History Matters: Modelling Path Dependence on a Spreadsheet

Chris Hand
School of Marketing, Kingston University

Abstract

This paper explores methods of simulating Polya Urn Models using a spreadsheet. The Polya Urn model provides a statistical model of path dependent behaviour where the final equilibrium is determined by random chance or historical accident.

Introduction

In both network markets and information-based markets the attractiveness of a given product or product standard increases with the number of users of that product or product standard, in other words there is a positive externality. For example, either Betamax or VHS could have been the dominant standard for Video Cassette Recorders (VCRs). The more people who chose VHS, the more films would be released on VHS video (as there is a greater potential audience); this increased choice of films makes VHS VCRs more attractive, and so on. Such a process is variously described as displaying increasing returns or path dependency. Under these circumstances the success or failure of products of near equal quality may depend on chance; it is impossible to predict which alternative will win. A consequence of such a process is that a sub-optimal outcome may arise with an inferior technology becoming locked in. Such stochastic processes are called non-ergodic processes and do not converge to a fixed-point distribution of outcomes. Which outcome is arrived at depends on earlier events which may be dominated by chance or by systematic factors; in other words, history matters. Examples of this include the prevalence of light-water nuclear reactors in the United States when gas-cooled reactors are arguably superior and

the dominance of the petrol engine over steam-driven alternatives (Arthur, 1989; Cowan, 1990), the QWERTY keyboard (David, 1985), the narrow gauge of British railways (Kindleberger, 1983) and the prevalence of DOS over CP/M and Apple Macintosh's PC operating systems (Arthur, 1996).

Such path dependent processes can be described mathematically using what are known as urn models in probability theory. The model outlined here is known as Polya's urn model. Usually this model is described using an imaginary urn and two differently coloured sets of balls. This convention is employed here, but one could equally well describe the model in terms of, for example, VCRs (VHS versus Betamax) or industrial location (e.g. software companies clustering in Silicon Valley or on Route 128 circling Boston).

The Polya urn model

Imagine an urn containing one red ball and one blue ball. One ball is drawn at random and then replaced with a ball of the same colour. The probability of drawing a blue ball is given by the number of blue balls in the urn divided by the total number of balls in the urn (similarly, the probability of drawing a red ball is given by the number of red balls in the urn divided by the total numbers of balls in the urn). If the first ball drawn was blue, that blue ball is replaced in the urn and a second blue ball is also placed in the urn. The probability that the next ball drawn is blue is then 2/3 whilst the probability that the next ball is drawn is red is 1/3. In general, if k of the first r balls drawn were blue, the urn at the time of draw $r + 1$ would contain $k + 1$ blue and $r - k + 1$ red balls. The probability that the next ball drawn is blue is given by $(k + 1)/(r + 2)$ (Ross, 1997). This is known as the Polya urn model (named after the mathematician George Polya). This urn model has an intriguing property: all possible proportions of blue balls in the urn are equally likely to occur. Which equilibrium the system settles down to depends on chance and in particular on the early draws made. In other words, the 'selection' of an equilibrium displays path dependency. Such models have recently been used to simulate firm growth rates (Bottazzi and Secchi, 2003) and the formation of social networks (Skyrms and Pemantle, 2000).

The model can of course be extended to include more than two choices, which yields the probability distribution known as the Bose-Einstein distribution (Ross, 1997). Alternatively, the probability of drawing a ball of colour c may vary non-linearly with the proportion of c coloured balls (for a discussion of such models see Arthur, Ermoliev and Kaniovski, 1987). In the latter case, the number of fixed-point outcomes is limited, depending on the precise relationship between the proportion of balls of a given colour and the probability of drawing a ball of that colour.

However, it should be noted that the Polya urn model can only be seen as a simplified model of competition between products/standards. Specifically it does not incorporate strategic action by either firm such as pricing to gain market share, predatory pricing or promotional activity. Rather, it assumes that future choices depend only on past choices (analogous to the information cascades theory suggested by Bikhchandani, Hirshleifer and Welch, 1992). Sterman (2000) suggests a more realistic simulation model of competition between standards which includes the attractiveness of each product, the installed base of each product, market share, network/compatibility effects and total demand in a number of feedback loops. Windrum (2004) uses a development of the urn model, a coupled urn model, to simulate the standards battle between Microsoft and Netscape.

Matthews (2001) discusses the use of path dependent models in teaching statistics for economists, illustrating that under path dependence, the law of large numbers breaks down. Matthews (2001) also provides MAPLE code to simulate random draws and draws with limited and substantial positive feedback. However, a package such as MAPLE is not necessary to run simulated draws; the model can also be implemented on a spreadsheet. Arguably, there are advantages to introducing students to an unfamiliar concept (a model with unpredictable multiple equilibria) using a familiar software package. A second, more general, advantage of spreadsheet models is that the students can examine all the formulae used which can help remove any notion of the model operating as a 'black box' (for a range of spreadsheet models of business problems, see Barlow, 2005; for applications in economics, see Judge, 2000).

The spreadsheet model

The Polya urn model can be implemented on a spreadsheet using a column of random numbers and the IF command. Using the spreadsheet's random number generator create a column of 100 (or more) random numbers which range between 0 and 1 in column C. Copy the random numbers and then paste the values back into the same column using Paste Special (this will prevent the random numbers changing every time the spreadsheet is changed). The total number of balls in the urn is also needed. We begin with one ball of each colour and one ball is added after each draw. Hence in the first period the urn contains 2 balls, 3 in the second period, 4 in the third period and so on. Enter this series of numbers in column D. Columns A and B will contain the number of blue and red balls in the urn after each draw. In the first cell of each of these two columns enter a 1. The draws from the urn

are simulated using a series of random numbers and the proportion of blue balls in the urn. The probability of drawing a blue ball depends on the proportion of blue balls in the urn; the more blue balls there are in the urn, the greater the probability that the next ball drawn will be blue. Assume there are 2 blue balls and 1 red ball in the urn. The probability that the next ball drawn is blue is 2/3. The draws can be simulated using random numbers. If the random number in the second row is less than the proportion of blue balls (2/3), a blue ball is drawn and hence the number of blue balls increases by one; if the random number is greater than the proportion of blue balls, a red ball is drawn. If the number of blue balls is entered in column A, the number of red balls in column B, the random numbers in column C and the total number of balls in column D then the draws can be simulated using the following formula in cell A3:

=IF(C3<(A2/D2),(A2+1),(A2+0))

Cell C3 contains a random number while the expression A2/D2 contains the proportion of blue balls in the urn, which like the random numbers can only take values between 0 and 1. This formula can then be copied down the screen to simulate further draws. Note that as the value of An/Dn (n = 2…N) increases it becomes more likely that the next ball drawn will be blue. For example when the first draw is made A2/D2 = 0.5 which is equivalent to saying that a blue ball will be drawn with probability of 0.5. Assuming a blue ball was drawn in the first draw, A3/D3 = 0.66, so the probability that a blue ball is drawn in the second draw equals 2/3. The number of red balls equals the total number minus the number of blue balls, so in cell B3 enter the formula D3-A3. The screenshot below shows the initial set up of the model.

The formulae in columns A and B can then be copied down the screen.

If the draw is simulated several times with a different set of random numbers each time it will be seen that the proportions settle down to a fixed value over time, but what that proportion will be or even whether there will be more blue than red balls or vice versa cannot be predicted. Figure 2 shows the histories of six simulated draws from the spreadsheet model.

As was noted above, *a priori*, the value the simulated series will converge to is unknown, all that can be said is that the series will converge to a stable value. As a way of

Figure 1. Linear Polya urn model spreadsheet

demonstrating this emergent order in these models, Matthews (2001) suggests asking students each to run the simulation once on their own as a classroom exercise. They will agree that the system settles down to a stable value, but will not agree on what that stable value is. (This uniform distribution of outcomes only occurs when one ball is added at a time and the urn begins with one ball of each colour; for a comprehensive discussion of urn models, see Johnson and Kotz, 1977.)

The spreadsheet model could also be used to investigate the effects of first mover advantage or late entry into a market characterised by increasing returns. To make discussion of this clearer, assume that the series in columns A and B are cumulative sales made by firms X and Y. Assume first mover advantage confers an advantage on firm X allowing it to sell more of its product in the first period, this then increases the probability that the next sale will be made by firm X. Instead of starting the simulation with each firm having 1/2 of the market (or 1 sale each), the simulation could be run with firm X holding 2/3 of the market (X begins with 2 sales while Y begins with 1). Under path dependence, such an initial advantage can be quickly eroded, although the greater the initial advantage given to X, the less likely that Y will capture much of the market. The model outlined above is limited to duopoly situations. To introduce more competitors (more colours) the spreadsheet model must be extended.

The extended linear model

As with the Polya urn model, in the extended model (which has more than two colours/competitors) all possible outcomes are equally likely. The probability distribution of the outcomes is sometimes called the Bose–Einstein distribution. This distribution is more familiar to particle physicists than economists, as it describes the distributions of particles known as bosons over energy states. However, as Hill and Woodroofe (1975) have shown, it also has an application in the social sciences as the well-known Zipf law (rank–frequency relationship) can arise from a system following Bose–Einstein dynamics. As De Vany and Walls (1996) note, the Bose–Einstein distribution is uniform. Usually, a uniform distribution implies that, given n colours of balls, the probability that any one colour is drawn is $1/n$. However, the Bose–Einstein distribution arises from sequential draws from an urn. For each draw, the probability that a given colour is drawn depends on the proportion of previous draws of that colour (as with the Polya urn model).

Imagine an urn this time containing one of each of m types of balls. One ball is drawn at random and replaced with a ball of the same type. If in the first n drawings from the urn, x_j type j balls were drawn, the probability of drawing a j type ball in the $n + 1$ draw is equal to $(x_j + 1)/(m + n)$ (that is the number of balls of type j divided by the total number of balls in the urn). A more concrete example may be films competing with each other at the box office; the more one film is chosen, the more people there are to recommend it to others, so in turn a greater number of people are likely to choose that particular film (a model suggested by De Vany and Walls, 1996). Ross

(1997) provides a more detailed explanation and derivation of this model.

Again, a column of random numbers is used to simulate the draws from the urn. We begin with 5 balls in the urn, one each of red, orange, yellow, green and blue. The probability that any one is drawn is 0.2 (1/5). If the first random number is less than 0.2 a red ball is drawn, if it is less than 0.4 but greater than 0.2 an orange ball is drawn, if it is less than 0.6 but greater than 0.4 a yellow ball is drawn, if it is less than 0.8 but greater than 0.6 a green ball is drawn, finally, if it is less than 1 but greater than 0.8 a blue ball is drawn. This is entered in the spreadsheet as outlined below.

In the first row of columns A, B, C, D and E type the names of the 5 colours/firms. Label column F 'random' and column G 'total'. Using the random number generator, create a list of random numbers in column F. In column G enter the total number of balls in the urn at each draw. This can be done by simply typing in the number or by summing the values in each row of columns A to E. It is a good idea to use both of these methods; if they do not produce the same value, there is an error in one of the formulae. We begin with one ball of each colour so enter a 1 in cells A2, B2, C2, D2 and E2. In cells A3, B3, C3, D3 and E3 enter the formulae shown in Table 1.

Taking the formula in A3 as an example, F3 is the random number and A2/G2 is the proportion of red balls in the urn prior to the first draw being made (i.e. 1/5). If the random number is less than A2/G2 a red ball is drawn and replaced and an extra red ball is added to the urn (A2+1), if the random number is greater than A2/G2 the ball drawn is not red so the number of red balls does not change. The formula in cell B3 uses exactly the same logic. If the random number is less than A2/G2 a red ball is drawn so the number of orange balls does not change. However, if the random number is less than the proportion of red balls plus the proportion of orange balls, an orange ball is drawn and replaced and an extra orange ball is added to the urn. The formulae in cells C3 and D3 are constructed in exactly the same way. A somewhat simpler formula is entered in cell E3. If the number of red, orange, yellow and green balls has not changed, the ball drawn must be blue. These formulae can then be copied down the screen to simulate further draws. The screenshot below (see Figure 3) shows the model set up in Excel.

Note the two columns headed 'total'. One contains the sum of the values in columns A to E, the other the number that should be obtained (5 before the first draw is made, 6 after the first draw and so on). As these columns agree, the formulae have been entered correctly.

The extended spreadsheet model could be used to illustrate the competition between different types of nuclear reactor investigated by Cowan (1990). Cowan shows that there were three main types of reactor: Gas Graphite, Light Water and Heavy Water. Although the first nuclear reactor to be used to generate electricity was a Gas Graphite reactor, more was known about Light Water reactors as a result of the US Navy's research into nuclear power for submarines. This can be modelled on the

Table 1 Extended urn model formulae

Colour	Cell	Formula
Red	A3	=IF(F3<(A2/G2),A2+1,A2)
Orange	B3	=IF(F3<(A2/G2),B2,IF(F3<((A2/G2)+(B2/G2)),B2+1,B2))
Yellow	C3	=IF(F3<((A2/G2)+(B2/G2)),C2,IF(F3<((A2/G2)+(B2/G2)+(C2/G2)),C2+1,C2))
Green	D3	=IF(F3<((A2/G2)+(B2/G2)+(C2/G2)),D2,IF(F3<((A2/G2)+(B2/G2)+(C2/G2)+(D2/G2)),D2+1,D2))
Blue	E3	=IF((A2+B2+C2+D2)=(A3+B3+C3+D3),E2+1,E2)

spreadsheet as a three-ball urn model. Given the stock of knowledge regarding Light Water reactors, Light Water starts with an advantage over the other two types so set the initial value for light water to 4 and the initial values for the other two types to 1. Note that the advantage given to Light Water is arbitrary; the point here is to illustrate the effect of prior advantage. Giving Light Water an advantage of 3 over the others makes it likely, but does not guarantee that it will become the dominant technology. This model is shown in Figure 4.

Figure 3. Five-ball linear probability urn model screenshot

Figure 4. Nuclear reactors example

At the end of the 31-year period covered by Cowan's study, Light Water accounted for approximately 70% of all reactors. Running 30 simulated draws ten times yielded the shares of the market shown in Table 2.

In this example, Light Water's initial advantage allows it to take the lion's share of the market, although its dominance is not guaranteed. Furthermore, it is only the initial advantage that allows light water to dominate the market; the relative costs and efficiency are not part of the model.

The non-linear model

In the urn models described above, the probability of drawing a blue ball is a linear function of the proportion of blue balls in the urn and every possible outcome is an equilibrium. If the relationship between the probability of drawing a red ball and the proportion of red balls in the urn were non-linear, this would change; the model would have a limited number of stable and unstable equilibria and would rapidly lock in to one of these equilibria. Figure 5 shows this more clearly.

The continuous line shows a linear relationship between the proportion of red balls in the urn and the probability that the next ball drawn will be red. Every point on this line is a possible equilibrium. The broken line shows a non-linear relationship with three possible equilibrium points labelled A, B and C. At points A and B are stable equilibria; if the urn contains only blue balls (point A) or only red balls (point B) it will remain in that state. If the proportion of red balls is one-half (point C) the probability of drawing a red ball is 50% so, on average, the proportion of red balls will remain constant. If another red ball is added, however, it dramatically increases the probability that a red ball will be drawn next time, moving the system towards point B. Conversely, if a blue ball is added it disproportionately increases the probability that the next ball drawn will be blue, moving the system towards point A. Hence point C is an unstable equilibrium. Under this relationship, the urn will quickly 'lock-in' to one extreme or the other.

The spreadsheet model can be easily adapted to incorporate a nonlinear probability relationship. The dotted curve in Figure 5 resembles the normal cumulative distribution function. In the linear probability model, draws were simulated by comparing the proportion of blue balls in the urn with a random number. If the proportion was greater than the random number, a blue
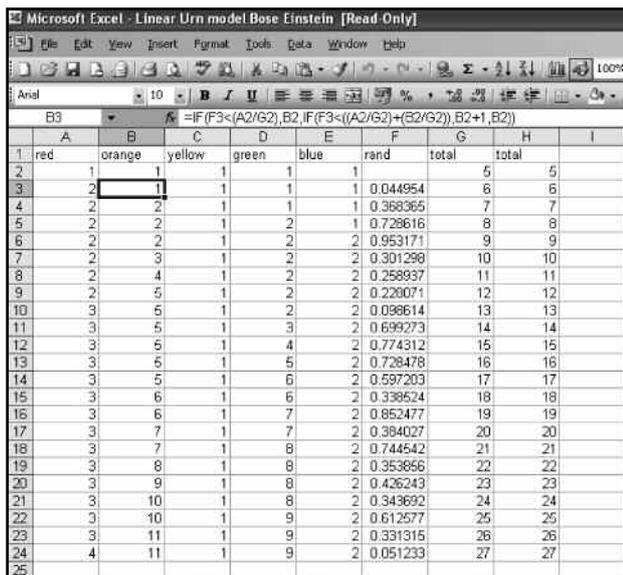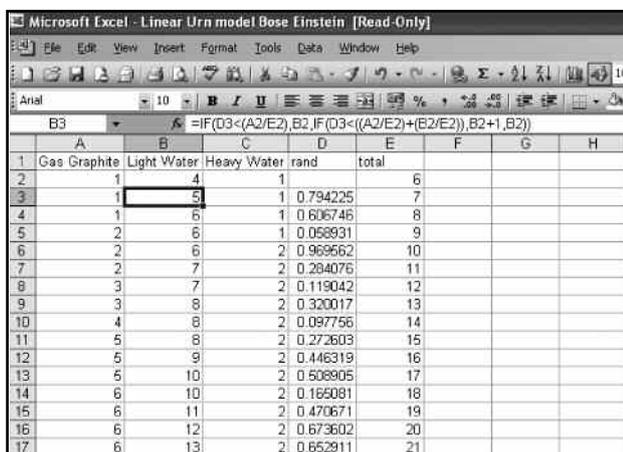
Table 2. Reactor simulations

| Simulation	Market share (%)		
	Gas Graphite	Light Water	Heavy Water
1	19.4	72.2	8.3
2	33.3	61.1	5.6
3	5.6	66.7	27.8
4	11.1	80.6	8.3
5	2.8	72.2	25.0
6	8.3	47.2	44.4
7	19.4	77.8	2.8
8	19.4	58.3	22.2
9	33.3	27.8	38.9
10	16.7	80.6	2.8
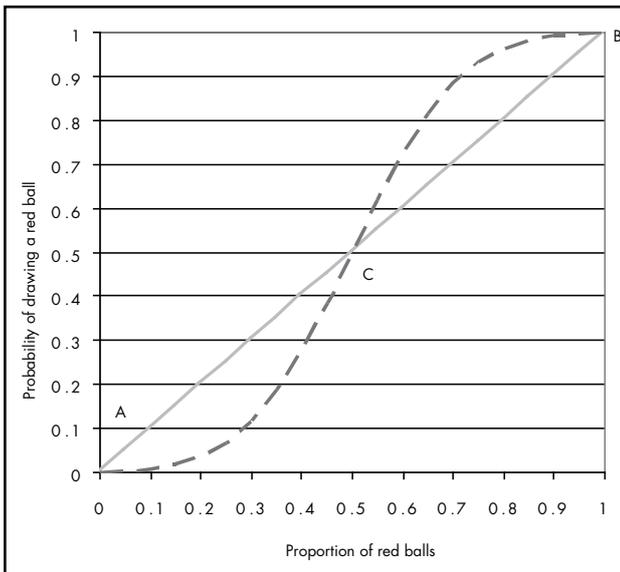
Figure 5. Linear and non-linear probabilities
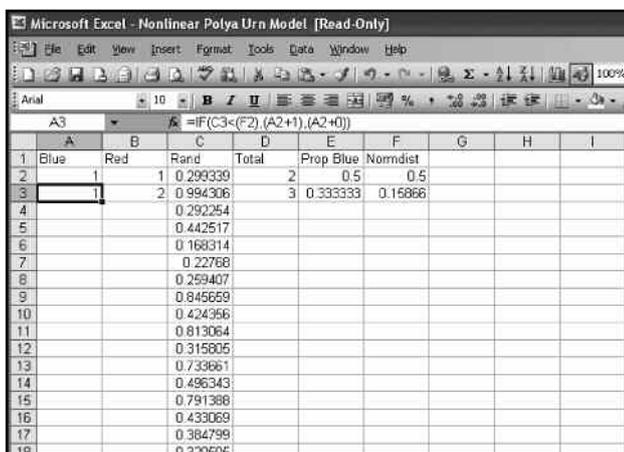
Figure 6. Non-linear probability Polya urn model

ball was drawn; if the proportion was smaller, a red ball was drawn. In the linear model, the probability of drawing a blue ball next time depended on the proportion of blue balls in the urn, represented in Figure 5 by the solid line. In the nonlinear case, the random number is compared to the value of the broken line. This is given by the value returned by Excel's '=NORMDIST' function for the proportion of blue balls (0.5 initially) using a mean of 0.5 and standard deviation of 1/6. (The values were selected to ensure the formula would only return values between 0 and 1 and would trace out a curve like that shown in Figure 5.) To turn a linear urn model into a non-linear urn model, we need to change the formula in cell A3 to read '= IF (C3<(F2), (A2+1),(A2+0)'. The formula '=NORMDIST(E2,0.5,1/6,TRUE)' should be entered in cell F2. Both of these can then be copied down as before. Figure 6 shows a screenshot of the nonlinear urn model with the revised formula in column A.

Discussion and conclusion

The notions of increasing returns and path dependence are now firmly established in economics. The spreadsheet models described here are intended to provide an intuitive introduction to a mathematical model of path dependence. Although the Polya urn model itself is relatively accessible and can be described using words alone, implementing it in a spreadsheet environment provides a clear demonstration of the implications of the model. To economics students, familiar with competitive equilibrium models where the equilibrium is unique, the implications of a model displaying multiple equilibria, the distribution of which is a random variable will be, to say the least, counter-intuitive. In such a situation, rather than being presented with an abstract theoretical model, a spreadsheet-based simulation allows students to explore how the model works and see its implications in a familiar computing environment. Although it has been described here in abstract terms, the model can be used to simulate

competition between standards and information and experience goods.

The main conclusion which can be drawn from all these models however, is that in markets which are characterised by increasing returns, network externalities or lock-in, which of the possible equilibria the system arrives at is a matter of chance. Or, to put it another way: history matters.

Acknowledgements

References

Arthur, W. B., Ermoliev, Y. M. and Kaniovski, Y. M. (1987) 'Path-dependent processes and the emergence of macro-structure', *European Journal of Operational Research*, 30, 294–303.

Arthur, W. B., (1989) 'Competing Technologies, Increasing Returns, and Lock-In by Historical Events', *Economic Journal*, 99(394), 116-31.

Arthur, W. B. (1996) 'Increasing returns and the new world of business', *Harvard Business Review*, July–August, 100–109.

Barlow, J. (2005) *Excel Models for Business and Operations Management*, 2nd edn., Chichester: John Wiley and Sons.

Bikhchandani, S., Hirshleifer, J. and Welch, I. (1992) 'A theory of fads, fashion, custom and cultural change as information cascades', *Journal of Political Economy,* 100(5), 992–1026.

Bottazzi, G. and Secchi, A. (2003) 'Why are distributions of firm growth rates tent-shaped?', *Economics Letters*, 80, 415–420.

Cowan, R. (1990) 'Nuclear power reactors: A study in technological lock-in', *Journal of Economic History*, 50(3), 541–567.

David, P. (1985) 'Clio and the economics of QWERTY', *American Economic Review Proceedings*, 75, 332–337.

De Vany, A. and Walls, W. (1996) 'Bose–Einstein dynamics and adaptive contracting in the motion picture industry', *Economic Journal*, 439(106), 1493–1514.

Hill, B. and Woodroofe, M. (1975) 'Stronger forms of Zipf's law', *Journal of the American Statistical Association*, 70(349), 212–219.

Johnson, N. and Kotz, S. (1977) *Urn Models and their Application*, New York: John Wiley.

Judge, G. (2000) *Computing Skills for Economists*, Chichester: John Wiley.

Kindleberger, C. (1983) 'Standards as public, collective and private goods', *Kyklos*, 36, 377–396.

Matthews, P.H. (2001) 'Positive feedback and path dependence using the law of large numbers', *Journal of Economic Education*, 32(2), 124–136.

Ross, S.M. (1997) *Introduction to Probability Models*, 6th edn., London: Academic Press.

Skyrms B. and Pemantle R., (2000), 'A dynamic model of social network formation', *Proceedings of National Academy of Science*, 97, 9340–9346.

Sterman, J. D. (2000) *Business Dynamics: Systems thinking and modelling for a complex world*, Boston: McGraw-Hill.

Windrum, P. (2004) 'Leveraging technological externalities in complex technologies: Microsoft's exploitation of standards in the browser wars', *Research Policy*, 33, 385–394.

Contact details

Dr Chris Hand
Senior Researcher
School of Marketing
Kingston University
E-mail: c.hand@kingston.ac.uk