
Richard McDowall
System Dynamics Group,
Department of Geography, University of Bergen, 5020 Bergen, Norway
richard.mcdowall@gmail.com

Abstract
UK prisons have been overcrowded for over 15 years, and the prison population continues to rise at a faster rate than prison capacity. The rising prison population has been attributed to harsher sentencing rather than a real increase in crime. Yet calls for "short sharp shocks" for offender groups, continue to be made without reference to prison capacity. The London rioters are one of these offender groups and are used here as a case study. Though the numbers of rioters facing prison are a small fraction of the prison population we are interested in how the punishment of the rioter’s will affect the prison system as a whole. We analyse the pathways of the rioters using a system dynamics model designed to provide a dynamic hypothesis for the growth of the UK prison population. The paper considers non-linear dynamic feedbacks related to the certainty and severity of punishment, and the problems of rehabilitation implementation caused by capacity constraints.

Keywords: Criminal justice, Prisons, System Dynamics, Crime, Re-offending, Rehabilitation, Recidivism, Punishment, Riots, Penal Populism.

Introduction
England is an increasingly punitive society which views prison as the greatest deterrent to crime (Roberts 2003). According to the English interpretation of the principle of deterrence, increasing the certainty and the harshness of a punishment will increase its effectiveness (Becker 1974). Efforts to deter certain offender groups, especially persistent offenders, have therefore concentrated on increasing the probability of a custodial sentence as a punishment (Stephenson 1999, Mair and Mills 2009). The policy of targeting of offenders in this way is referred to as delivering a "short sharp shock"(Cavadino 2007). The success of the "short sharp shock" penalty is questionable - in recent years recidivism has remained above 60% and the prison population has been rising steadily each year (Figure 1). This paper takes the London riots as a case study to analyse how the application of “short sharp shock” penalties affects the resource limited prison environment.

Figure 1: Prison population and two year recidivism rates 1987 to 2008
The London riots of 2011 lasted for four days, and spread to cities throughout the country. The scale of the damage caused was severe and the unrest caused serious concern over the rule of law. The Prime Minister called for a clear message to be sent to the rioters through harsh sentencing as a deterrent to future outbreaks of civil disobedience (Guardian 2011a). Over 3000 arrests were made. The arrested rioters experienced far higher incarceration rates, and received longer sentences than would be usual for the types of property crimes committed during the rioting (Guardian 2011c).

The association of prison governors has been vocal in its concerns over the prison systems ability to cope with the increased population. The prison system has been put on alert, with plans to increase cell occupancy in the already overcrowded service (Guardian 2011b). Previous research has indicated that high capacity utilization\(^1\) in the prison system can worsen the problem of recidivism (ECPT 2005, MoJ 2008). Operating prisons at overcapacity can form a vicious cycle where increasing recidivism, increases the prison population, worsening overcrowding and increasing recidivism even further (McDowall 2010). Despite highly publicised concern, no equation-based model has been developed to study the impact of this wave of new entrants to the prison system.

Due to the rapid nature of change in the prison system and the number of different hypotheses for these changes there are few equation based analyses of policy decisions made before or during their implementation. The objective of this study is to create a relatively simple equation based simulation model of the impact of the 2011 riots on the prison system. The model consists of two sub-structures, one of the current prison population and another isolating the new riot offender group. This approach delivers new insights into the effects of punishment, both on the riot offender group and the wider existing prison population.

The strength of the modelling approach is the ability to test out different hypothesized causal relationships. Policy analysis within the criminal justice sector requires such highly adaptive tools as the one presented in this model. Criminal justice problems often have multiple hypotheses as causal relationships are often unclear. The ability to both communicate and simulate these hypotheses is essential to the debate within criminal justice To this end we present one simple hypothesis interpreted from one possible reading of the literature. The results can give some information about the possible effect of the riots but more importantly this type of exploratory system dynamics model can be a guide to help improve our understanding of how criminal justice systems function.

Hypothesis

We look at the introduction of a new population into the prison system and explore how the imprisonment rates, sentence lengths and recidivism rates develop over time. In order to link these variables it was necessary to develop a simple hypothesis of the rules ordinarily governing the development of the prison population. The hypothesis must be

\(^1\) The capacity shortages represented by capacity utilization are a proxy for staff shortages, underfunding and poor conditions which could result in a reduction of prison's capacity for rehabilitation.
capable of explaining the links between the type of offender and the certainty of punishment, the links between population size, capacity usage and sentence lengths.

Offenders can be grouped in varying ways. Our hypothesis uses two offender types: first offenders and repeat offenders (or recidivists). As recidivists have been highlighted as the source of a majority of the crime in the country they are specifically targetted for prison sentences (Carter 2003). Recent analysis of sentencing decisions in the UK (Mair et al 2009) has indicated a judicial preference for the use of prison in cases where the accused has been sentenced previously (whether to prison or being diverted to a non custodial sentence). The observed effect of the preference for the use of prison for recidivists has been the slow increase in the overall certainty of imprisonment being used as a punishment (Mair et al. 2009). We interpret this to mean that the more recidivists there are the higher the probability of imprisonment for all offenders.

The increase in imprisonment, will increase the prison population, and the likelihood of capacity utilization problems. Once capacity utilization problems arise the active use of rehabilitation to reduce reoffending\(^2\) becomes a secondary priority to harsh punishment and the secure housing of offenders. Without rehabilitative interventions, the number of recidivists will increase, forcing the prison population to rise. Capacity utilization will be problematic, until new prison places are available to cope with the increased number of offenders.

Capacity utilization problems occur during the inevitable delay between increasing prisoner numbers and the construction of new prison places. If we ignore budget constraints the actual process of building a prison takes on average five years. To cope with these capacity shortages the system has responded with periodic reductions in sentence lengths and the increase of early release of prisoners to make room for new arrivals (Cavadino 2007). The early release of these prisoners can have the counterintuitive effect of increasing the prison population in the long term, as higher numbers of potential recidivists are released at once.

Any increase in recidivism, whether due to sentence lengths, or reduced rehabilitation will increase the pre-existing judicial incentive to be tougher on crime and criminals, placing yet more stress on prison capacity.

\textbf{Causal Loop Hypothesis}

The central feedback is the increase of prison population through recidivism. The more prisoners are released the more recidivists there will be (Loop R1; Figure 2). The more recidivists the more punitive judges will be and the greater the certainty of imprisonment (Loop R2; Figure 2).

\(^2\) Candidates for effective rehabilitative actions could include ensuring housing on release, counciling, and education (Harper and Chitty 2004)
In this model we use the concept of capacity utilization as a proxy for general conditions in prison. So one consequence of high capacity utilization is a reduction in the effectiveness of punishment (Loop R3; Figure 2) as an overcrowded prison system is less efficient in the provision of rehabilitative programs. With a lower focus on helping prisoners correct offending behaviour, recidivism will rise and the prison population will grow and the situation will worsen.

As capacity utilization is a function of both prison population and prison capacity, it will increase when the population rises and decrease when prison capacity rises (Figure 3). Prison capacity increases only occur when prison planners are aware that a rise in the population has taken place and the construction of new prison capacity takes even more time. As a result prison capacity will always be a delayed reaction to prison population growth (Loop C1; Figure 3). Loop C1 counteracts the reinforcing loop R3 and can be considered as a coping strategy for the growth of the prison population. The delays inherent in building new prison capacity severely weakens the effect of loop C1.

A secondary coping strategy is the release prisoners earlier than planned in order to bring the prison population back down (Loop C2; Figure 4). Examples of this relationship in the real world include bulk releases of prisoners at the end of their sentence, and the provisions of the 1991 Criminal Justice Act ensuring that certain prisoners will only ever serve half of the sentence given to them by the presiding judge (Cavadino 2007). Too low capacity utilization could have the opposite effect: with less pressure from prison capacity utilization prisoners can be held for longer. All other things being equal, longer sentence lengths will cause the prison population to grow.
The top down model of issuing a commands and expecting a defined change, contains the implicit assumption that policy makers have control over the operation of the prison system (Elmore 1979). In reality the implementation of a policy can result in very different outcomes to those intended. The modeling approach used to create our tentative structural hypothesis takes a bottom up or backward mapping approach (Elmore 1979, Wheat and Shi 2011). We begin with what is being changed and develop an operational hypothesis for how it can be changed. Our hypothesis is that the true value of a policy can be better understood through creating an understanding of the role of policy resistance in the implementation of a change. We concentrate on what is being changed rather than how we want to change it.

Our interpretation of the backward mapping approach for change in the prison system starts with the central stock of interest (the prison population) and works backwards to find out what can alter that stock. According to the formal modeling methodology of system dynamics only flows can change stocks. So the prison population increases through the addition of inflows for first offenders, and recidivists and decreases through the outflow of offender release (Figure 5).

From there we take a step back and define how the flows change (Figure 6). The offender release rate is a function of the population level and the sentence length. The recidivist entry rate is in fact two flows. The main recidivism flow is dependant upon the current rate of recidivism and the number of potential recidivists. In order to control the number of recidivists we therefore need new structure to determine the number of released offenders. The flows out of the new released offenders stock are determined by
the time it takes to detect recidivism\(^3\), the fractional recidivism rate and the fraction of recidivists who are imprisoned (Figure 6).

![Figure 6: The second stage of the backward mapping approach: What changes the flows?](image)

The few steps taken so far provide us with a highly simplified understanding of the "plumbing" of the problem (Richmond 1994). The model now traces the critical path from the first reception into prison, through to release, and incorporates feedbacks of recidivists re-entering the prison system. The simplicity of the initial model creates an illustrative and workable structure to base our exploration of the dynamics of the main prison population (Figure 6). The next section will expand further upon the definition of these variables.

**The imprisonment fraction**

We take our earlier interpretation of the literature regarding judicial decision making to build a tentative structural hypothesis (Figure 7). We use the total number of cases and the total of all the recidivism flows to establish the percentage of cases that involve recidivists. We then create a judicial perception of recidivism that changes over time to reflect the imperfect nature of the information available to the judiciary. The effect of the perception is termed judicial punitiveness, which we use to calculate the probability of imprisonment on all the recidivism flows.

---

\(^3\) Over a two year period after release it is assumed that a fraction of released offenders will become recidivists. There will be offences committed after this two year period, but the peak occurs during it.
Figure 7: Tentative hypothesis for the imprisonment fraction

The fractional recidivism rate

The fractional recidivism rate is in this model dependent upon the level of capacity utilization and the ex-prisoners perception of the profitability of crime. Though rehabilitation training is often provided by outside providers, there are basic limitations imposed by capacity, for example offenders may be moved between prisons on multiple occasions in order to make best use of available capacity - a problem known as "the churn" (Cavadino 2007). When moved it is highly unlikely that rehabilitation programs (which tend to be regional) will be continued at the new prison. So the higher capacity utilization, the higher the probability of recidivism after release - corresponding to loop R3 in Figure 4. In addition we assume that ex-prisoners interpret general public perceptions of the risk of crime and use these to calculate whether they should commit a crime. The higher the perception of crime the higher the prisoner perception of the potential benefit of crime (Figure 8).

Figure 8: Tentative hypothesized structure for recidivism
The sentence length

Sentence lengths are also a function of capacity utilization (Figure 9). Whenever there are severe capacity shortages the system reacts by shortening the sentence lengths of prisoners - a reduction in the severity of punishment. Decreases in sentence length act to increase the flow out of the stock. This reflects the observed tendency of releasing prisoners earlier due to capacity utilization pressure (Cavadino 2007).

The unintended consequence of early release is that there are more people capable of becoming recidivists, so the prison population will grow after a short delay (Loop C2; Figure 4 and Figure 7). As the population grows there is more overcrowding and sentence lengths have to be reduced further. One important addition is that the length of sentence served by offenders is also affected by the social perception of crime - which we derive from the current crime rate. The effect of this is to create the rise in sentence lengths that we have seen over the last two decades.

Prison Capacity

The final step in the backward mapping of this system is to establish what lies behind capacity utilization; namely population and capacity (Loop C1; Figure 9). We are already aware of the structures behind population growth so we can concentrate on capacity. As prison projections are run according to a 5 year forecast we have added a forecast structure that populates the system with the required number of cells.

The prison capacity structure is rooted in John Sterman’s (2000) inventory model. Figure 10 is the simplest stock and flow explanation of the prison system. Two stocks are used: Prisons Under Construction, and Prison Cells. Prison cells have to be ordered before construction can begin. It takes a prescribed amount of time for cells to be constructed. Once in service prison cells depreciate, according to their average lifetime.
The full structure (Figure 11) details how the order rate is decided upon. The indicated order rate aims to replace cells lost through depreciation, and to close the gap between the forecast number of prison cells and the forecast number of prisoners. Within the formulation of the indicated order rate provision is also made to allow for the cells which are currently under construction.

The prison population forecasts span for five years. Trends in population growth are used to make a five year forecast for the prison population:

\[ \text{Population} \times \text{PRISONER\_TOTAL} \times (1 + \text{perceived pop growth percentage}/100)^5 \]

Trends in the growth of cells are also used to make a forecast of the number of cells that will be required in five years time:

\[ \text{Prison} \_\text{Cells} \times (1 + \text{perceived cell growth percentage}/100)^5 \]

The gap between these two forecasts is used to create the forecast number of cells that will be ordered. The final influence on the order rate is the percentage of the gap that is funded. As we have not explicitly modeled the economy we have simulated swings and shifts in economic performance and the political prioritization of the prison system with a 10 year sinwave.

**Diversion from custody**
For the sake of completeness we have added in a structure for diversion from custody (or non custodial punishments). It is essentially the mirror image of the structure seen in Figure 6. Recidivism rates for diverted populations have been given a constant value in this model\(^4\).

The Rioter Sub-Structure

The rioter sub-structure is an exact copy of the fundamental structure used for the rest of the prison system (see Appendix). The sub-structures interact through the judicial perception of the recidivist fraction which we use to describe judicial punitiveness (Figure 10). The judicial perception is a function of all the recidivism flows and the total number of rioter and non-riot cases. The total population of the two substructures is combined to define the desired prison capacity. As rioter sentences have been proportionally longer than other sentences a simple multiplier has been used increase their sentencing in the period between August 2011 and August 2012 - by which time we assume everyone will have received their initial sentence for the riot offences.\(^5\)

---

\(^4\) A later version of the model should build in a dynamic recidivism capacity component for diversions. This has not been included here due to the lack of sufficient data. The addition of this sector will have important policy implications.

\(^5\) One could also make the case that the sentence length increase should continue until 2013 to ensure the initial rioters will serve the entire of their sentence. This was not done here due to the numbers of rioters appealing their sentences.
Analysis

The main population structure outlined in the previous section (see Appendix for entire structure) is capable of reproducing something very close to the reference mode behaviour for the main prison population (Figure 14). In addition historic trends of parameters such as the sentence length, recidivism and the number of prison cells were reproduced with a reasonable degree of accuracy. Recreating the capacity utilization trend proved to be more problematic due to slight deviations in the reproduction of the population and capacity. This problem was also encountered in an earlier study (McDowall 2010). To ensure that the results from our model were still valid we independently tested the model with historic inputs for capacity, then population and then for capacity utilization.

![Figure 14: Simulation of the Historic Behaviour of the Prison Population between 1987 and 2007](image)

Equilibrium testing

To test the effect of the rioter arrests on the prison population we set the model to run in equilibrium and ran the system from time 2000 to time 2040. We used the rioters to shock the system at 2011, and included the rioters imprisoned whilst awaiting trial. In order to achieve equilibrium recidivism rates and sentence lengths were reduced to 0.5.

The rioters

One of the strengths of the simulation results was its ability to display the true effect of recidivism on the rioter population. The original average sentence length for the riots had been 11 months. In the simulation run there were still 15 rioters in prison 20 years after the riots and the last rioter does not exit the system until 25 years after the riots (Figure 15). The reason for this is the level of recidivism in the system. As real fractional recidivism rates are nearer 0.65 we would expect a higher number of rioters to become recidivists after release.

![Figure 15: The long term outlook for the rioter population](image)
The ordinary prison population

When we examine the prison population with and without the rioter population we can see that though the total population rises the number of prisoners in the ordinary population actually drops in 2011 (Figure 16). The total population falls sharply between 2012 and 2013 (as the majority of rioters are released) before rising once more. The second rise in the population is almost entirely due to an increase in the ordinary prison population.

Factors behind the rise

According to our hypothesis the rise in the number of prisoners should cause capacity utilization to rise (Figure 17 and 18). Capacity utilization falls when prison capacity responds to the input of new prisoners. As the input of rioters is high but very short it results in a mismatch between capacity and population – and a situation of under capacity operations.

Any increase capacity utilization should force sentence lengths down and recidivism up. Sentence length changes happen very quickly (Figure 19). The change in recidivism is delayed as prisoners do not show the effects of their sentence until after their release.
The effect of the sentence reduction is to ease the population pressure (explaining the drop seen in Figure 16). Once capacity utilization drops sentence lengths climb up again. The secondary effect of reduced sentence lengths is an increase the stock of released offenders. The combination of an increased number of released prisoners and a rise in recidivism help to cause the increase in the prison population observed in Figure 16.

Cutting Loops

Cutting out the effect of recidivism (loop C2 in Figure 4 and Figure 9) has only a slight effect on the prison population (Figure 20).

Cutting the sentencing loop results in a smoothing out of the prison population development and a lower overall prison population. If sentencing is held as constant recidivism takes a lower value(Figure 21).

Sensitivity to budget forecast changes

The system behaviour is very different if we assume that the riot population is not included in the forecast for new prison cells - as they are only going to be a short term
on time addition (Figure 22). The total prison population spikes after 2011 but does not rise to levels reached if the extra prison population was included in the forecast cell demand. Prison capacity actually drops as after the spike ends in 2011, the prison population is actually lower than before. As a result capacity utilization is higher than it would have been if the rioters had been included in the forecast.

The above graphs (Figure 22) would suggest that including the riot arrest in the forecast for the prison population may be unadvisable. However when we look at the number of released offenders, the length of sentence and the rate of recidivism there is a very different picture (Figure 23). Due to the lower level of resources available in a prison system with high capacity utilization, sentence lengths drop, and recidivism rises. As a result there are far more offenders being released.

Again, due to the low level of resources fractional recidivism worsens and we see higher numbers of recidivists being sent to prison (Figure 24).
Budget sensitivity

Sensitivity runs on the level of funding assigned to the riot sentences showed that lower levels of funding could be effective in the reduction of recidivism (Figure 25).

Multiple shocks

Finally we ask the question: what happens if there is another riot event in 10 years. If there is an allowance made in the forecast for new cells, recidivism drops further. If no allowance is made recidivism climbs higher (Figure 26).

Discussion

This initial paper and modelling effort has been a combination of tasks. Firstly the causal loop diagramming method has been used to show a hypothesized set of feedback loops that govern the behaviour of the prison population over time. The formal modelling methodology of system dynamics has then been used to show how the
behaviour generated by these feedback loops would affect the structure of the system. In our analysis section we explored the role of forecasting policies.

The simulation results show punishments designed as "short sharp shocks" for certain offender groups actually result in long-term unintended consequences to the operation of the system as a whole. Overcrowding pressures initially force the early release of prisoners from the main prison population. But in the long term any reduction in capacity utilization reduces the incentive for shorter sentences and the prison population will rise. This is in line with observations that efforts to encourage the use of non-custodial penalties are often short-lived⁶ (Cavadino 2007). For the prison system to be able to run efficiently policy makers and judges need to be making decisions with reference to the goals of the prison system and its capacity to deliver.

The model results show that sentence lengths and recidivism are dynamic, and at least partially endogenously driven. In turn it is these two variables that drive both population and capacity development. From our equilibrium tests it was discovered that sentence length adjustments due to capacity constraints were the main lever of control in the system as we envision it. Recidivism plays a role in the level of the prison population but surprisingly overcrowding played a much lesser role in the development of recidivism than expected. Our research suggests that factors outside of prison could play a larger role in the reduction of recidivism than prison itself - this could mean that continued follow up efforts after release could have a greater rehabilitative effect than reducing overcrowding. A future study should be conducted focusing solely on recidivism, taking the analysis to the effects from wider society, perhaps working with a single prison.

The final part of our analysis centered on whether the rioters should be included in prison population forecasts and budgets for prison building. Currently there are large scale budget cuts throughout the public sector. Little extra funding is available. From our equilibrium testing we can see that unless funding is made available fractional recidivism will rise, as will the number of recidivists. A rise in recidivism rates would suggest that even if the short sharp shock policy successfully deters further rioting, it fails to achieve the objective of reducing crime. If the policy fails to deter further rioting and there is no funding made available we will have significantly higher recidivism rates for a longer period of time.

The structure used in this paper was based on one possible set of causal relationships developed from previous study. The variables and links used were chosen for the close fit to the widest stream of literature whilst maintaining a simple and understandable structure. Other relationships can and do exist. The strength of the modelling approach taken here is its simplicity and adaptability. Different hypotheses can quickly be combined with the fundamental structure shown in Figures 4 and 5.

⁶ The current "go to" policy in this area is to suggest the provision of more non-custodial penalties, yet the capacity to deliver non-custodial penalties is rarely referred to. It may suffer from the same capacity utilization complications as the prison system, and has been partially explored in McDowall (2010).
Ultimately this modeling approach could play a central role as a decision support tool for policy makers and prison governors alike, providing a much needed method to explore new policy before it has been implemented.
Reference List


Appendix A:

Figure 27: The main prison population

Figure 28: The rioter population