Component 2: Draft Detailed Design

## Introduction

The purpose of this design report is to further detail the intended design for an agent based model of Sheffield, investigating the effects of three new policies on the number of commuters in Sheffield who cycle. The policies are to be introduced by Sheffield City Council with the aim of increasing the proportion of cycling commuters from 2% to 11%. The three new policies to be investigated are:

1. Construct cycle routes that are separate to the road infrastructure.
2. Increase the amount of road space dedicated to cyclists.
3. Invest in subsidised pay-per-use docked electric bicycles that require less cycling effort on Sheffield’s hills.

The agents shall simulate groups of people of similar characteristics, with the choices made by those agents based partially on these characteristics, partially on temporal events and partially on a global perceived road safety metric. The agents shall represent both cyclist and non-cyclist commuters, with initial proportions matching that of the 2011 Sheffield census data.

## Policy Responses

In order to model the effects of these policies, they have been abstracted to a set of influences on the characteristics of the agents:

1. In creating new cycle routes separate to the road infrastructure, the cyclists would have an increased level of isolation from motor vehicles. This option would therefore set the perceived road safety metric to 0. However, building new routes around the existing road system may result in a longer route for cyclists. Therefore, a given journey distance will likely be increased by a constant factor.
2. In increasing the amount of road space dedicated to cyclists, the effective vehicle density is decreased, which would increase perceived road safety. To model this, the road safety function shall likely be increased by a constant factor for cyclists.
3. Finally, investing in subsidised electric bikes on a pay-per use system would reduce the physical effort required in going up hills. This can be represented in the simulation by setting any negative impact in choice due to hills to zero. However, a cost function will be added to the decision making process, providing a constant negative value.

## System Conceptual Schematic

Figure 1 represents the flow of data and the multiplicities of data components involved in the system. First, an initial conditions file will be read as input data to the model, which shall remain constant throughout the simulation. This shall contain configuration data of the agents or groups of agents, describing their home location and job locations. This is to model how commute distance and socio-economic status - based on their home location - might impact the effects of the new policies being implemented. The socio-economic rating shall be based upon secondary data relating to average house prices, crime statistics, council investments and infrastructure quality and will be based on a predefined grid map of Sheffield. This data shall also be supplied through the initial configuration file. Finally, the file will contain settings providing which policy or policies are to be implemented for a given run of the model.

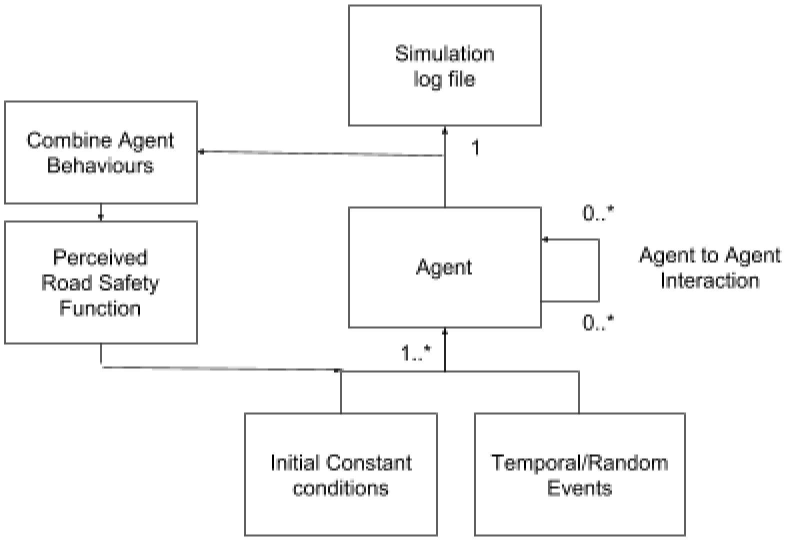

Figure 1

There shall also exist a temporal and random events generator. This will simulate seasonal effects through the 20 year simulation time of the model, such as weather conditions impacting willingness to cycle or seasonal infrastructure work such as road improvements before and after winter. This would simulate less predictable temporal events such as government investment into previously low investment areas, improving both socio-economic status over time and local infrastructure. It will also account for random events such as unplanned road closures and accidents which may affect traffic flow.

The configuration data will be fed into the agents at start time, whereas the temporal data shall be fed in per simulation cycle or “tick”.  The agents shall run, making their decisions based on input data as well as potentially an agent-to-agent data flow.

Per tick, the agents’ decision data shall be passed to the simulation log file and to a global combine function. This function, which will  effectively be another agent, shall be executed once per tick, but at a half tick delay to the commuter agents. This is to force them all to finish, and produce their output data before it is then processed by the combine function agent. This function sums the agents’ behaviours and creates a perceived road safety metric, which shall then be fed back into the agents on the next tick.

**System UML Class Diagram**

Figure 2 shows a basic program layout, not including any details for the Repast HPC functions. This shows the commuter agent is part of the composition of the main function. The commuter instances will be encapsulated into the main function and will be destroyed upon exit of the main function. The fields of the commuter agent mainly

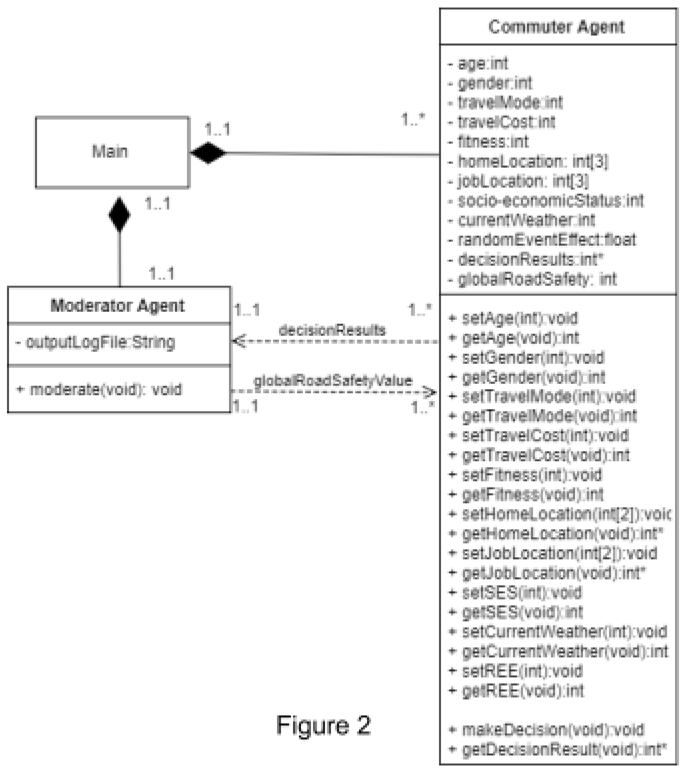

Figure 2

consist of the configuration data that each agent shall use in its per-tick-decisions. There is also a decision results variable which will be where the result of each tick simulation shall be stored. All data fields shall be accessed and modified using “get” and “set” functions, which are good practice in object-oriented programming. They help to limit access to variables when an object does not want or cannot allow them to be changed. The only other function relating to the commuter agent is “makeDecision”, which is the code that each agent shall execute once per simulation tick and will be scheduled by Repast HPC.

The moderator agent performs two functions. Firstly, it combines the agents’ behaviours as outlined above and also executes a global perceived safety function which creates a global perceived safety value. This takes the data from the commuter agents’ decision results using the respective “get” function. Secondly, it will log all relevant data to the simulation log file once per tick.

**Other Model Details: Simulation Details and Additional Desirable Features**

Given the model is intended to simulate a 20 year period, each tick of the simulation will correspond to a single working day in one week. As such, there will be a total of 1,040 ticks over the 20 year simulation. This value was selected because a daily simulation would produce more than the required level of data to show any significant trends which might develop over the 20 years. However, the chosen value provides a fine enough level of detail to observe the effect of seasonal changes, which are planned to be a key set of temporal input data to the model.

To further explore the effects of a working day on commuters, it may be worth creating a two-tick system to represent a morning and evening commute within each working day cycle. This would allow the simulation to account for differing factors between the two commutes, such as tiredness.

The final number of agents has yet to be decided as there is little available data to base this value on at this early stage. The literature review undertaken for this project has yet to provide useful information regarding use on desktop sized computers for models including this level of decision making. It is therefore hard to estimate presently the amount of compute time required for this model on the particular hardware planned for running the simulation. In order to account for this, the number of active agents within the model will be increased with each stage of development. This would mean that any later addition of features which drastically reduce simulation speed below a desirable level could be removed.

However, real world data could be used to set upper limit parameters for agent numbers within the simulation. According to 2011 census data, there are 244,000 working age adults between 16-74 living in Sheffield. The Sheffield City Region Transport Strategy document also specified that, from this 2011 census figure, an estimated 2% or 4,880 people cycle in order to commute to work.. As such, this would be a desirable upper target for the model in terms of the number of initial active cyclists.

As this project will be managed under an agile development strategy, following further research into Repast HPC and feedback from the previous component of this assignment, there are a number of features which could be added depending on available time and computing resources.  Firstly, it may be possible to link socio-economic status to home and work locations in Sheffield, as mentioned earlier in this report. Given location data is readily available, the model could be developed by adding road maps between common areas of the city. This would be used alongside Repast HPC's spatial simulation functions to look at transport densities. The transport densities data, along with a probabilistic function for calculating collision rates, could help develop a more accurate value for the global perception of road safety generated by the moderator agent.  The road location data would be taken from Ordnance Survey providers, and contains the real-world frequency of use for all the roads which can be mapped within the simulation. This could be taken as a template for the model's initial conditions.

Also relating to spatial simulation, the commute path could be considered. This would take into account the socio-economic status of areas travelled through where low scoring areas would put people off cycling for fear of personal safety. The temporal events generator could help the model account for variables such as government investment into infrastructure or area improvement. This could improve the global perceived road safety amongst agents, allowing the model to reflect any positive effects in increasing numbers of commuting cyclists as a result of temporal changes. Conversely, the model would also be able to account for any decline in global perceived road safety as a result of negative socio-economic impacts such as rising levels of crime. This would be reflected in a lower number of agents willing to commute via bike through these areas of the city.

Finally, implementing a fuzzy logic-based decision making system would be desirable. A fuzzy logic system would allow multi-dimensional continuous or non-continuous input data to be considered and programmed in an intuitive way, mimicking the process in which agents would act in the simulation. This could better represent the rationale of a person in their decision-making process without requiring the training that other neural networks would. A fuzzy logic system could ideally produce a simulation more accurate to real world decision-making, which would also make any subsequent data derived from it more significant.