Geographical modelling of happiness and well-being using population surveys and remote sensing data

Dimitris Ballas[1] and Steffen Fritz[2]

1 Department of Geography, University of Sheffield, England; e-mail: d.ballas@sheffield.ac.uk.
2 Forestry Program, International Institute for Applied Systems Analysis, Austria; e-mail: fritz@iiasa.ac.at

Abstract: In recent years there have been numerous attempts to define and measure happiness in various contexts and pertaining to a wide range of disciplines, ranging from neuroscience and psychology to philosophy, economics and social policy. However, it is interesting to note that although there are now a considerable number of happiness studies from various perspectives, there is a paucity of work by geographers in this area. This paper presents work in progress that aims at building geographical models of happiness capable of providing information on the different degrees of happiness and well-being attained by people in different regions and localities, under alternative scenarios and happiness definitions. In particular, it will show how a spatial microsimulation model of subjective happiness and well-being (Ballas, 2007; see http://www.uptap.net/project10.html) can be enhanced with the use of remote sensing very high resolution data in order to determine the extent to which geographical proximity to entities pertaining to the natural and built environment (e.g. living near a river, lake, by the sea, a park, a motorway etc.) may be affecting subjective well-being and happiness, in addition to socio-economic, demographic and contextual variables. The proposed model will be based upon the analysis of secondary socio-economic data, such as the household panel surveys and the censuses of population, as well as satellite very high resolution imagery, which will be combined with the use of GIS-based spatial modelling.

1. Introduction

Human perceptions of happiness vary and depend on a wide range of factors. Efforts to define and understand happiness date back to the work of Socrates, Plato and Aristotle. In particular, Aristotle, in his work *Nicomachean Ethics*, attempted to give an answer to the question: *what is the good life?* (Lear, 1988; Ross, 1923). For Aristotle, happiness is the highest good achieved by human action. Aristotle suggested that the attainment of happiness involves the satisfaction of the human desires that are necessary to live a full and rich life (Lear, 1998). However, Aristotle believed that the question of what is a full and rich life cannot be answered for an individual in abstraction from the society in which he lives (Lear, 1998). The meaning of happiness varies through space and time. There have been numerous attempts to understand and define happiness since the original work of Aristotle. Attempting to determine the factors that make individuals happy has long represented a great research challenge. There have been numerous studies of happiness and well-being issues across academic disciplines and from different perspectives. Sumner (1996) pointed out that happiness is subjective and that no objective theory about the ordinary concept of happiness has the slightest plausibility. Nevertheless, there have been several researchers who suggested that happiness can be measured (e.g. see Griffin, 1986; Broome, 1999) and there has been on-going debate (Little, 1957; Sen, 1982 and 1987). Furthermore, recent research (Clark and Oswald, 2002) attempted to measure the importance of material goods, expressed in the form of money and wealth, in determining personal happiness. They arrived at the conclusion that *money can buy happiness*, using data from the British Household Panel Survey (BHPS), a representative sample of some 18,000 individuals living in Britain in the 1990s. This survey includes a question that asks whether the respondents have been recently unhappy or depressed, and a number of straightforward questions that seek to measure individual contentment, such as whether respondents feel "able to enjoy normal day-to-day activities"; whether they have been losing self-confidence; whether they are losing sleep. Clark and Oswald (2002) developed and used statistical regression models of happiness that measured the impact of different life events upon human well being. They concluded that their new statistical method could, in principle, be used to value any kind of event in life. More recently, the UK Prime Minister’s Strategy Unit published a report on what makes people satisfied with their lives (Strategy Unit, 2003. It can be argued that there is a need to critically review this kind of work and evaluate it under different conceptions of well-being and definitions of happiness over a person’s lifetime. Further, it would be interesting to add a geographical dimension to the measurement of perceived levels of happiness (Dorling, 1996).

This paper presents a research framework for a geographical approach to measuring happiness, which is part of a wider investigation into the factors and life events (Balals and Dorling, 2007) that make different types of individuals happy and how these affect the overall structure and cohesion of society. The paper presents a spatial microsimulation model of happiness and well-being which is capable of estimating happiness levels at different geographical scales. It also presents a framework for the combination of the spatial microsimulation model outputs with data obtained using remote sensing in order to determine the extent to which geographical proximity to entities pertaining to the natural and built environment (e.g. living near a river, lake, by the sea, a park, a motorway etc.) may be affecting subjective well-being and happiness, in addition to socio-economic, demographic and contextual variables.

2. Building a spatial microsimulation models of happiness

> "Simulation is a critical concept in the future development of modelling because it provides a way of handling complexity that cannot be handled analytically. Microsimulation is a valuable example of a technique that may have increasing prominence in future research."
>
> (Wilson, 2000:98)

Simulation-based spatial modelling is an expanding area of research, which has enormous potential for the evaluation of the socio-economic and spatial effects of major developments in the regional or local economy. Spatial microsimulation methodologies involve the merging of census and survey data to simulate a population of individuals within households, whose characteristics are as close to the real population as it is possible to estimate. Dynamic spatial micro-simulation involves forecasting past changes forward to produce as best an estimate as possible of individual's circumstances in the future - were current trends to continue- or under different policy scenarios. The research presented here builds on on past and on-going spatial microsimulation work (Ballas and Clarke, 2000; Ballas, 2004; Ballas and Clarke, 2001; Ballas *et al.*, 2005) by developing and using a spatial microsimulation methodology to define personal happiness and quantify and estimate its degree for different types of individuals, living in different areas. As Sen (1987) points out:

> "A person who has had a life of misfortune, with very little opportunities, and rather little hope, may be more easily reconciled to deprivations than others reared in more fortunate and affluent circumstances. The metric of happiness may, therefore, distort the extent of deprivation in a specific and biased way."
>
> (Sen, 1987: 45)

It can be argued that since the degrees of well-being vary significantly between different individuals (different people are made happy by different things, life-courses etc.), microsimulation is an ideal methodology to study and quantify happiness at the individual level. Further, one of the main advantages of microsimulation is the ability to link data sets from different sources. For the purposes of this study the microsimulation method is being used to link the British Household Panel Study (BHPS) mentioned in the introduction to Census small area outputs (building on on-going work on how this link can be satisfactorily achieved). In this manner a geographical dimension can be added to the existing BHPS research (such as the research on happiness by Clark and Oswald, 2002; Blanchflower and Oswald, 2004; Layard, 2005). In particular, in the context of the research presented here, a spatial microsimulation model has been developed in order to estimate the geographical distribution of individual contentment through the 1990s. The model links the first wave of the BHPS (1991) to Census Small Area Statistics on the basis of socio-economic variables. The BHPS is an annual survey of the adult population of the UK, drawn from a representative sample of over 5000 households (Berthoud and Geshuny, 2000). In the context of this paper the reweighting methodology described in Ballas (2004) and Ballas *et al.* (2005) has been employed to re-adjust the weights of the records of the BHPS households so that they would fit census small area statistics tables in 1991 and 2001. The simulated database was then used to estimate subjective happiness at different geographical levels.

It should be noted that all the households in the BHPS are given a weight that compensates for error, bias, refusals etc. These weights can be readjusted in order to fit small area descriptions, such as census small area data (the weights can be readjusted so that they would add up to these small area descriptions). An example of how such a readjustment can be carried out is described in tables 1-4 (after Ballas *et al.*, 2005). In particular, table 1 gives a hypothetical individual microdata set comprising 5 individuals, which fall within two age categories. Further, table 2 depicts a small area statistics table for a hypothetical area, whereas table 3 depicts a cross-tabulation of the hypothetical microdata set, so that it can be comparable to table 2. Using these data it is possible to readjust the weights of the hypothetical individuals, so that their sum would add up to the totals given in table 1. In particular, the weights can be readjusted by multiplying them by the value of the cell in table 2, which denotes the category in which they belong over the respective cell in table 3. This can be expressed as follows:

$$n_i = w_i \; x \; s_{ij}/m_{ij} \qquad (1)$$

where n_i is the new household weight for household *i*, w_i is the original weight for household *i*, s_{ij} is element *ij* of table *s* (small area statistics table, which is the equivalent of table 2) and m_{ij} is element *ij* of table *m* (reproduced table using the household microdata original weights, which is the equivalent of table 3 in the example). Table 4 depicts how this simple formula is used to readjust the weights of the individuals in the example.

Table 1: A hypothetical microdata set (original weights: table ***w***)

Individual	Sex	Age-group	Weight
1st	Male	Over-50	1
2nd	Male	Over-50	1
3rd	Male	Under-50	1
4th	Female	Over-50	1
5th	Female	Under-50	1

Table 2: Hypothetical small area data tabulation (table ***s***)

Age/sex	Male	Female
Under-50	3	5
Over-50	3	1

Table 3: The hypothetical microdata set, cross-tabulated by age and sex. (table ***m***)

Age/sex	Male	Female
Under-50	1	1
Over-50	2	1

Table 4: Reweighting the hypothetical microdata set in order to fit table 2.

Individual	Sex	age-group	Weight	New weight
1st	Male	Over-50	1	1 x 3/2 = 1.5
2nd	Male	Over-50	1	1 x 3/2 = 1.5
3rd	Male	Under-50	1	1 x 3/1 = 3
4th	Female	Over-50	1	1 x 1/1 = 1
5th	Female	Under-50	1	1x 5/1 = 5

The above process can then be used to reweight the individuals to fit another table. In the context of this paper this reweighting procedure was adopted iteratively to readjust the BHPS households weights so that they would fit the electoral wards of areas in Wales on the basis of the following census statistics tables in 1991 and 2001:

1. Household tenure status
2. Occupation of head of household
3. Number of cars
4. Household type (single, married, lone parent)

The generated weights for each household represent the probabilities of BHPS households to “live” in a given area.

After generating BHPS household weights for each small area, the next step was to convert the decimal weights into integer weights. This conversion was carried out with the implementation of an algorithm that maximised the likelihood of households with the highest decimal weights to be represented in a small area (for more details see Ballas *et al.*, 2005).

For the purposes of this paper the first wave (1991) of the BHPS was used. The implementation of the methodology described above resulted in a microdata set at the ward level for Wales, which contained a number of non-census variables, including happiness questions such as these described in table 5.

Table 5: Measuring subjective well-being in the BHPS – the GHQ set of questions as they appear on the BHPS questionnaire: Here are some questions regarding the way you have been feeling over the last few weeks. For each question please ring the number next to the answer that best suits the way you have felt. Have you recently:

GHQ questions / responses	**1**	**2**	**3**	**4**
1. Been able to concentrate on whatever you are doing?	Better than usual	Same as usual	Less than usual	Much less than usual
2. Lost much sleep over worry?	Not at all	No more than usual	Rather more than usual	Much more than usual
3. Felt that you are playing a useful part in things?	More than usual	Same as usual	Less so than usual	Much less than usual
4. Felt capable of making decisions about things?	More so than usual	Same as usual	Less so than usual	Much less capable
5. Felt constantly under strain?	Not at all	No more than usual	Rather more than usual	Much more than usual
6. Felt you could not overcome your difficulties?	Not at all	No more than usual	Rather more than usual	Much more than usual
7. Been able to enjoy your normal day-to-day activities?	Much more than usual	Same as usual	Less so than usual	Much less than usual
8. Been able to face up to your problems?	More so than usual	Same as usual	Less able than usual	Much less able
9. Been feeling unhappy and depressed?	Not at all	No more than usual	Rather more than usual	Much more than usual
10. Been losing confidence in yourself?	Not at all	Not more than usual	Rather more than usual	Much more than usual
11. Been thinking of yourself as a worthless person?	Not at all	No more than usual	Rather more than usual	Much more than usual
12. Been feeling reasonably happy all things considered?	More so than usual	About same as usual	less so than usual	Much less than usual

It should be noted that the accuracy of the estimation of these variables depends on the degree of their correlation with the census variables which were used as constraints in the simulation. For instance, assuming that subjective happiness is to an extent correlated to the "constraint" small area statistics tables describe above, it is possible to estimate the geographical distribution of happiness. Figures 1 and 2 show the estimated geographical distribution of happiness in Wales in 1991 and 2001 respectively, (aggregated to parliamentary constituency level) on the basis of such assumptions.

Figure 1: Estimated geographical distribution of happiness (% happy more than usual) in Wales, 1991

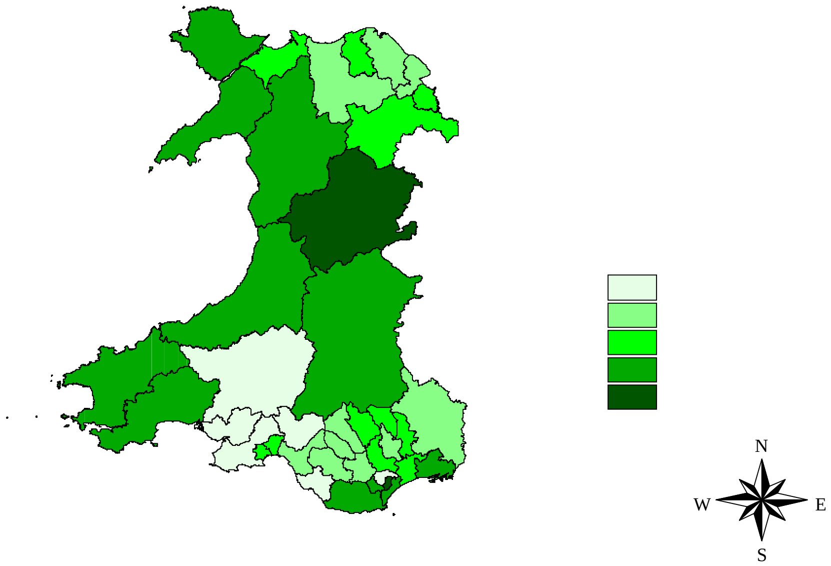
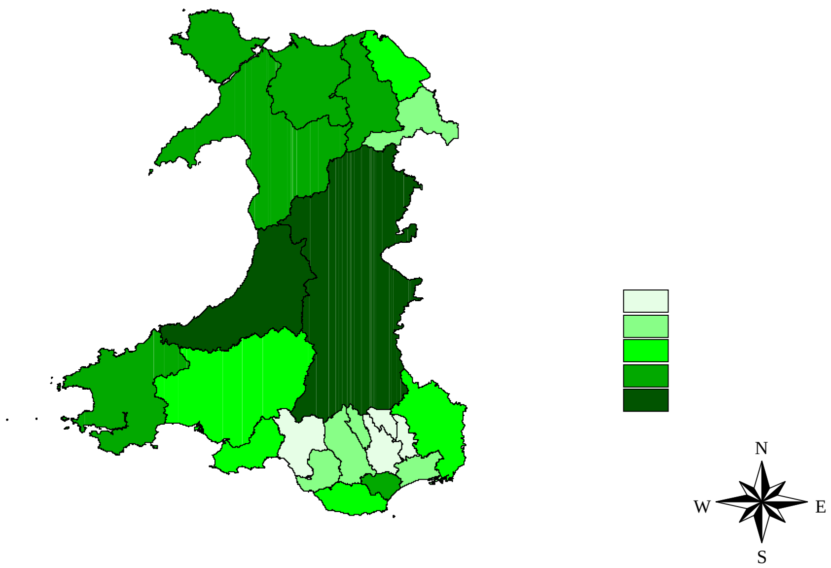


Figure 2: Estimated geographical distribution of happiness (% happy more than usual) in Wales, 2001



3. Combining spatial microsimulation model outputs with remote sensing data

The analysis briefly described above can be enhanced with the use of remote sensing very high resolution data in order to determine the extent to which geographical proximity to entities pertaining to the natural and built environment (e.g. living near a river, lake, by the sea, a park, a motorway etc.) may be affecting subjective well-being and happiness, in addition to socio-economic, demographic and contextual variables. In the context of this paper we are presenting a modelling framework to achieve such a much building on past work (Ballas et al., 2000). This modelling framework involves the combination of secondary socio-economic data, such as the household panel surveys and the censuses of population, as well as simulated data (microsimulation outputs) with satellite very high resolution imagery. These can be combined with the use of GIS-based spatial modelling.

One difficulty at present with spatial microsimulation models such as the model described in the previous section is that it is based on probabilities that are calculated from known distributions (provided by data sources such as the Census of population) at the small area level (e.g. the ED level in the UK). It is not possible to know precisely whereabouts within a small area such as the Census Enumeration district or Census Output Area of a particular household (high income or low income) is actually located. For many policy purposes that is not a major problem – it is the overall effects on the locality that is most important. However, it can be argued that for certain applications this would be a worthwhile addition – especially, and in the context of the work presented here, when looking at the impact of the natural and built environment upon happiness. Using remote sensing techniques it is possible to obtain a point data set of houses, which would contain the housing type attribute, with the use of remote sensing methodologies (Ballas et al., 2000). These point data set can then be linked to spatially disaggregated (at the ED level) microsimulated households, in order to disaggregate the simulated population at the ED level. In other words, the task of this modelling exercise would be to populate the remotely sensed residential properties with attribute data. Table 6 lists the attributes that can be used as a link between the remote sensing generated database and the microsimulation output.

Further, figure 3 depicts schematically and in a simplified manner the geographical databases that are typically generated by microsimulation models and remote sensing methodologies and how these can be linked. As can be seen, these databases can be joined on the basis of the fields that they have in common, such as the housing type and house size. However, it can be argued that all the attributes listed in table 6 can be used to build an index of similarity between a remotely sensed house and a microsimulated synthetic household. Moreover, the linkage between the two databases can be achieved with the use of statistical matching or data fusion techniques. It should be noted that although statistical matching (also known as data fusion) has a relatively long history, its theoretical basis is somewhat narrow and there is no established, tested and widely applied methodology (Paas, 1986; Sutherland *et al.*, 2001). Data fusion involves the statistical matching of data for statistically similar records from two or more micro-databases.

The approach suggested here draws on past research in quite different contexts, such as the work of Radner (1981), who used statistical matching techniques in the estimation and analysis of the size distribution of family unit personal income. In addition, Paas (1986) discusses data matching in a statistical theory framework and describes ways of testing empirically the quality of matching methods. One of the ways of combining spatial microsimulation and remote sensing methodologies for the estimation of spatially disaggregated population microdata is to employ a statistical matching approach in order to derive the data set:

$$S = (ed,h,g, ed,h,x_1,x_2,\ldots,x_n) \qquad (2)$$

from:

$$M = (ed,h, x_1,x_2,\ldots,x_n) \qquad (3)$$

$$R = (ed,h,g) \qquad (4)$$

where:

- S: estimated spatially disaggregated population microdata set at the *house level*.
- M: microsimulation output - spatially disaggregated population microdata set
- R: remotely sensed data set
- *ed*: the Enumeration District location of each household

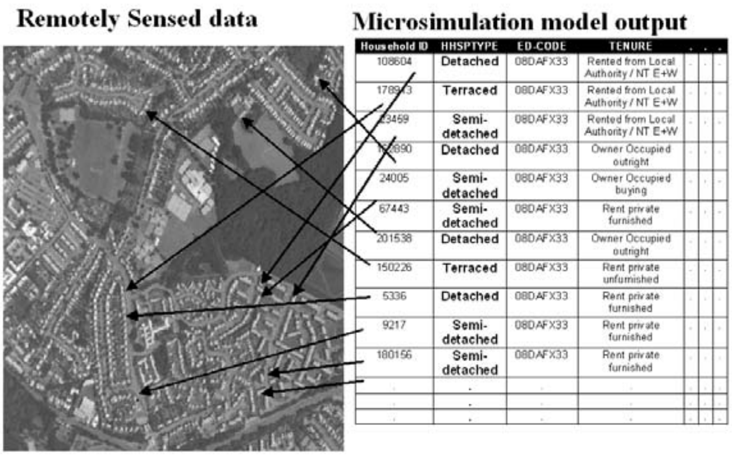
- h: the housing type
- g: the exact geographical co-ordinates of each house
- $x_1 \ldots x_n$: socio-economic and demographic attributes

In the context of a statistical matching procedure, each record of the S microsimulation data set could be assigned to one record (house) of data set R, in order to obtain S. This assignment would be based upon the similarities of the common variables (e.g. housing type), which could be expressed by some form of distance function. Further, house size can be used as a surrogate of socio-economic attributes, so that it can be taken into account when joining the two data sets.

Spatial microsimulation output	Remotely sensed data
No. of residents in household (as a proxy to house size)	Land use
House type	Property size
Number of cars (as a proxy to house size)	House type
Number of rooms in household space (as a proxy to house size)	…
…	

Table 6: Database attributes that can be used for the linkage

Figure 3: Combining spatial microsimulation and Remote Sensing (Ballas, Barr and Clarke, 2000)



4. Concluding comments

This paper presented a new framework for the combination of remotely sensed data with secondary and simulated data sets in order to provide a powerful database for the geographical analysis of subjective happiness and well-being, building on a rapidly growing body of inter-disciplinary research in this field. It should be noted that such a framework can provide very interesting insights into the local factors that may be affecting happiness and well-being. It may also be very useful for the analysis of local policy outcomes and it could also inform local debates on issues such as green-spaces and the geographical allocation and extent of geographical features that may be affecting happiness and local well-being. In addition, the framework presented here may also offer potential for calibration and for dynamic modelling of populations. The visualisation of the area being modelled would provide useful additional diagnostic information and would allow comparisons to be made between simulated households and real households. In addition, the framework suggested here would add much to the potential of remotely sensed data. It would be possible to put estimations on the types of buildings in terms of housing types and characteristics of their inhabitants (and their estimated happiness and well-being levels). Clearly, it is not possible to categorically say what types of families were in each building. However, it may be possible to give an estimation of the types of families within blocks thus giving very detailed portraits of small areas of our cities.

Acknowledgements: Funding from the Economic and Social Research Council (research fellowship grant number RES-163-27-1013) is gratefully acknowledged by Dimitris Ballas. The British Household Panel Survey data were made available through the UK Data Archive. The data were originally collected by the ESRC Research Centre on Micro-social Change at the University of Essex, now incorporated within the Institute for Social and Economic Research. All responsibility for the analysis and interpretation of the data presented in this paper lies with the authors.

References:

Ballas, D. (2004) Simulating trends in poverty and income inequality on the basis of 1991 and 2001 census data: a tale of two cities, *Area* 36**.**2, 146-163

Ballas D, Barr, S. & Clarke G. P., (2000), *Combining census data and remotely sensed data for the generation of population microdata: a spatial microsimulation approach*, paper presented at the 40th European Regional Science Association (ERSA) congress, Barcelona, 29 August – 1 September 2000

Ballas, D., Clarke, G.P. (2000), GIS and microsimulation for local labour market policy analysis, *Computers, Environment and Urban Systems*, *24*, 305-330

Ballas, D., Clarke, G.P. (2001), Modelling the local impacts of national social policies: a spatial microsimulation approach, *Environment and Planning C: Government and Policy, 19,* 587 – 606

Ballas, D**,** Dorling, D (2007), Measuring the impact of major life events upon happiness, *International Journal of Epidemiology,* 36, 1244-1252, doi:10.1093/ije/dym182

Ballas, D., Rossiter, D, Thomas, B, Clarke, G.P, Dorling, D (2005) *Geography matters: simulating the local impacts of national social policies*, Joseph Rowntree Foundation, York

Broome, J (1999), *Ethics out of economics*, Cambridge University Press, Cabridge

Blanchflower, D, Oswald A (2004), Well-Being Over Time in Britain and the USA, *Journal of Public Economics*, Vol. 88, pp. 1359-1386

Clark, A, Oswald, A, (2002), A Simple Statistical Method for Measuring How Life Events affect Happiness, *International Journal of Epidemiology*, 2002, 31(6), 1139-1144.

Dorling, D. (1996) Be cheerful, strive to be happy, identifying disadvantaged areas: health wealth and happiness, *Radical Statistics*, 62, 8-21

Griffin, J (1986), *Well-being: its meaning, measurement and moral importance*, Clarendon, Oxford

Layard, R. (2005), *Happiness: Lessons from a New Science*, Allen Lane

Lear, J (1988), *Aristotle: the desire to understand*, Cambridge University Press, Cambridge

Little, I M D (1957), *A Critique of Welfare Economics*,Clarendon Press (2nd ed.), Oxford

Paas, G. (1986), Statistical match: evaluation of existing procedures and improvements by using additional information, in G.H. Orcutt, J. Mertz, H. Quinke (eds) *Microanalytic Simulation Models to Support Social and Financial Policy*, North-Holland, Amsterdam, 401-421

Ross, W D (1923), *Aristotle*, Metuhen & Co, London

Sen, A (1982), *Choice, Welfare and Measurement,* Blackwell, Oxford

Sen, A. (1987), *On Ethics and Economics*, Blackwell, Oxford

Sutherland, H., Taylor, R., Gomulka, J. (2001), *Combining household income and expenditure data in policy simulations*, Working paper MU0101, The Microsimulation Unit, Department of Applied Economics, University of Cambridge, Cambridge CB3 9DE

Strategy Unit (2003), *Life Satisfaction: the state of knowledge and implications for Government*, London: Cabinet Office [Web: http/www.cabinet-office.gov.uk/innovation/2001/futures/attachments/ls/paper.pdf], Accessed 17 February 2003

Sumner, L W (1996), *Welfare, Happiness and Ethics*, Clarendon Press, Oxford

Wilson, A.G. (2000), *Complex Spatial Systems: the modelling foundations of Urban and Regional Analysis*, Prentice Hall, London