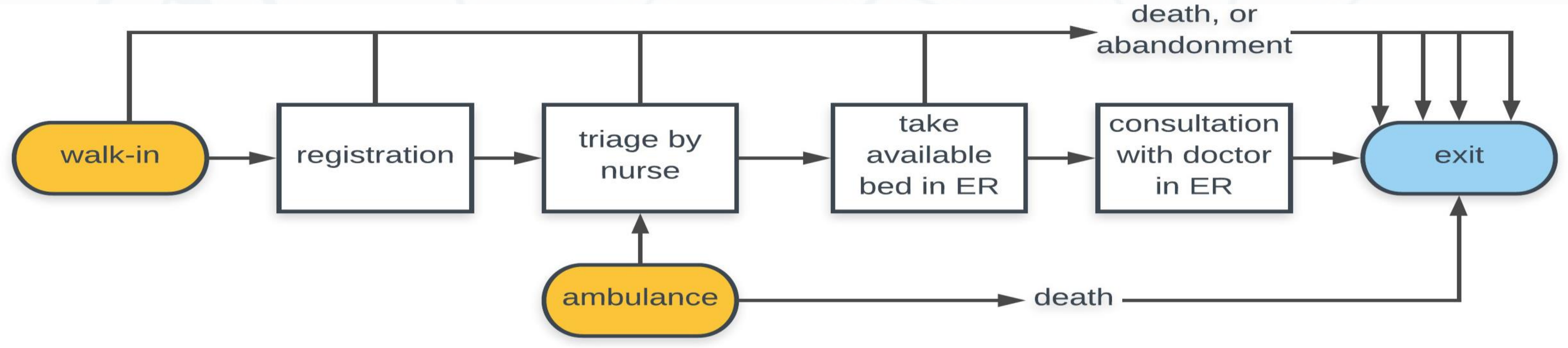
We have the final model!

Unfortunately, it's time for implementation...

What are the agents in the model?

Or rather, what should they be?

What are the Agents?



What would their characteristics (states) be?

Doctor **Nurse (ER)** **Patient** **Receptionist** **Bed** **Nurse (triage)**

Communication

The system of communication should at least answer all of the "Reporter's Questions".

Who?
What?
When?

Where?
Why?
How?

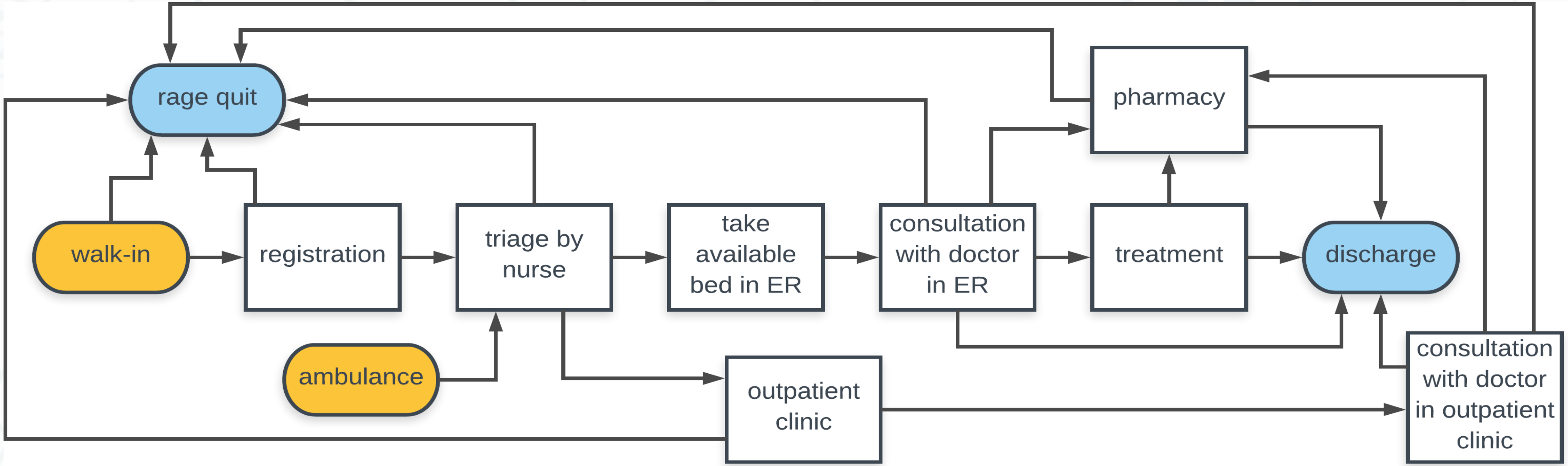
This is for the purpose of providing local information to the agents.

Communication

- Who: Which agents do we need to communicate with?
- What: What information will the agent need?
- When: At what points during the simulation will the agents talk to each other?
- Where: Will agent interaction constrained to any specific part of the network at any point in the simulation?
- Why: What will the new information be used for?
- How: Is there a specific order for agent communication, or can it be done randomly?

Question 5

What inter-agent communication was there in the family of models discussed so far?



The Environment

- a structure for communication among agents, called a *topology*
- all global variables and parameter values for the simulation
- contains the passive objects in the simulation

Environment

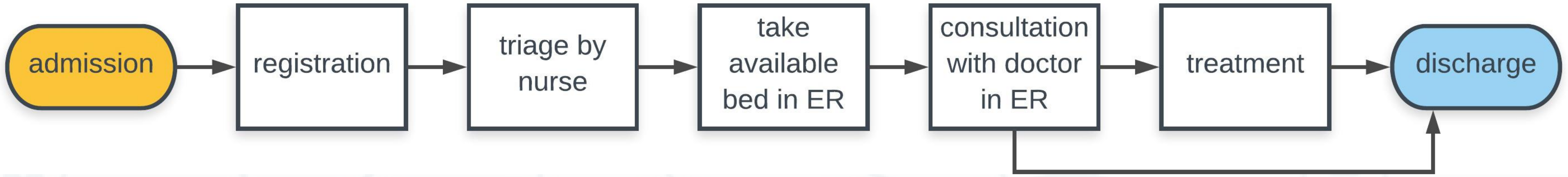
- What is the topology of the network?
- How will the model be parametrised? What are the global variables?
- Will the objects be placed on a grid (or some other structure), or is spatial positioning not fixed/important?
- What can possibly be changed by the agents?

The System (II)

- What constraints to we place on the agents?
  - Individual behaviour
  - Interaction
  - Influence on the environment (ability to change the structure of the simulation)
- Is there any extrinsic data to be incorporated into the model? Does this alter the structure of the proposed simulation?

Enter Statistics, Stage Right

Ms. Geller has just given us a table of net usage statistics for the RGH, in the next slide. She thinks that's what we're working towards, for this model. Before we use this data, we have to sift through it.



Question 5:
Which rows are irrelevant?

Number	Patient Numbers (daily average)	ER	Outpatient Clinic
1	Walk-ins	80	-
2	Patients arriving by ambulance	20	-
3	Patients dying	~ 1%	-
4	Patients leaving without being seen or treated	6	-
5	Patients referred to outpatient clinic	~ 7%	-
	Average Times		
6	Registration	5 mins	-
7	Triage	5 mins	-
8	Sanitise equipment and replace medical supplies used in the ER	6 mins	~ 30 sec
9	Pharmacy	10 mins	-
	Staff		
10	Number of doctors	5	2
11	Average number of doctors needed for a single patient	1	1
12	Doctor average time spent on one patient	40 mins	10 mins
13	Number of Nurses	40	3
14	Average number of nurses needed for a single patient	2	1
15	Nurse: average time spent by the first nurse on one patient in ER	40 mins	5 mins
16	Nurse: average time spent by the second nurse on one patient in ER	55 mins	-
17	Break time allotted to doctors and nurses in the ER between patient interactions	4 mins	-
18	Number of Pharmacists	1	-
	Resources		
19	Number of doctors total	23	-
20	Number of nurses total	40	-
21	Number of Pharmacists total	4	-
23	Total number of beds available for use	60	-
22	Capacity of beds in the ER	16	-
24	Number of Ambulances	10	-

Computation (I)

- Does the simulation *need* to be as fast as possible?
- How expensive is computational time?
- How good are you at coding?
- What’s the best process?
- Which programming language to use?
- Are there any available prebuilt packages to your purpose?

Computation (II)

- What do you think the scale of the final model will be?
- Is the complexity of the model reducible?
- Should you prototype the model first, or just build at full scale?
- How will the model be tested?
- How will the data be collected, output, post-processed (if necessary) and analysed?

Verification and Validation

Validation and Verification

Validation - “Did we use the right model???”

Verification – “Does the model do what we want it to do?”

These questions appear to be very similar (and are sometimes blurred in computational modelling), but actually need to be done distinctly.

“… model validation deals with building the right model ... model verification deals with building the model right.”

- Balci (1994)

Validation (I)

"The process of assessing the degree to which a computer model is an accurate representation of the real world from the perspective of the models [sic.] intended applications."

Ormerod, Rosewell (2009)

"... validation concerns whether the simulation is a good model of the real system, 'the target'. A model which can be relied on to reflect the behaviour of the target is 'valid'."

Abdou, Hamill, Gilbert (2012)

Validation (II)

Can be done by comparing the output of the simulation to real data, or the output of other validated models. Statistical testing or Visual Heuristic Analysis may be used.

DO NOT OVERFIT THE MODEL.
DO NOT OVERCOMPLICATE THE MODEL.

If the model is too complex (in an attempt to model the real world), or the parameters are tuned too tightly, you will potentially "black box" the model, and lose the ability to accurately extrapolate from the data.

Verification (I)

"The process of determining that a computational software implementation correctly represents a model of a process"

Ormerod , Rosewell (2009)

"The process of checking that a program does what it was planned to to is known as 'verification'. "

Abdou, Hamill, Gilbert (2012)

Verification (II)

Only the probability distribution of the results of the model can be predicted, so you must average an ensemble before you can establish a behaviour of the model.

Use a bunch of tests (edge cases, null cases, torture etc.) with known (or common sense) results to test the outputs of the simulation.

There may be a missing decimal point somewhere in your code, or a potential division by zero.

Concerns

There's no rest for the wicked…

Computational Concerns (I)

How big should the model be?

- Depends on what is being modelled, the research question, and the desired results.
- Agents in the simulation must also have enough communication to generate meaningful results. From the literature using ABMs, N = 10 000 seems to hit the spot, most times.
- Bear in mind that the scale of the model may impact the environment of the ABM, and the global parameter values chosen.

Computational Concerns (II)

How many time steps should I let the model run for?

- Depends on what a time step represents in the model.
- Time steps should be chosen according to the "shortest" process occurring *each time step*.
- Generally, more runs give more confidence in the results, but may be unnecessary, or may complicate data post-analysis and representation

Computational Concerns (III)

How big should the ensemble be?

- Depends on the behaviour you're looking for.
- Lots of papers use 20 trials, statistical theory says ~30, and 50 is safe for inexpensive simulations.
- The number of trials needed for stable results may also vary with respect to the parameter regime chosen.
- Report all results by stating *both* the mean and the standard deviation of the measurements taken over all the runs.

Computational Concerns (IV)

How do I choose an initial state / parameter regime for the model?

- There's really no good general answer here (that I've found).
- Choose reasonable parameter values, preferably based on previous literature. For others, take initial values that "make sense", and see what comes out the other end.
- For example: infection rate for a paediatric communicable disease – 80% ?
- This is a part of model validation (kinda…the distinction is blurred).

Question 6

a) What would a time step represent in his model?

b) How many runs should we do for each set of parameter values?

c) What specific pieces of data do we want to gather?

Here's that table again…

Number	Patient Numbers (daily average)	ER	Outpatient Clinic
1	Walk-ins	80	-
2	Patients arriving by ambulance	20	-
3	Patients dying	~ 1%	-
4	Patients leaving without being seen or treated	6	-
5	Patients referred to outpatient clinic	~ 7%	-
	Average Times		
6	Registration	5 mins	-
7	Triage	5 mins	-
8	Sanitise equipment and replace medical supplies used in the ER	6 mins	~ 30 sec
9	Pharmacy	10 mins	-
	Staff		
10	Number of doctors	5	2
11	Average number of doctors needed for a single patient	1	1
12	Doctor average time spent on one patient	40 mins	10 mins
13	Number of Nurses	40	3
14	Average number of nurses needed for a single patient	2	1
15	Nurse: average time spent by the first nurse on one patient in ER	40 mins	5 mins
16	Nurse: average time spent by the second nurse on one patient in ER	55 mins	-
17	Break time allotted to doctors and nurses in the ER between patient interactions	4 mins	-
18	Number of Pharmacists	1	-
	Resources		
19	Number of doctors total	23	-
20	Number of nurses total	40	-
21	Number of Pharmacists total	4	-
23	Total number of beds available for use	60	-
22	Capacity of beds in the ER	16	-
24	Number of Ambulances	10	-

Summary

for TL;DL purposes

Summary (I)

- Clearly define the research problem, and search the literature for anything similar, both to inform the model and to provide data for model comparison and validation.
- Draft the system first.
- Choose agents and their characteristics based on what normal people would do in the real system.
- Simplify as much as possible, for tractability and meaningful results.
- See what computational resources you have, and choose your weapon carefully.

Summary (II)

- Choose how the agents interact, and the parameters of the system.
- Write your code carefully. Verify and validate.
- Test your code before running it full scale!
- Think about how many trials you actually need for each parameter value.
- Name your data files carefully so they don't get overwritten
- **ALWAYS BACK UP YOUR DATA REGULARLY**
- Be conservative with extrapolations and inferences.

DO NOT OVERFIT THE MODEL

DO NOT OVERCOMPLICATE THE MODEL

Your model will fail, maybe harder than Google Flu did.