The Impact of Choice on Public Services.

Dr. Vincent Knight[1], Professor Paul Harper[1]

Cardiff University,Cardiff, UK

Abstract

It is well observed that individual behaviour can have an effect on the efficiency of queueing systems. The impact of this behaviour on the economic efficiency of public services is considered in this paper where we present results concerning the congestion related implications of decisions made by individuals when choosing between facilities. The work presented has important managerial implications at a public policy level when considering the effect of allowing individuals to choose between providers. We show that in general the introduction of choice in an already inefficient system will not have a negative effect. Introducing choice in a system that copes with demand will have a negative effect.

Keywords: Game Theory, Queueing Theory, OR in health services

1. Introduction

What damage do the selfish choices of some afflict on to the welfare of all? The work presented in this paper answers this question in the context of public service systems.

There is a substantial quantity of literature on the subject of equilibrium behaviour of a queueing system where congestion is a factor influencing behaviour [1, 2, 3, 4, 5, 6, 7]. This can be traced back to a series of short communications between Leeman [8, 9] and Saaty [10]. This paper builds on this literature by considering the problem of public service systems. For most

Email addresses: knightva@cf.ac.uk (Dr. Vincent Knight), harper@cf.ac.uk (Professor Paul Harper)

URL: www.vincent-knight.com (Dr. Vincent Knight), www.profpaulharper.com (Professor Paul Harper)

public services, congestion is a negative aspect of service quality. Examples of this are healthcare systems (waiting lists), transports systems (traffic jams) and/or schools (overcrowding of class rooms).

The degree of central control that should be exercised is a very important question to be considered by governments and/or policy makers. What is the effect of allowing individuals to choose service provider?

Of course, the motivation for the introduction of choice is to create competition in the hope that this would improve overall service quality. The aim of the work presented here is to use a game theoretical approach to measure the immediate effect of choice (i.e. prior to the effect of competition). The approach proposed is based on a measure called *the price of anarchy* [11, 12, 13, 14, 15, 16, 17, 18, 19, 20]. First introduced in [21] (a conference paper that has been reprinted in [22]), the price of anarchy is the ratio of the costs of the worst possible Nash equilibrium and the social optimum, and can therefore be interpreted as a measure of the efficiency of a system.

The main contributions of this paper are as follows:

- A novel connection is made: placing choice between public services within the formulation of routing games;
- Theoretical results are obtained as to the effect of demand and worth of service;
- It is shown that in a public service system with an adequate capacity to provide the perceived worth of service, a high price of anarchy is to be expected;
- A numerical approach based on heuristics is proposed to calculate the price of anarchy in a real world setting;
- The above ideas are demonstrated with a large scale real world case study.

The paper is organised as follows: Section 2 will give a brief overview of routing games; Section 3 will interpret choice of public services as a routing game; Section 4 will study a particular model that gives an insightful conclusion as to efficiency of general public service systems under choice; Section 5 looks at an application using hospital performance data for elective knee replacement surgery in Wales; Section 6 makes concluding remarks.

2. Routing Games

This section gives a brief introduction to routing games following very closely [23]. For a thorough overview of routing games the reader is encouraged to view [23, 18].

A *non atomic* routing game (atomic routing games will not be considered in this work), is defined on a network $G = (V, E)$, with vertex set V and edge set E, as well as a set of *source-sink* pairs: $\{(s_1, t_1), (s_2, t_2), \ldots, (s_k, t_k)\}$. These pairs are called *commodities*. The set of all possible (s_i, t_i) paths (for $i \in [k]$) is denoted as $\mathcal{P}_i$. Thus, $\mathcal{P}_i$ denotes the set of all the possible routes that player i may take to get from source s_i to sink t_i. Only networks with $\mathcal{P}_i \neq \emptyset$ for all $i \in [k]$ are considered and $\mathcal{P} = \bigcup_{i\in[k]} \mathcal{P}_i$. The graph G is allowed to have multiple edges and a vertex can participate in multiple source-sink pairs.

The routes taken by traffic are called *flows*, where $f \in \mathbb{R}^{|\mathcal{P}|}_{\geq 0}$ denotes a particular flow and f_P is interpreted as the quantity of traffic of commodity i choosing path P for $P \in \mathcal{P}_i$. A flow f is called *feasible* for $r \in \mathbb{R}^k_{\geq 0}$ if and only if $\sum_{P\in\mathcal{P}_i} f_P = r_i$ for all $i \in [k]$. Thus, the vector r denotes a prescribed quantity of traffic that must travel from sources to sinks.

What is now needed is some way of differentiating the various paths (indeed some paths may be *better* than others). Each edge $e \in E$ of G has a cost function $c_e : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ and it is assumed that c_e is non-negative, continuous, convex and non-decreasing. Using this, one can quantify the efficiency of a flow f; we define the cost of a flow $C(f)$:

$$C(f) := \sum_{P\in\mathcal{P}} c_P(f) f_p$$

where $c_P(f)$ naturally denotes the cost incurred by the traffic choosing path P. A routing game is then defined by the triple (G, r, c).

Importantly this cost function takes into account the quantity traveling through a particular path. It is immediate to give an equivalent definition:

$$C(f) := \sum_{e\in E} c_e(f) f_e \tag{1}$$

where f_e corresponds to the quantity of traffic using edge e.

Using this we give the following definitions:

Definition 2.1. *For the routing game (G, r, c), the flow f^* is an* optimal flow *if and only if f^* minimizes C (as given by (1)) over all feasible flows f.*

The next definition corresponds to an absence of central control. Note that the term "Wardrop equilibrium" [24] is also used however we choose to use the term "Nash flow" in line with [18].

Definition 2.2. *For the routing game (G, r, c), the flow $\tilde{f}$ is a* Nash flow *if and only if for every commodity $i \in [k]$ and every pair of paths P_1, P_2 with $\tilde{f}_{P_1} > 0$ we have:*

$$c_{P_1}(\tilde{f}) \leq c_{P_2}(\tilde{f})$$

Thus, $\tilde{f}$ is a Nash flow if and only if all used paths have minimum possible cost. This ensures that no user can improve their situation. We now state (without proof) two very powerful results obtained in [25].

Theorem 2.3. *The flow f^* is an optimal flow for (G, r, c) if and only if f^* is a Nash flow for the instance (G, r, c^*) where:*

$$c_e^*(x) = \frac{d}{dx} x c_e(x) = c_e(x) + x \frac{d}{dx} c_e(x) \tag{2}$$

The cost $c_e^*(x)$ is called the *marginal cost* for e. This powerful result shows that mathematically, Nash flows and optimal flows are analagous. The next result simply confirms this.

Theorem 2.4. *The flow $\tilde{f}$ is a Nash flow for (G, r, c) if and only if $\tilde{f}$ minimizes Φ where:*

$$\Phi(f) = \sum_{e \in E} \int_0^{f_e} c_e(x) dx \tag{3}$$

The function Φ is called the potential function for (G, r, c). As stated we give these results without proof but encourage the reader to see [23, 18].

The object of the work in this paper is to present a measure of the efficiency of a system, the measure we shall use is given by the following definition.

Definition 2.5. *For the routing game (G, r, c), the* price of anarchy *denoted by $PoA(G, r, c)$ is given by:*

$$PoA(G, r, c) = \frac{C(\tilde{f})}{C(f^*)}$$

Note that the definition taken here differs slightly to the common definition taken as the worst case ratio. The price of anarchy quantifies the inefficiency created by choice. We illustrate these ideas through a famous example, known as, Pigou's example [26].

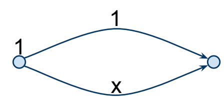

Figure 1: Pigou's example.

The network of Figure 1 corresponds to the routing game, where traffic has a choice of two paths to reach the sink. The upper arc corresponds to a large highway and the travel time is independent (say 1 hour) of the quantity of traffic using that highway. The lower arc corresponds to a much smaller road, that is heavily affected by congestion and the time spent on this road is equivalent to the proportion of traffic that uses it. It is immediate to note that the Nash flow of this game is given by $\tilde{f} = (0, 1)$, since all traffic will go along the smaller road (thus incurring an hour of travel time), in the hope that at least a small quantity of traffic will use the larger road. The cost function for this game is:

$$C(f_1, f_2) = C(1 - x, x) = 1 - x + x^2$$

thus, the optimal flow is $f^* = \left(\frac{1}{2}, \frac{1}{2}\right)$. It is then straightforward to calculate the price of anarchy: $PoA(G, r, c) = \frac{4}{3}$.

Note that the potential function for this game is:

$$\Phi(f_1, f_2) = \Phi(1 - x, x) = \int_0^{1-x} 1dx + \int_0^x xdx = 1 - x + \frac{x^2}{2}$$

minimising Φ gives $\tilde{f} = (0, 1)$ as required. For this simple game, calculating the Nash flow does not require using the potential function. However, optimising Φ gives an algorithmic approach. This will be the method used when calculating the price of anarchy for large systems.

In the next section we show how these ideas will be used to measure the efficiency of a public service system.

3. Public service choice and routing games

When choosing amongst public service systems it is reasonable to assume that individuals take into account three aspects: reputation, distance and congestion. As will become clear, our game theoretical model will take these factors in to account.

The first of these is a subjective aspect that could depend on individual satisfaction, word of mouth, and/or an official rating. It is accepted that reputation itself is dependent on the sequences of the various service centre visits, however the way in which the opinion of a particular individual diffuses through its social network will not be considered in this work. As such we assume that reputation is a constant parameter. This is in line with the previous mentioned idea of modelling choice for a given system (without letting competition play a role). Travel distances, although (arguably) subject to individual perception are also to be assumed constant (i.e. do not vary with the decisions made by individuals). Finally, waiting time is an extremely dynamic factor that depends on the choices of individuals in the system. This is what justifies using a game theoretical approach. The model we propose takes these three factors in to account in a linear way to create a non-atomic routing game. We allow for multiple demand and service nodes. Each demand node will have a unique demand rate whilst each service node will offer a different level of attractiveness (potentially based on reputation, congestion and expected length of stay). The pairing of these nodes is heterogeneous in nature due to travel distances.

In particular let m and n be the number of demand and service nodes respectively. We have a distance matrix $d \in \mathbb{R}_{\geq 0}^{m\times n}$, where d_{ij} denotes the distances between demand node i and service node j. Every demand node has attributed to it a demand rate Λ_i for all $i \in [m]$. This demand rate is to be measured in procedures per unit time. Every service node has attributed to it a cost function $w_j : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ with arguments corresponding to the arrival rate at that node. We assume that w_j is a continuous, non-negative, convex and non-decreasing function. This assumption is not a restrictive one. Non-negativity is a question of scale, and the non-decreasing constraint simply implies that facilities become less attractive with increased congestion. It can be argued that facilities with high volumes of a particular service deliver better quality but how facilities respond to high volumes corresponds to the modelling of competition that arises as a product of choice and is not the purpose of this work.

Finally one last input to our model is a reward for service $\beta_i \in \mathbb{R}_{\geq 0}$ as considered in [1, 2, 4, 27, 6, 7]. By indexing the vector β by demand nodes the model we propose allows a partitioning of the population not only by geography but also by gender, morbidity and/or economic status which are all factors that will influence the choices of individuals [28, 29, 30]. The important modification to be made here is in the interpretation of β_i. As mentioned in Section 2, to allow for this situation to be considered as a routing game one needs all latency functions to be non-negative. For this reason we call β_i a *cost of balking* as opposed to a reward for service. There are two immediate interpretations that can be given to β_i:

- Individuals experiencing a dissatisfaction as a result of deciding that none of the options where of high enough quality (i.e. low enough cost).
- Individuals opting to seek treatment through a private service at a particular financial cost.

Note that these interpretations of β_i lend themselves well to considerations of public service systems, where equality is often as important as efficiency [31]. For example, in countries with public health systems such as the UK a series of waiting time targets are often imposed which may prove insightful when seeking to parametrize β.

The decision variables are the λ_{ij}; the arrival rate from demand node i to service node j. The balking rate from node i is: $\Lambda_i - \sum_{j=1}^{n} \lambda_{ij}$. Figure 2 gives a pictorial representation of the general system.

This can be interpreted in the form of a routing game with the latencies of each edge given, as shown in Figure 3.

Note that the sources of the routing game of Figure 3 correspond to the demand nodes, however there is only one sink. This sink corresponds to the state of "having made a choice". The choice options in question correspond to the various paths. All commodities have $n+1$ possible paths, corresponding to the n service nodes as well as the balking option. From (1) and (3) we have the cost function and potential function given by:

$$C(\lambda) = \sum_{i=1}^{m} \alpha_i \sum_{j}^{n} d_{ij}\lambda_{ij} + \sum_{j=1}^{n}\sum_{i=1}^{m} \lambda_{ij} w_j \left(\sum_{i=1}^{m} \lambda_{ij}\right) + \sum_{i=1}^{m} \beta_i \left(\Lambda_i - \sum_{j=1}^{n} \lambda_{ij}\right) \tag{4}$$

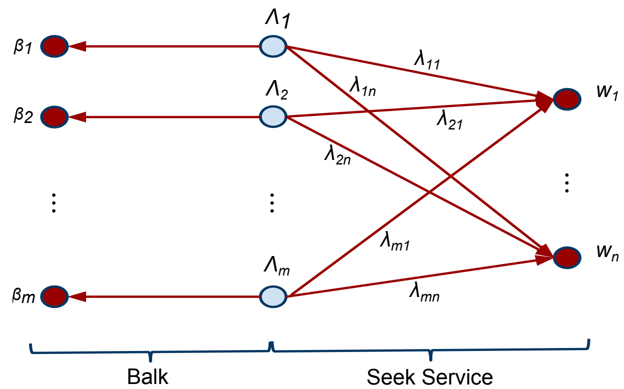

Figure 2: Pictorial representation of choice strategies

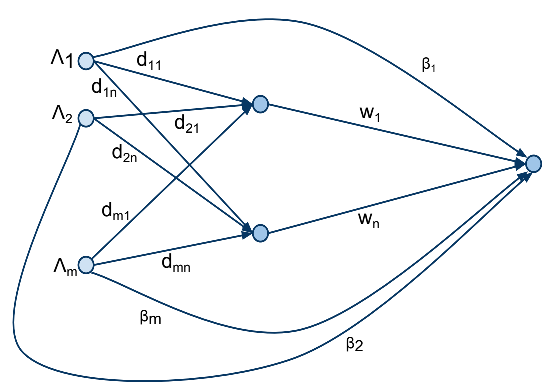

Figure 3: Routing game representation of choice strategies

and:

$$\Phi(\lambda)=\sum_{i=1}^{m}\alpha_i\sum_{j}^{n}d_{ij}\lambda_{ij}+\sum_{j=1}^{n}\int_0^{\sum_{i=1}^{m}\lambda_{ij}}w_j(x)dx+\sum_{i=1}^{m}\beta_i\left(\Lambda_i-\sum_{j=1}^{n}\lambda_{ij}\right) \quad (5)$$

The constant $\alpha_i \in \mathbb{R}_{\geq 0}$ is simply a weighting statistic for the relative importance of travel distances to the other factors (once again allowing for this coefficient to be dependent on population partitioning). Using this formulation of choice we obtain insights in to the asymptotic behaviour of the price of anarchy. The first considers the sensitivity to the value of service:

Theorem 3.1. *Assuming constant Λ we have:*

$$\lim_{\beta_i\to\infty} PoA(\beta)<\infty \text{ for all } i\in[m]$$

The price of anarchy increases with worth of service, up to a point.

Proof. Recalling Definition 2.5 we have:

$$PoA=PoA(\beta)=\frac{C(\lambda^*)}{C(\tilde{\lambda})}=\frac{C(\lambda^*(\beta))}{C(\tilde{\lambda}(\beta))}$$

Importantly we also have:

$$\lim_{\beta_i\to\infty}\lambda^*(\beta)=k^* \quad (6)$$

and $k^*\in\mathbb{R}_{\geq 0}^{m\times n}$. Indeed, as β_i increases, the arrival rate from demand node i increases until the value of service is large enough such that there are no more balkers. Similarly:

$$\lim_{\beta_i\to\infty}\tilde{\lambda}(\beta)=\tilde{k} \quad (7)$$

Thus we have:

$$\lim_{\beta_i\to\infty}\sum_{i'=1}^{m}\left(\sum_{j=1}^{n}\lambda^*_{i'j}-\Lambda_{i'}\right)=\lim_{\beta_i\to\infty}\sum_{i'=1}^{m}\left(\sum_{j=1}^{n}\tilde{\lambda}_{i'j}-\Lambda_{i'}\right)=0 \quad (8)$$

Recalling (4) we have:

$$PoA(\beta)=\frac{\sum_{i=1}^{m}\alpha_i\sum_{j}^{n}d_{ij}\lambda^*_{ij}+\sum_{j=1}^{n}\sum_{i=1}^{m}\lambda^*_{ij}w_j\left(\sum_{i=1}^{m}\lambda^*_{ij}\right)+\sum_{i=1}^{m}\beta_i\left(\Lambda_i-\sum_{j=1}^{n}\lambda^*_{ij}\right)}{\alpha\sum_{j}^{n}\sum_{i=1}^{m}d_{ij}\tilde{\lambda}_{ij}+\sum_{j=1}^{n}\sum_{i=1}^{m}\tilde{\lambda}_{ij}w_j\left(\sum_{i=1}^{m}\tilde{\lambda}_{ij}\right)+\beta\left(\sum_{i=1}^{m}\left(\Lambda_i-\sum_{j=1}^{n}\tilde{\lambda}_{ij}\right)\right)}$$

Recalling (6),(7) and (8) we get:

$$\lim_{\beta_i \to \infty} PoA(\beta) = \frac{\sum_{i=1}^{m} \alpha_i \sum_{j}^{n} d_{ij} k^*_{ij} + \sum_{j=1}^{n} \sum_{i=1}^{m} k^*_{ij} w_j \left(\sum_{i=1}^{m} k^*_{ij} \right)}{\sum_{i=1}^{m} \alpha_i \sum_{j}^{n} d_{ij} \tilde{k}_{ij} + \sum_{j=1}^{n} \sum_{i=1}^{m} \tilde{k}_{ij} w_j \left(\sum_{i=1}^{m} \tilde{k}_{ij} \right)} < \infty$$

□

Note (as demonstrated by the proof) that the case of a high value of β_i corresponds to a model with no balking allowed. Such a model was studied in [2, 15] where homogeneous agents are presented with a choice of multiple $M/M/1$ facilities with no option of balking. In these two papers, geographic location of demand is not taken in to account and congestion is thus the only factor dictating the choices. In [15] it is shown that the price of anarchy is bounded by the total number of facilities. Whether or not such an upper bound can be obtained for the more general model we have presented is an interesting research question.

The following result considers the effect of demand on the price of anarchy.

Theorem 3.2. *Assuming constant $\beta \in \mathbb{R}^m_{\geq 0}$ we have:*

$$\lim_{\Lambda_i \to \infty} PoA(\Lambda) = 1$$

In a system with insufficient capacity the price of anarchy is low.

Proof. We have:

$$\lim_{\Lambda_i \to \infty} \lambda^*(\Lambda) = l^* \tag{9}$$

for some $l^* \in \mathbb{R}^{m \times n}_{\geq 0}$. Indeed, as demand increases, each facility has an upper bound for it's total arrival rate after which, due to β_i it becomes more advantageous for agents to balk. Similarly

$$\lim_{\Lambda_i \to \infty} \tilde{\lambda}(\Lambda) = \tilde{l} \tag{10}$$

for some $\tilde{l} \in \mathbb{R}^{m\times n}_{\geq 0}$. We can write the expression for the PoA as:

$$PoA(\beta) = \frac{\frac{1}{\Lambda_i}\left(\sum_{i'=1}^{m}\alpha_{i'}\sum_{j}^{n}d_{i'j}\lambda^*_{i'j} + \sum_{j=1}^{n}\sum_{i'=1}^{m}\lambda^*_{i'j}w_j\left(\sum_{i'=1}^{m}\lambda^*_{i'j}\right) + \sum_{\substack{i'=1\\ i'\neq i}}^{m}\beta_{i'}\left(\Lambda_{i'} - \sum_{j=1}^{n}\lambda^*_{i'j}\right) - \lambda^*_{ij}\right) + \beta_i}{\frac{1}{\Lambda_i}\left(\sum_{i'=1}^{m}\alpha_{i'}\sum_{j}^{n}d_{i'j}\tilde{\lambda}_{i'j} + \sum_{j=1}^{n}\sum_{i'=1}^{m}\tilde{\lambda}_{i'j}w_j\left(\sum_{i'=1}^{m}\tilde{\lambda}_{i'j}\right) + \sum_{\substack{i'=1\\ i'\neq i}}^{m}\beta_{i'}\left(\Lambda_{i'} - \sum_{j=1}^{n}\tilde{\lambda}_{i'j}\right) - \tilde{\lambda}_{ij}\right) + \beta_i} \quad \text{for all } i \in [m]$$

and so (recalling (9) and (10)) taking $\Lambda_i \to \infty$ gives the required result. □

This shows that for a bad system, i.e. a system that is unable to cope with the total demand, choice does not create a relatively large level of inefficiency. The fact that the price of anarchy is low when the congestion in the system grows is a phenomenon identified in [13]. Theorem 3.2 seems to indicate that there is a tipping point at which the price of anarchy starts to decrease and tends to one. In Section 4 we will give a result that identifies this tipping point for a particular case and observe it empirically in Section 5.

Note that Theorems 3.1 and 3.2, extend the work presented in [15, 17, 18, 19]. In the next section we consider a particular case of the routing game model which corresponds to a general case of public service systems.

4. Price of anarchy in a single source single facility setting

Consider the simplest situation of Figure 3 i.e. a single source and a single facility. The options offered to individuals are thus to seek service or to balk. This is illustrated in Figure 4:

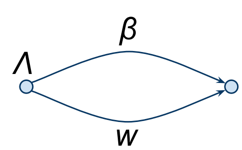

Figure 4: Routing game corresponding to a single source and a single facility.

This scenario, which could arise with a single population and a single service centre does not necessitate the inclusion of distance which can be taken in to account when setting a value of β and as such we assume $\alpha = 0$ (note that for simplicity of notation we suppress the indices i and j). Importantly this simple model can be used to model behaviour at a much higher level at which individuals are not choosing to use a particular facility but are at a national level choosing whether or not to use the public system. Individuals have the choice between balking and a freely provided service that becomes less attractive as congestion increases. Recalling (4) and (5), we have:

$$PoA(\Lambda) = \frac{\tilde{\lambda} w(\tilde{\lambda}) + \beta(\Lambda - \tilde{\lambda})}{\lambda^* w(\lambda^*) + \beta(\Lambda - \lambda^*)} \tag{11}$$

We give the following result that identifies the tipping point discussed after the proof of Theorem 3.2.

Theorem 4.1. *Assuming the particular model of Figure 4 we have:*

$$\sup_{\Lambda \geq 0} PoA(\Lambda) = PoA(\tilde{x})$$

where $\tilde{x}$ is the solution of $w(x) = \beta$.

Choice causes the highest level of inefficiency when the capacity of the system matches the perceived worth of service.

Proof. It is immediate to note that when $\Lambda > \tilde{x}$ we have $\tilde{\lambda} = \tilde{x}$, as when demand is sufficiently high for the congestion to imply a cost of facility equal to β, then all extra demand will result in balkers. Similarly we have for $\Lambda > x^*$, $\lambda^* = x^*$ where x^* is a solution of $w^*(x) = \beta$ (recalling (2) and Theorem 2.3).

Note that we also have due to the non decreasing and convexity properties of w: $w(x) < w^*(x)$ for all $x > 0$ which implies $x^* < \tilde{x}$ as illustrated by Figure 5.

We now consider (11) over the three intervals: $[0, x^*[$, $[x^*, \tilde{x}[$, $[\tilde{x}, \infty[$

- For $\Lambda \leq x^*$ we have $\tilde{\lambda} = \lambda^* = \Lambda$ so that:

$$PoA(\Lambda) = 1 \text{ for } \Lambda \leq x^* \tag{12}$$

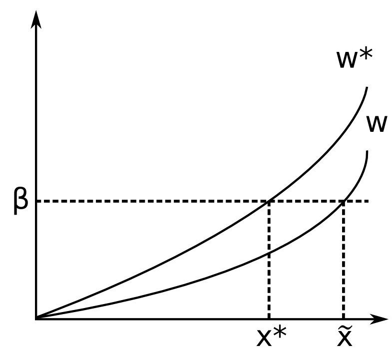

Figure 5: Graphical proof of $x^* < \tilde{x}$

- For $x^* < \Lambda \leq \tilde{x}$, recalling (11) we now consider the derivative of $PoA(\Lambda)$:

$$\frac{\partial PoA}{\partial \Lambda} = \frac{\partial}{\partial \Lambda} \frac{\Lambda w(\Lambda)}{x^* w(x^*) + \beta(\Lambda - x^*)} = \frac{w^*(\Lambda)(x^* w(x^*) + \beta(\Lambda - x^*)) - \Lambda w(\Lambda)\beta}{(x^* w(x^*) + \beta(\Lambda - x^*))^2} \tag{13}$$

Evaluating the sign of (13) is not straightforward however recalling that $PoA(\Lambda) \geq 1$ (by definition) and (12) it is sufficient to show that the sign of (13) does not change over $x^* < \Lambda \leq \tilde{x}$ to imply:

$$\frac{\partial PoA}{\partial \Lambda} > 0 \text{ for } x^* < \Lambda \leq \tilde{x}$$

The derivative of the numerator of (13) is given by:

$$\frac{\partial}{\partial \Lambda}(w^*(\Lambda)(x^* w(x^*) + \beta(\Lambda - x^*)) - \Lambda w(\Lambda)\beta) = \frac{\partial w^*(\Lambda)}{\partial \Lambda}(x^* w(x^*) + \beta(\Lambda - x^*)) \tag{14}$$

and as $\Lambda > x^*$ and $\frac{\partial w^*(\Lambda)}{\partial \Lambda} > 0$ this implies as required that $PoA(\Lambda)$ increases for $x^* < \Lambda \leq \tilde{x}$.

- For $\Lambda > \tilde{x}$, we have:

$$\frac{\partial PoA}{\partial \Lambda} = \frac{\partial}{\partial \Lambda} \frac{\beta \Lambda}{x^* w(x^*) + \beta(\Lambda - x^*)} = \frac{\beta x^*(w(x^*) - \beta)}{(x^* w(x^*) + \beta(\Lambda - x^*))^2} \tag{15}$$

However as shown in Figure 5, $w(x^*) < \beta$, which completes the proof as summarised by Figure 6.

$\square$

This result is insightful as it shows that the worst case equilibria is at a point at which the demand saturates the service to the point of worth of service. Recalling Theorem 3.2, we see that a system with a very high level of demand has a low price of anarchy (a fact noted in [13]), as does a system with a low level of demand. This leaves us with the situation shown in Figure 6 implying that a system with capacity that provides a level of service near the perceived worth of service will have a high price of anarchy.

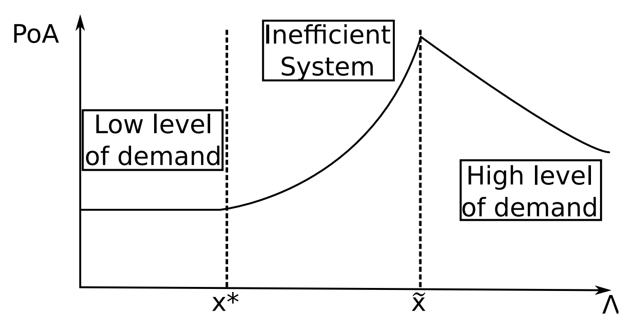

Figure 6: Plot of the price of anarchy for a generalised public service system

Figure 7 gives an initial confirmation of the above result through an example equivalent to a fictitious model of a public schooling system serving three neighbourhoods with two schools. Is is assumed that the "cost" (congested related utility) of each of the schools is the same and given by $w(x) = e^{x-20}$ so that for a class congested by less than 20 students we have a relatively low cost. In this case the cost of balking is equivalent to choosing to send students to a different option than the 2 schools immediately available. This game was solved numerically using [32].

In the next section we propose a further example: an applied case study in a healthcare setting.

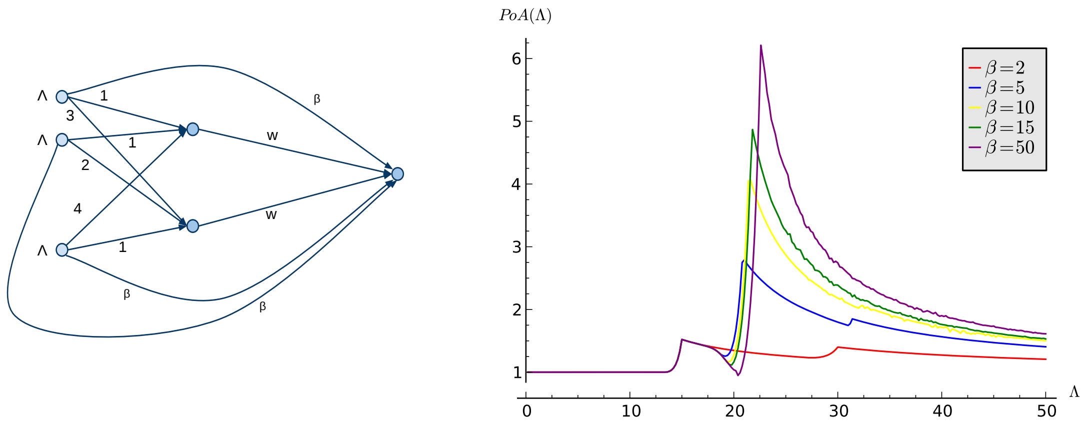

Figure 7: Education Example

5. Price of anarchy in the Welsh health service

Since 1948, the National Health Service (NHS) in the United Kingdom has given free healthcare to all patients at the point of service. Through the years this service has evolved into what is currently the largest employer in the UK. One of the most potentially significant changes was introduced in 2009, when free choice of service provider was offered to all patients in England [33]. In [34], the Welsh First Minister Rhodri Morgan claimed that NHS choice is 'not relevant to Wales'. This claim has motivated the work presented in this paper: *How does choice effect the efficiency of a public healthcare system?*

The ideas of the previous section are now demonstrated on a large network of service and demand nodes. To the author's knowledge, this is the largest such study presented in the literature. The case study chosen is the choice of facility for elective knee replacement surgery across Wales. Using hospital data for such procedures spanning a period from 29/01/2008 to 29/03/2010 (26 months) 14 service nodes and 22 demand nodes are identified. They are shown in Figure 8 and summarized in Tables 1 and 2.

For the purposes of this case study the measure we will use for the attractiveness of each facility is the expected time spent in the system (note that this includes time spent on the waiting list).

We assume that every facility is an $M/M/c$ queue. For the readers unfa-

Figure 8: Service nodes (crosses) and demand nodes (flags) in Wales

miliar with Kendall's notation, this implies that a facility acts like a service centre with random (negative exponential distribution) inter arrival and service times and c servers. The service time distribution corresponds to the length of stay. The c servers correspond to the capacity of each facility (i.e. the number of beds), as shown in Figure 9.

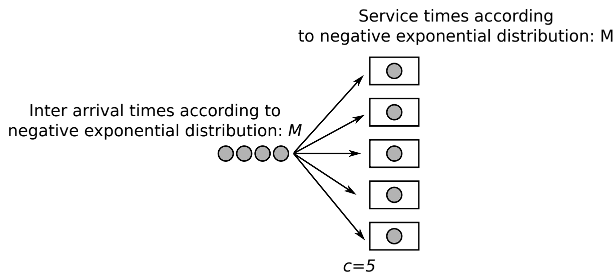

Figure 9: Pictorial representation of an $M/M/c$ queue.

Given an average arrival rate λ and an average service rate μ, such that $0 \leq \rho = \frac{\lambda}{c\mu} < 1$, the mean time spent in the system for an $M/M/c$ queue is given by ([35]):

$$W = W(\lambda, \mu, c) = \frac{1}{\mu} + \frac{\frac{\left(\frac{\lambda}{\mu}\right)^c}{c!c\mu\left(1-\frac{\lambda}{c\mu}\right)^2}}{\sum_{n=0}^{c-1}\frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} + \frac{\left(\frac{\lambda}{\mu}\right)^c}{c!\left(1-\frac{\lambda}{c\mu}\right)}} \tag{16}$$

We thus have:

$$w_j(\lambda) = \frac{1}{\mu_j} + \frac{\frac{\left(\frac{\lambda}{\mu_j}\right)^{c_j}}{c_j!c_j\mu_j\left(1-\frac{\lambda}{c_j\mu_j}\right)^2}}{\sum_{n=0}^{c_j-1}\frac{\left(\frac{\lambda}{\mu_j}\right)^n}{n!} + \frac{\left(\frac{\lambda}{\mu_j}\right)^{c_j}}{c_j!\left(1-\frac{\lambda}{c_j\mu_j}\right)}} \text{ for all } j \in [n] \tag{17}$$

where μ_j and c_j are the service rates and number of servers at service node $j \in [n]$.

Convexity of (16) is a known result [36] and so combining functions (4), (5) and (17) with Theorem 2.4 gives two mathematical programs with solutions; the optimal flow and Nash flow for any public healthcare facility system: OPTMP and NASHMP.

OPTMP: minimise (4) | NASHMP: minimise (5)

such that:

$$\sum_{j=1}^{n} \lambda_{ij} \leq \Lambda_i \text{ for all } i \in [m] \tag{18}$$

$$\lambda_{ij} \in \mathbb{R}_{\geq 0}^{m \times n} \text{ for all } i \in [m],\ j \in [n] \tag{19}$$

$$\sum_{i=1}^{m} \lambda_{ij} < c_j \mu_j \text{ for all } j \in [n] \tag{20}$$

Constraints (18) and (19) ensure non-negativity of all flows. Constraint (20) ensures that all our facilities are in steady state, i.e. that the arrival rate at all facilities is less than the corresponding service rate. This later constraint is not restrictive given a sensible choice of β. In fact constraint (20) has an interpretation where it is assumed that the expected wait at a facility that is not running at steady state is extremely large.

An approximate capacity for each of the 14 units is calculated, but due to sharing of wards between specialties, it is difficult to obtain precise bed numbers for a particular surgery/unit. The 22 demand nodes correspond to the 22 unitary areas of Wales and demand rates for each are obtained from the data set. The properties of each hospital and demand node are given in Tables 1 and 2. A distance matrix for these nodes was obtained using Google Maps Javascript API. For the purpose of this study a value of $\alpha = .8$ is taken, this is based on research presented in [28, 37].

The two mathematical programs OPTMP and NASHMP thus require the optimisation of (4) and (5) over 308 variables. A genetic algorithm implemented in [38] is used. A population size of 200 is used with crossover rate of .5 and a dynamic mutation rate dependent on the overall progress of the algorithm. Further research could be directed at finding a direct method for solving the routing games considered similar to the work of [39].

Two distinct aspects are considered. Firstly we aim to quantify the role played by the cost of balking β. For the demand rates of Table 2 the price of anarchy is calculated for varying levels of β and this simply examines the effect of the service value on the price of anarchy. The results are shown in

Table 1: Service node properties

Service Node	Capacity (beds)	Service Rate (patients per day)
Llandough Hospital	14	0.147
Royal Gwent Hospital	9	0.125
Prince Philip Hospital	7	0.132
Neath Port Talbot Hospital	4	0.151
Ysbyty Gwynedd	4	0.156
Morriston Hospital	4	0.148
Nevill Hall Hospital	4	0.136
Abergele Hospital	4	0.123
The Royal Glamorgan Hospital	4	0.141
Withybush General Hospital	3	0.170
Wrexham Maelor Hospital	2	0.153
Prince Charles Hospital	2	0.125
Bronglais General Hospital	1	0.171
Princess of Wales Hospital	1	0.100

Figure 10 which is an illustration of Theorem 3.1 as the price of anarchy is increasing but seemingly converging to some upper bound.

With an ageing population the demand for healthcare services is ever increasing. The next experiment considers the effect of demand on the price of anarchy. Using the same distribution of demand (as given by Table 2) sensitivity analysis for an overall demand rate $\Lambda = \sum_{i=1}^{m} \Lambda_i$ and for $\beta \in \{40, 70, 100\}$ is conducted. The results are presented in Figure 11.

6. Conclusions and further work

The work presented here has attempted to quantify the effect of choice on a public service system. Using theoretical results from routing game theory, generalised conclusions are obtained regarding the price of anarchy and an algorithmic approach for obtaining the price of anarchy is demonstrated. This measure takes in to account the geographical disposition of the demand and service nodes, as well as various factors that influence choice such as the congestion of a particular service centre.

Table 2: Demand node properties

Demand Node	Inter Arrival Rate (patients per day)
Cardiff	1.22
Carmarthenshire	0.92
Rhondda Cynon Taff	0.71
Caerphilly	0.61
Pembrokeshire	0.61
Swansea	0.51
Bridgend	0.51
Newport	0.51
Conwy	0.41
Gwynedd	0.31
Neath Port Talbot	0.31
The Vale of Glamorgan	0.31
Monmouthshire	0.31
Torfaen	0.31
Flintshire	0.31
Isle of Anglesey	0.31
Denbighshire	0.31
Ceredigion	0.31
Wrexham	0.31
Blaenau Gwent	0.20
Powys	0.20
Merthyr Tydfil	0.20

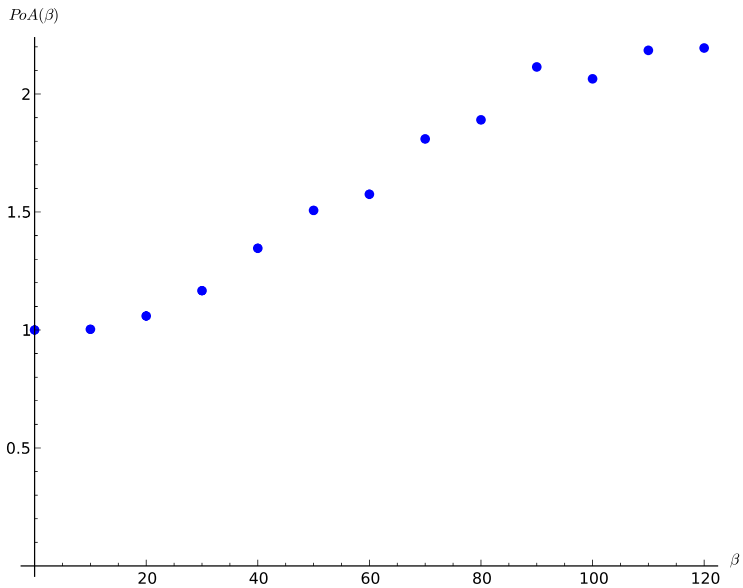

Figure 10: Effect on the Price of Anarchy of Increasing Value of Service

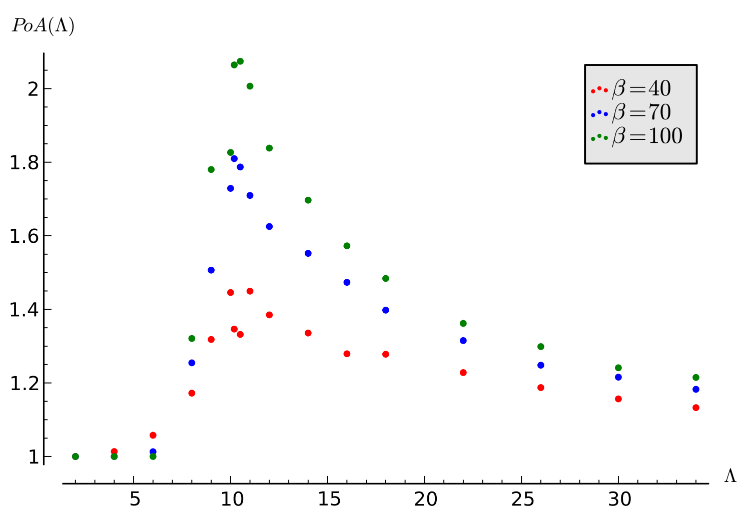

Figure 11: Effect on the Price of Anarchy of Increasing Demand

The 3 following results are proved analytically:

- The price of anarchy increases with worth of service, up to a point.
- In a system with insufficient capacity the price of anarchy is low.
- Choice causes the highest level of inefficiency when the capacity of the system matches the perceived worth of service.

A large scale healthcare case study is given. Two aspects are considered. Firstly we confirm that increasing the perceived worth of knee replacement surgery monotonically increases the price of anarchy to some upper bound. Secondly, the effect of demand is considered. It is once again confirmed that the price of anarchy is bounded. Importantly we note that for the current demand in Wales a high price of anarchy is obtained.

This paper brings with it various avenues for further work. In particular, as in [11] it would be interesting to consider the presented models with elastic demand. Furthermore it would be worth investigating whether the particular problem properties of OPTMP and NASHMP lend themselves to a better solution methodology than the generic one presented here.

The main potential insight gained from the price of anarchy work is at a strategic level where systems can be designed so as to minimise the price of anarchy. As such, new facilities can be introduced in a way as to counteract the negative effects of choice, thus leaving competition to act as a potential vector for improvement.

The authors would like to conclude by re-iterating that this work takes in to account a single dimension of the introduction of choice. No consideration is given to potential benefits of competition. Whether or not individuals are happier with choice is a further subject for debate (one that the authors cannot claim to have sufficient expertise on). For example, in [40], a very negative essay on the subject of choice is given whereas in [41] it is reported that patients' are happier when given a choice.

[1] I. Adler, P. Naor, Social optimation versus selfish optimistion in waiting lines (1969).
URL `http://www.stormingmedia.us/89/8927/0892796.pdf`

[2] C. Bell, S. Stidham, Individual versus Social Optimization in the Allocation of Customers to Alternative Servers, Management Science 29 (7) (1983) 831–839. doi:10.1287/mnsc.29.7.831.
URL http://mansci.journal.informs.org/cgi/doi/10.1287/mnsc.29.7.831

[3] N. Edelson, Congestion tolls under monopoly, The American Economic Review 61 (5) (1971) 873–882.
URL http://www.jstor.org/stable/1813147

[4] N. Knudsen, Individual and social optimization in a multiserver queue with a general cost-benefit structure, Econometrica 40 (3) (1972) 515.
URL http://ideas.repec.org/a/ecm/emetrp/v40y1972i3p515-28.html

[5] I. Luski, On partial equilibrium in a queueing system with two servers, Review of Economic Studies 43 (3) (1976) 519–525.

[6] P. Naor, The regulation of queue size by levying tolls, Econometrica: journal of the Econometric Society 37 (1) (1969) 15–24.
URL http://www.jstor.org/stable/1909200

[7] U. Yechiali, Customers' optimal joining rules for the GI/M/s queue, Management Science 18 (7) (1972) 434–443. doi:10.1287/mnsc.18.7.434.
URL http://www.jstor.org/stable/2629362

[8] W. Leeman, The Reduction of Queues Through the Use of Price, Operations Research 12 (5) (1964) 783–785.

[9] W. Leeman, Comments on preceding note, Operations Research 13 (4) (1965) 680–681.
URL http://www.jstor.org/stable/167860

[10] T. Saaty, Burdens of queuing charges - comments on a letter by Leeman, Operations Research 13 (4) (1965) 679–680.

[11] C. Chau, The price of anarchy for non-atomic congestion games with symmetric cost maps and elastic demands, Operations Research Letters 31 (5) (2003) 327–334. doi:10.1016/S0167-6377(03)00030-0.

URL http://linkinghub.elsevier.com/retrieve/pii/S0167637703000300

[12] R. Cole, Y. Dodis, T. Roughgarden, How much can taxes help selfish routing?, Journal of Computer and System Sciences 72 (3) (2006) 444–467. doi:10.1016/j.jcss.2005.09.010.
URL http://linkinghub.elsevier.com/retrieve/pii/S0022000005001121

[13] J. Correa, a. Schulz, N. Stiermoses, A geometric approach to the price of anarchy in nonatomic congestion games, Games and Economic Behavior 64 (2) (2008) 457–469. doi:10.1016/j.geb.2008.01.001.
URL http://linkinghub.elsevier.com/retrieve/pii/S089982560800033X

[14] G. Gilboa-Freedman, R. Hassin, Price of anarchy in the Markovian single server queue, preprint.
URL http://alexandria.tue.nl/repository/books/640285.pdf

[15] M. Haviv, T. Roughgarden, The price of anarchy in an exponential multi-server, Operations Research Letters 35 (4) (2007) 421–426. doi:10.1016/j.orl.2006.09.005.
URL http://linkinghub.elsevier.com/retrieve/pii/S0167637706001052

[16] J. Musacchio, S. Wu, The price of anarchy in competing differentiated services networks, in: Communication, Control, and Computing, 2008 46th Annual Allerton Conference on, IEEE, 2009, pp. 615–622.
URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4797615

[17] T. Roughgarden, The price of anarchy is independent of the network topology, Journal of Computer and System Sciences 67 (2) (2003) 341–364. doi:10.1016/S0022-0000(03)00044-8.
URL http://linkinghub.elsevier.com/retrieve/pii/S0022000003000448

[18] T. Roughgarden, Selfish routing and the price of anarchy, MIT Press, 2005.
URL http://books.google.com/books?id=K_qheq8dlS0C&pgis=1

[19] T. Roughgarden, E. Tardos, Bounding the inefficiency of equilibria in nonatomic congestion games, Games and Economic Behavior 47 (2) (2004) 389–403. doi:10.1016/j.geb.2003.06.004.
URL http://linkinghub.elsevier.com/retrieve/pii/S089982560300188X

[20] B. Skinner, The Price of Anarchy in Basketball, Journal of Quantitative Analysis in Sports 6 (1) (2010) 3.
URL http://works.bepress.com/context/brian_skinner/article/1004/type/native/viewcontent

[21] E. Koutsoupias, C. Papadimitriou, Worst-case equilibria, in: Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science, 1999, pp. 404–413.
URL http://linkinghub.elsevier.com/retrieve/pii/S1574013709000203

[22] E. Koutsoupias, C. Papadimitriou, Worst-case equilibria, Computer science review 3 (2) (2009) 65–69. doi:10.1016/j.cosrev.2009.04.003.
URL http://linkinghub.elsevier.com/retrieve/pii/S1574013709000203

[23] N. Nisan, Algorithmic game theory, Cambridge Univ Pr, 2007.
URL http://books.google.com/books?hl=en&lr=&id=YCu2alSw0w8C&oi=fnd&pg=PR13&dq=Algorithmic+game+theory&ots=aKGIyWkpB6&sig=y145UQmYVTtweF5-DMbYeoUbdpY

[24] J. Wardrop, Some theoretical aspects of road traffic research, in: Institute of Civil Engineers, 1952, pp. 325–378.
URL http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Some+theoretical+aspects+of+road+traffic+research#3

[25] J. Beckmann, C. McGuire, C. Winstein, Studies in the Economics of Transportation, Tech. rep., Yale University Press, New Haven, Conn. (1956).

[26] A. Pigou, The economics of welfare, Transaction Publishers, 1920.
URL http://books.google.com/books?hl=en&lr=&id=bQUDtmh7uxQC&pgis=1

[27] S. Lippman, S. Stidham, Individual versus social optimization in exponential congestion systems, Operations Research.
URL http://www.jstor.org/stable/169827

[28] A. Dixon, R. Robertson, J. Appleby, P. Burge, N. Devlin, H. Magee, Patient Choice: How Patients Choose and How Providers Provide, Tech. rep. (2010).

[29] N. Koizumi, A. Rothbard, E. Kuno, Distance Matters in Choice of Mental Health Program: Policy Implications for Reducing Racial Disparities in Public Mental Health Care, Administration and Policy in Mental Health and Mental Health Services Research 36 (6) (2009) 424–431. doi:10.1007/s10488-009-0233-z.
URL http://www.springerlink.com/index/b0q70517227h2pn6.pdf

[30] L. Miller, D. May, Patient choice in the NHS: How critical are facilities services in influencing patient choice?, Facilities 24 (9/10) (2006) 354–364. doi:10.1108/02632770610677637.
URL http://www.emeraldinsight.com/journals.htm?articleid=1562411&show=abstract

[31] D. McAllister, Equity and efficiency in public facility location, Public facilities planning (2006) 46.
URL http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Equity+and+efficiency+in+public+facility+location#0

[32] www.sagemath.org, Sage mathematical software V 4.7.1 .

[33] NHS, NHS Constitution Interactive version (2010).
URL http://www.nhs.uk/choiceintheNHS/Rightsandpledges/NHSConstitution/Pages/Overview.aspx

[34] BBC, NHS choice 'not relevant to Wales' (2004).
URL http://news.bbc.co.uk/1/hi/wales/3835135.stm

[35] D. Gross, C. Harris, Fundamentals of Queueing Theory, John Wiley & Sons, 1998.
URL http://www.zentralblatt-math.org/zmath/en/search/?an=0949.60002

[36] H. L. Lee, M. A. Cohen, A Note on the Convexity of Performance Measures of M/M/c Queueing Systems.
URL http://www.jstor.org/pss/3213606

[37] V. A. Knight, J. Williams, I. Reynolds, Modelling Patient Choice in Healthcare Systems : Development and Application of a Discrete Event Simulation with Agent-Based Functionality, Accepted for publication in Journal of Simulation (2010) 1–24.

[38] http://www.palisade.com/evolver/, Palisade Evolver V 5.7.

[39] T. Harks, I. Kleinert, M. Klimm, H. M. Rolf, Computing Network Tolls with Support Constraints (2011).

[40] B. Schwartz, The Paradox of Choice: Why More is Less, HarperCollins, 2005.
URL http://www.amazon.co.uk/Paradox-Choice-Why-More-Less/dp/0060005696

[41] V. Zigante, Subjective Well-being as a Measure of Welfare and Equity: The Case of Choice Policies in Health Care, CESifo Economic Studies 57 (4) (2011) 715–739. doi:10.1093/cesifo/ifr022.
URL http://cesifo.oxfordjournals.org/cgi/content/abstract/ifr022v1