Supplementary material I: Variable definition and calculation

Table S1: Definition and calculation of important variables used in the modelling

Variable	Definition	Calculation
DEM	Representation of a continuous terrain surface height	The terrain factors are derived using the 'Surface' function in ArcGIS
Slope	Representation of the steepest downhill slope	
Aspect	Identification of the downslope direction	
Distance to … (schools, transport, amenity, existing urban, roads…)	All the distance-related factors refer to the Euclidean distance between a target cell and the mentioned entity, such as a school, a bus stop, an existing urban pattern, a road, and etc.	The results of amenity and proximity factors are derived using the "Euclidean Distance" function in ArcGIS
CBD influence range	Three types of CBD are considered, the central CBD of Auckland, sub-district CBD, and suburban CBD†	$IR_{CBD} = W_1*CBD_1 + W_2*CBD_2 + W_3*CBD_3$ $(W_1+W_2+W_3 = 1)$
Population density	The number of people per square kilometre in a target cell, at the area unit scale due to the lack of data at the meshblock scale.	*Population / Area*
Ethnicity diversity	Represented by the clustering index of ethnicity for a community in a cell.	*Max (ethnic group) / Total Population*
Religious identity	Similar to ethnicity diversity, is also represented by the religion clustering index for each cell.	*Max (religion) / Total population*
Employment rate	The number of people employed in a census unit.	*Employed people / Total population*
Educational level	The proportion of population receiving a high education (higher than level four) in a cell, as shown in NZ census data.	*Sum (Edu. level 4 – Edu. level 9) / Total population*

†: In the 1990s, the CBD area of Auckland centred around Queen Street had a significant effect at all ranges. Back then there were no other CBDs, thus $IR_{CBD1990} = CBD_1$ (W_1 = 100%, W_2 = 0, W_3 = 0). As time passed, new sub-CBDs emerged at various locations, and they weakened the influence of the previous CBD. And these new sub-CBDs would influence more people nearby. Then the weight could be set to 0.3, 0.4, and 0.4, respectively, to calculate the new CBD influence at that time.

Supplementary material II: ANN Processing

Artificial Neural Network is a widely used ML approach with self-adapting, self-organizing, and self-learning attributes. This Adaboosted ANN model is trained and tested with random samples from the study area by applying a supervised learning procedure and the back-propagation (BP) algorithm to control the error-correction process. Half of the data samples are changed land use and the other half are unchanged, 70% of which are used for training, 15% for validation, and the remaining 15% for testing the accuracy of the model. The training is implemented iteratively with a target mean square error (MSE) of 0.005. After an initial weight is assigned to each input variable and each weak classifier, the ANN starts to "learn" by adjusting weights continuously between neurons in response to the MSE between the modelled and the observed values. The training will be terminated when the specified MSE threshold (0.005) is reached and became stabilized. The optimized weight for every variable and classifier representing agents' preference are finalized, and the trained ANN can be applied to predict the agents' decisions based on the input variables.

Supplementary material III: Garson's algorithm

The connecting weight W_{ij} (between input and hidden layers) and W_{jk} (between hidden and output layers) derived from the ANN can represent the input variables' weights on the output results. The Garson's algorithm is adopted to partition and quantify the connection weights between the input layer and the hidden layer, and between the hidden layer and the output layer. This processing is essential to determine the relative importance of the input variables and explore the heterogeneity of the agents. Garson's algorithm comprises a few steps. The first is to calculate the contribution of each input neuron to the output via a hidden neuron (Eq. A1):

$$C_{ji} = W_{ji} * W_{ki} \qquad \text{(A1)}$$

Where W_{ji} stands for the connection weight between input variable i and hidden node j; W_{jk} is the connection weight between hidden node j and output node k. The relative input of neuron m to the output of each hidden neuron is calculated as:

$$R_{jm} = |C_{jm}| / (|C_{j1}| + |C_{j2}| + \ldots + |C_{jk}|) \qquad \text{(A2)}$$

The contribution of all hidden notes to input neuron l is then summed as:

$$S_l = R_{1l} + R_{2l} + \ldots + R_{jl} \qquad \text{(A3)}$$

The final relative importance of input variable n is determined as:

$$RI_n = S_n / (S_1 + S_2 + \ldots + S_k) * 100 \qquad \text{(A4)}$$

Where RI_n is the relative importance value of the nth input variable.

Supplementary material IV: Model Environment

Computer type: Dell PC (Desktop computer) - Window 10 Enterprise

Processor: Inter(R) Core(TM) i15-7600 CPU @ 3.50Hz 3.50Hz

Installed memory (RAM): 8.00GB (7.86GB usable)

System type: 64-bit Operating System, x64 based processor

GIS Tool: ArcGIS 10.5.1 ESRI®

Simulation Tool: Matlab R2017b, Python 3.5