# SOCIAL INEQUALITIES AND OCCUPATIONAL STRATIFICATION

## METHODS AND CONCEPTS IN THE ANALYSIS OF SOCIAL DISTANCE

PAUL LAMBERT
DAVE GRIFFITHS

# Social Inequalities and Occupational Stratification

Paul Lambert • Dave Griffiths

# Social Inequalities and Occupational Stratification

## Methods and Concepts in the Analysis of Social Distance

palgrave
macmillan

Paul Lambert
University of Stirling
Stirling, UK

Dave Griffiths
University of Stirling
Stirling, UK

ISBN 978-1-137-02252-3 ISBN 978-1-137-02253-0 (eBook)
https://doi.org/10.1057/978-1-137-02253-0

Library of Congress Control Number: 2018933593



Printed on acid-free paper

This Palgrave Macmillan imprint is published by the registered company Macmillan Publishers Ltd. part of Springer Nature.
The registered company address is: The Campus, 4 Crinan Street, London, N1 9XW, United Kingdom

# Contents

# List of Figures

# List of Tables

# 1

# Introduction

## 1.1 Social Distance Between Occupations

In what ways can the analysis of social connections between individuals inform us about social stratification and inequality? Our approach in this book it to connect data on the *occupations* held by people, with records of their social connections. This can tell us about the 'social distance between occupations', that is, the extent to which different occupations are linked to each other by higher or lower volumes of social interactions. We show in particular that our own and others' analyses of the social distance between occupations can reveal important and sometimes unanticipated patterns—particularly concerning stability—in social inequalities.

We focus particularly on occupations (see also Chaps. 2 and 3). In sociology in particular, the occupational order is commonly regarded as a consistent marker of long-term position within the structure of economic inequality, not least because resource distribution through occupational pay is a major source of economic inequalities (e.g. Wright 2005; Parkin 1972). Occupations are also helpful indicators because they are reasonably easy to measure for most people, and there are numerous alternative occupation-based measures of stratification available for use (e.g. Rose and Harrison 2010).



P. Lambert, D. Griffiths, *Social Inequalities and Occupational Stratification*,
https://doi.org/10.1057/978-1-137-02253-0_1

We report a consistent empirical link between the social connections that are held by the incumbents of occupations and the wider social structure of inequality or 'social stratification'. Our text describes methodologies that can be used to explore and interpret data on this link and discusses why and how this link arises. Chief amongst the methods that we cover is a tradition known as the 'CAMSIS' approach to analysing social connections between occupations ('Cambridge Social Interaction and Stratification scales, see www.camsis.stir.ac.uk). Both authors have worked in projects related to the CAMSIS tradition for a number of years, and our text summarises recent research and ongoing debates related to that tradition. We also explore alternative ways of analysing data on social connections and occupations and undertake analyses that draw upon both the CAMSIS tradition and other important approaches.

Figures 1.1 and 1.2 provide two empirical illustrations of the relationship between the social distance between occupations, and wider social inequalities. They are derived from two representations of social structure

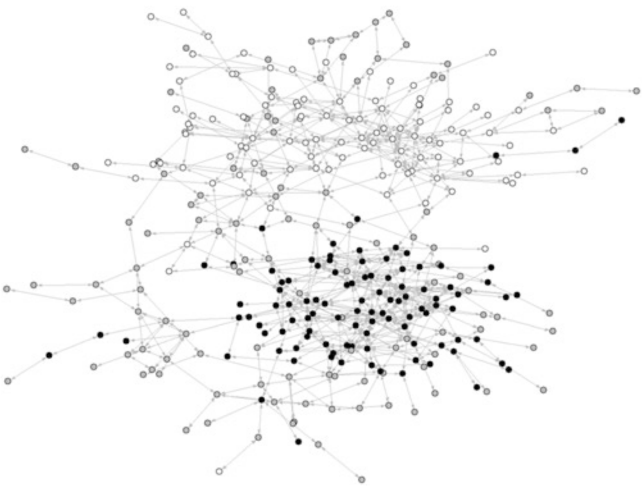


**Fig. 1.1** A sociogram depicting 'networked occupations' (i.e. links between those occupations whose incumbents are disproportionately likely to have social connections with each other). Source: IPUMS-I data (MPC 2015) for occupations in the USA in 2000; occupations shaded as 'professional' (dark), 'routine non-manual' (grey), and 'manual' (white)

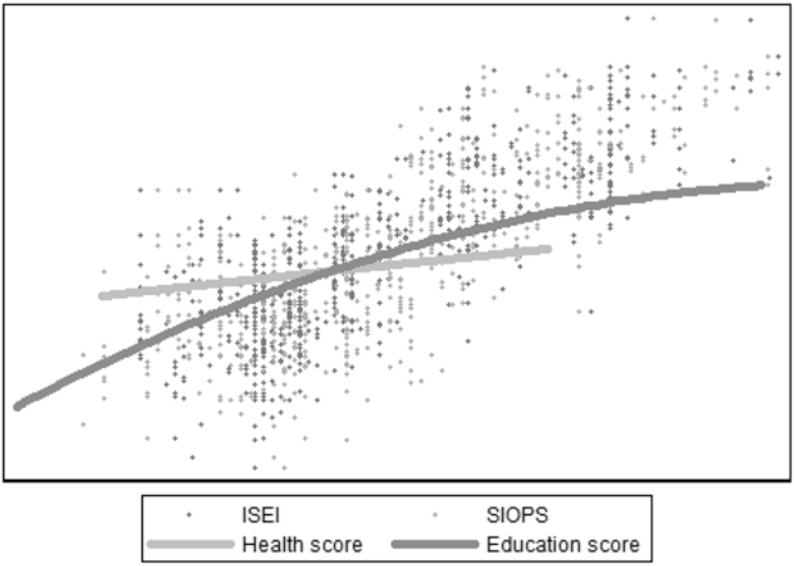



**Fig. 1.2** Depiction of the relationship between an occupation-based social interaction distance scale ('CAMSIS') and other measures of social circumstances. Source: Analysis of UK data from 1991 (BHPS, see University of Essex 2010). CAMSIS score on horizontal axis. Lines for 'Health' and 'Education' score show relation between measures based on those responses and CAMSIS. 'International Socio-Economic Index' ('ISEI') and 'Standard International Occupational Prestige Scale' ('SIOPS') are occupation-based indexes of 'socio-economic status' and 'prestige' (see Chap. 3)

that are used throughout this text, namely, social network analysis and social interaction distance analysis. Figure 1.1 is a network 'sociogram' which shows a structure of social connections between occupations—its points show occupations, and lines are drawn connecting those occupations that have disproportionately many social connections between them (see especially Chaps. 7 and 8). Like most sociograms, it shows a web-like structure in which some units are close together and others are considerably separated. However, ties are much more likely to arise between occupations whose incumbents occupy similar positions in the structure of social inequality, and so we would argue that the pattern of ties between occupations that is revealed in the figure can also be interpreted as a structural map of consequential social inequality.

Figure 1.2 shows the relationship between a 'CAMSIS scale for occupations' and other data about social inequality. The CAMSIS scale is a representation of social inequality obtained by finding the main dimension in empirical patterns of social interactions between occupations (see Chaps. 4, 5, and 6 for elaboration). Occupations are assigned scores that represent their position in that dimension. The social interactions which lead to the derivation of the CAMSIS scale can be thought of as examples of how social connections influence social inequalities. The figure illustrates how the derived CAMSIS scale is itself a meaningful representation of social inequality (because it is linked to many other social inequalities—the figure shows its correlation with measures related to income, education and health).

Figures 1.1 and 1.2 show that there is something interesting in the empirical relationship between social connections, occupations, and social inequality. Quite a lot of our text below covers technical issues that are concerned with accurately depicting these relationships in appropriate detail. However, there is also a wider contribution that emerges from the detailed empirical analyses: how we understand and theorise social inequality can ultimately be transformed when we better understand the relationship between social connections and key social markers such as occupations.

## 1.2 Social Distance in Everyday Life

Have several different members of your family held the same occupation at some point? Have many of your friends had fairly similar educational experiences to your own? When you take a holiday, do you often come across people of similar occupational and educational backgrounds at your destination? Many people answer 'yes' to these questions: in most circumstances, people who are amongst our family and friends are also to be found in social positions that have similar socio-economic circumstances, for instance, that involve similar jobs or educational backgrounds, to our own (e.g. McPherson et al. 2001).

The association between social stratification and social connections is a strong one, although it is actually a relationship that can work in more than one direction. One pathway of dependency goes from socio-economic situation to social relationships: for example, people often meet their friends and future family through their workplace, during their education, and so forth. Indeed, similarity of structural location is often regarded as a key criterion when forming social relationships (e.g. Skopek et al. 2011). However, the pathway might equally be reversed: for example, the job that a person holds may well be influenced by advice or support that they received from social connections at the time of recruitment (e.g. Christakis and Fowler 2010).

### 1.2.1 Osborne and Gascoigne

Readers from the UK will need little introduction to the politician and commentator George Osborne and to the former professional footballer Paul Gascoigne. Both men have been prominent actors in the UK's public sphere for several decades. Paul Gascoigne was regarded by many as one of the best footballers in the world during his playing career (1985–2004), but his spontaneous, often mischievous personality attracted even more extensive media attention and ensured his celebrity status both during his playing days and for many years after his retirement. George Osborne rose to prominence in the 2000s as a conservative politician who held one of the most senior governmental roles ('Chancellor of the Exchequer', or finance minister, between 2010 and 2016); aside from his political activities, Osborne also made several other public contributions, including authoring books, editing a major newspaper, and holding senior visiting appointments in the university sector.

An interesting feature to the stories of George Osborne and Paul Gascoigne is that popular accounts of their lives often portray their 'structural' experiences as a direct function of their social relationships. Consider the following salient characteristics of George Osborne.

### 1.2.2 Characterising George Osborne

- Occupations: editor of a national newspaper, executive financial advisor, former National Government minister
- Parents' occupations: Baronet and company executive; Baronet's wife and charitable work

- Friends with: ministers, mayors, bankers, colonels
- Leisure pursuits: polo, hunting
- Education: Eton college school and Oxford University

Osborne is routinely portrayed within the UK as a privileged 'toff'.[1] Educated at an elite private school, throughout his life Osborne has also socialised with the most advantaged individuals in the UK, and maintains social contacts with people from the most privileged circles—relationships that many believe helped him to secure one of the most powerful structural positions in the nation (e.g. Ganesh 2012).

### 1.2.3 Characterising Paul Gascoigne

- Occupations: unemployed, former professional footballer and coach
- Parents' occupations: building labourer, cleaner
- Friends with: footballers, roofers, bouncers, broadcasters
- Leisure pursuits: pubs, fishing, football
- Education: comprehensive school to age 16

Paul Gascoigne's circumstances contrast markedly with George Osborne's. Gascoigne grew up in a relatively deprived and challenging environment. Both during and after his sporting success, he continued to socialise with people from his original background, and it is often suggested that he maintained social ties and behaviours that inadvertently led to significant personal problems, including heavy drinking (e.g. Gascoigne 2005). For Paul Gascoigne, social interactions may have left the scars of serious health and welfare problems to the present day, whereas for George Osborne, maintaining the social contacts of his origins may have brought great personal privilege. Both have talents and skills that have contributed to their successes independently of their social surroundings, and our brief portrayal ignores other complexities, but these two stereotypical accounts serve to illustrate how social resources, social support, and long-term structural outcomes can often be connected.

It can seem obvious that social relationships and social support can matter to what people achieve; myriad sociological studies can be cited to support this point in more systematic detail (e.g. Ermisch et al. 2012; Christakis and Fowler 2010; Saunders 2010; Attewell and Lavin 2007; Devine 2004). However, when we think of the lives of George Osborne and Paul Gascoigne, we could also say that they illustrate the formation of social structure itself and the endurance and reproduction of social inequality that results from differential socialisation. George Osborne contributes to defining the most advantaged in society as *those echelons that feature people like George Osborne*. Paul Gascoigne's example suggests that *people like Paul Gascoigne* tend to define a less-advantaged social circumstance. The argument here is that the social structure comes to be shaped according to the social relations of the individuals who inhabit its different positions. Patterns of social connections—sometimes described as a 'space of social interactions' (e.g. Bottero 2005)—can serve to define other patterns of inequality.

A key feature to recognise about social interactions is that people tend on average, through their voluntary behaviours, to maintain existing social interaction patterns. The preference for stability arguably fosters social reproduction of structural outcomes (e.g. Bourdieu 1984). One example can be seen in self-exclusion from social activities. Most commonly, when people feel themselves to be a 'fish out of water', they withdraw from the social space and leave it to others to inhabit. We would anticipate that most readers of this book have at some point felt uncomfortable in a homogenous social environment which was not one that they usually inhabited. Self-exclusion might seem a trivial process, but the point is that social connections are influential for important outcomes related to social inequality. Therefore, when we acquiesce to existing divisions in social interactions, we are probably, whether consciously or not, contributing to the perpetuation or exacerbation of the socio-economic structures that sit beside them. Bourdieu (1984) poignantly described how when people from relatively disadvantaged circumstances retreat from social environments that are 'not for the likes of us', they are in effect 'refusing what they are refused'.

At the same time that outsiders self-exclude, insiders take comfort from maintaining and reproducing their social collective—irrespective of whether, in structural terms, the familiar is relatively advantaged or disadvantaged. Indeed, many authors have highlighted how people in relatively deprived environments nevertheless express strong preferences and rationalisations for maintaining their current situation and social connections rather than making changes (e.g. McKenzie 2015; Tyler 2013), and in some scenarios a socio-economically positive 'mobility experience' might nevertheless bring negative personal impacts (Friedman 2014). People might even turn down opportunities to foster social connections that might otherwise have offered some advantages. Consider the hypothetical example of a tax lawyer who moves into a new village. They might in principle expect almost any group of locals to welcome their presence, as they may represent a potentially valuable source of specialist information and advice. Nevertheless, this might not happen in practice—if they wander into a bar inhabited by building site labourers and their friends, for instance, they might well receive short shrift, perceived to be intruding into the 'wrong' social space.

Packard (1959) demonstrated individuals' deeply ingrained values about social behaviour that was appropriate to their social status. Observations illustrated how people would withdraw from social environments that they felt to be very much above or below them, but would expend great energy in participating in activities which involved people of similar status (or those of slightly higher-status levels, to which they realistically aspired). More recently, Sayer (2005) gave a compelling account of how social behaviour, and with it individuals' aspirations and life objectives, is shaped by a sense of 'class'-appropriate behaviour, manifested in a sense of 'shame' experienced when individuals are outwith the domains in which they are comfortable, such as in environments where other class behaviours are dominant. To avoid such problems, the solution is obvious: most people, most of the time, prefer to engage socially with others of similar circumstances to themselves; most people choose to like people who are like themselves. This behaviour has further social implications: the social structure, in turn, evolves in a manner that reflects the average patterns of separation between people in their social choices.

## 1.3 The Structure of This Book

Whilst social interactions are important in charting and explaining structures of social stratification and inequality, in our opinion they are often under-emphasised in relevant social science research. Some social science traditions concentrate upon explanations that are rooted in overarching social forces as if they applied to all individuals equally, apparently without recognition that interactions between people could readjust structural inequalities. Many empirically oriented social science studies take a largely 'individualist' approach, neglecting to use data or insights rooted in social interactions between individuals. In empirical research, it can seem relatively difficult and complicated to develop an analysis that does take account of social interactions, yet it is not impossible. The contents of this book involve making fuller use of social interaction data in the analysis of complex social inequalities and in particular in relationship to occupations. We ultimately argue that analyses that adopt this strategy can achieve a more reliable empirical characterisation, and more meaningful theoretical interpretation, of fundamental forms of social inequality in modern societies.

The forthcoming chapters expand upon how social connections between occupations can be investigated empirically, and the contributions to our understanding of social structure that emerge. In Chap. 2, we discuss underlying theories associated with social distance and social connections in relation to occupations. We include definitions of terms and wider theoretical perspectives on how social stratification works and why it is related to social distance. In Chap. 3, we provide an account of existing traditions in using data on occupations to measure social stratification. This gives some background on traditions of analysis that we engage with and highlights some technical and 'operational' aspects of this work that are sometimes overlooked but that can be quite important to empirical research.

Chapters 4, 5, and 6 review the CAMSIS approach and its contribution to studying social inequality, and Chaps. 7 and 8 discuss another strategy for summarising social connections between occupations using network analysis. In both cases, we provide an overview of the methodologies (within Chaps. 4 and 7, respectively), before elaborating upon the operationalisation and application of the approaches. In Chap. 5, we provide a review of the empirical features of CAMSIS scales, and in Chap. 6, we discuss at some length the elements of work involved in generating new CAMSIS scales based upon social interaction patterns (we hope that this chapter will be a useful reference to anyone interested in undertaking their own statistical research in this field). Using a different approach to network analysis, Chap. 7 defines ‘networked occupations’ as those with unusually high volumes of social connections between them, whilst Chap. 8 looks at methods for exploring patterns of networked occupations.

Chapters 4, 5, 6, 7, and 8 incorporate some technical materials that are concerned with empirical data resources and their analysis. We hope this content will enable some readers to implement the approaches that we cover in their own work—for example, in Chaps. 6 and 8, we include some brief examples of software command language ‘syntax’ and mention further downloadable files in the languages of four relevant packages—namely, Stata (Statacorp 2015), lEM (Vermunt 1997), R (R Core Team 2016), and Pajek (see de Nooy et al. 2011). We have tried to introduce these elements, when they arise, in an accessible way, with cross-references available for further details.

Our text then features three chapters that look at some specific issues in studying social inequality that are raised by our general approach of analysing data on social interactions between the incumbents of occupations. Chapter 9 focusses upon ways of modelling fine-grained occupational influences such as by exploiting ‘random effects’ or ‘multilevel’ models for this purpose. Chapter 10 explores the interplay between social interactions involving occupations, and data on educational qualifications and experiences, particularly in the context of educational expansion. In Chap. 11, we explore the extent to which the analysis of social stratification and inequality, when it draws upon data about social interactions, can or should take account of multidimensional structures of social inequality, such as the interaction between stratification inequalities, social distances, and social divisions like gender and ethnicity. Lastly, in Chap. 12, we conclude with comments on the empirical and theoretical implications of rooting an understanding of social stratification and inequality in data on social distances and occupations.

## Notes

1. A term used in the UK to refer disparagingly to people from very advantaged backgrounds. The term reflects a perception of socially distinct, inaccessible privilege.

## Bibliography


Attewell, P., & Lavin, D. E. (2007). *Passing the Torch: Does Higher Education for the Disadvantaged Pay Off Across the Generations?* New York: Russell Sage Foundation.

Bottero, W. (2005). *Stratification: Social Division and Inequality*. London: Routledge.

Bourdieu, P. (1984). *Distinction*. London: Routledge and Kegan Paul.

Christakis, N., & Fowler, J. (2010). *Connected: The Amazing Power of Social Networks and How They Shape Our Lives*. London: Harper Press.

de Nooy, W., Mrvar, A., & Batagelj, V. (2011). *Exploratory Social Network Analysis with Pajek*. Cambridge: Cambridge University Press.

Devine, F. (2004). *Class Practices: How Parents Help Their Children Get Good Jobs*. Cambridge: Cambridge University Press.

Ermisch, J., Jantti, M., & Smeeding, T. (Eds.). (2012). *From Parents to Children: The Intergenerational Transmission of Advantage*. New York: Russell Sage Foundation.

Friedman, S. (2014). The Price of the Ticket: Rethinking the Experience of Social Mobility. *Sociology, 48*(2), 352–368.

Ganesh, J. (2012). *George Osborne: The Austerity Chancellor*. Hull: Biteback Publishing.

Gascoigne, P. (2005). *Gazza: My Story*. London: Headline Press.

McKenzie, L. (2015). *Getting By: Estates, Class and Culture in Austerity Britain*. Bristol: Policy Press.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology, 27*, 415–444.

Minnesota Population Center. (2015). *Integrated Public Use Microdata Series, International: Version 6.4 [Machine-Readable Database]*. Minneapolis, MN: University of Minnesota.

Packard, V. (1959). *The Status Seekers*. London: Penguin.

Parkin, F. (1972). *Class Inequality and Political Order: Social Stratification in Capitalist and Communist Societies*. London: MacGibbon and Kee.

R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rose, D., & Harrison, E. (Eds.). (2010). *Social Class in Europe: An Introduction to the European Socio-economic Classification*. London: Routledge.

Saunders, P. (2010). *Social Mobility Myths*. London: CIVITAS: Institute for the Study of Civil Society.

Sayer, A. (2005). *The Moral Significance of Class*. Cambridge: Cambridge University Press.

Skopek, J., Schulz, F., & Blossfeld, H. P. (2011). Who Contacts Whom? Educational Homophily in Online Mate Selection. *European Sociological Review, 27*(2), 180–195.

StataCorp. (2015). *Stata Statistical Software: Release 14*. College Station, TX: StatCorp LLC.

Tyler, I. (2013). *Revolting Subjects: Social Abjection and Resistance in Neoliberal Britain*. London: Zed Books.

University of Essex, & Institute for Social and Economic Research. (2010, July). *British Household Panel Survey: Waves 1–18, 1991–2009 [Computer File]* (7th ed.). Colchester, Essex: UK Data Archive [distributor], SN: 5151.

Vermunt, J. K. (1997). *lEM: A General Program for the Analysis of Categorical Data*. Tilburg: Tilburg University.

Wright, E. O. (Ed.). (2005). *Approaches to Class Analysis*. Cambridge: Cambridge University Press.

# 2

# Homophily and Endogamy

## 2.1 Homophily and Endogamy

The phenomena of 'homophily' and 'endogamy' refer to the tendency of people, on average, to form social ties with others of similar circumstances. 'Homophily', which means the preference for similarity, is used in sociology to refer to similarities in circumstances between people who are linked by any form of social relationship (e.g. friendships, partnerships). The special case of homophily in marriage/cohabitation is usually referred to as 'homogamy'. Patterns of social contacts within group boundaries are known as 'endogamy'. 'Endogamy' in a literal sense refers to marriage/cohabitation within groups, but it is often generalised to other forms of relationship. It is well established that friendship and cohabitation patterns show homophily and endogamy consistently across societies and over time (e.g. McPherson et al. 2001; Kalmijn 1998; Mare 1991).

Many explanations for why homophily and endogamy are common social phenomena have been considered (Kalmijn 1998). People might deliberately seek them out. As one example, many cultures have traditions whereby other relatives, such as parents, influence relationship formations,



P. Lambert, D. Griffiths, *Social Inequalities and Occupational Stratification*,
https://doi.org/10.1057/978-1-137-02253-0_2

usually by matching up people on the basis of similar structural positions (cf. Penn and Lambert 2009, c9). In economics, it is argued that 'utility maximisation' tends to lead to homophily and endogamy. In this paradigm, if individuals have different levels of 'attractiveness' to others, it follows that the best strategy overall is to seek to form relationships with alters who possess similar level of attributes—'people on average do best in terms of their own welfare by marrying people like themselves' (Brynin and Ermisch 2008, p. 7).

Homophily and endogamy can also emerge through less deliberate practices. One mechanism is our tendency to form new relationships with people we meet at workplaces and educational institutions or through mutual connections brokered through such locations (e.g. Kalmijn and Flap 2001). We also tend, on average, to opt into other leisure and lifestyle activities that themselves mainly involve people from similar social positions (e.g. Bennett et al. 2009; Archer 2007). These tendencies stack the dice in favour that when we form new relationships, it will be with people of similar circumstances. There is also some evidence to suggest that individuals may be happier, on average, when they achieve homophily and endogamy (e.g. Brynin et al. 2008; Stutzer and Frey 2006; Lampard 1997)—although there are counterarguments that highlight the individual and societal benefits of non-homophilous diversity in relationships (e.g. Skvoretz 2013). Yet regardless of its consequences, it is certainly a common social habit to exhibit homophilous and endogamous behaviour.

The prevalence of homophily and endogamy means that empirical data about social connections between people can itself be used to define which positions in the social structure are similar. This logic of analysis is something that we draw heavily upon throughout this book. A nice example is found in Pearce and Gambrell's (2016) interactive graphical display, which allows a reader to browse over a long list of occupational units and identify the occupations that are most commonly held by the partners of their incumbents in the contemporary USA. When doing so it is obvious that the links between occupations are far from random—in general, when one occupation stands out as being unusually often connected to another, there is usually an evident similarity in the structural or socio-economic circumstances. However, Pearce and Gambrell's visualisation also provides a device to understand where the occupation itself sits in the social structure: if we knew little else about one particular occupation, we could draw conclusions about it simply by looking at the profile of occupations with which it is most commonly linked through marriage.

Naturally, there are numerous non-occupational forms of homophily and endogamy (e.g. Skvoretz 2013; McPherson et al. 2001). Indeed, the presence of homophily between categories arguably reflects that those categories are socially important (Smith et al. 2014). Non-occupational forms of homophily often emerge during the formation of social interactions—for instance, the tendency to date those of a similar ethnicity (Curington et al. 2015) or mechanisms whereby tastes in cultural consumption (Lizardo 2006) or social activities (Xu et al. 2000; Lampard 1997) bring people together. Sometimes, homophily may be created when individuals adapt to a friend or partner's interests (Upright 2004). Nevertheless, non-occupational forms of homophily are themselves sometimes rooted in occupational differences, for instance, through the correlation between social position and cultural interests (Bennett et al. 2009).

## 2.2 Social Connections and Social Distance

Data on 'social connections' allow us to study homophily and endogamy. We define 'social connections' as those links between people that are not wholly contractual, and that might reasonably be expected, in some way, to offer benefit to the people involved. Relationships more generally need not be non-contractual, nor mutually beneficial—making a similar definition, Hinde (1997, p. 37, as highlighted by Brynin and Ermisch 2008) considers social relationships merely to require 'interchanges [that] have some degree of mutuality, in the sense that the behaviour of each takes some account of the behaviour of the other'. We however are deliberately defining 'social connections' as a subset of social relationships.

Social connections might be of benefit immediately or prospectively and might involve support in many different forms (for instance, financial, practical, emotional or informational). We think that it helps to specify that social connections are not wholly contractual, to exclude certain types of social relationship from analysis—for instance, those that arise solely due to some economic arrangement, such as that between a teacher and their pupil, or between a wealthy householder and their maid, or even those relationships that are imposed upon one individual, such as between a probation officer and their client. By contrast, some social connections, such as of friendship, may be defined entirely by informal or 'private' normative expectations (e.g. Pahl 2000). Others, such as involving family members, are typically subject to a mix of formal institutional recognitions and accompanying informal norms. In our analyses, of the many ties that might constitute social connections, we generally focus on those that people actively choose or maintain (marriage and friendship), and we generally differentiate between those that involve family (defined as connections of marriage, cohabitation, descent or adoption) and those that involve friends and acquaintances that are not family—although even this boundary is sometimes fuzzy.

Social science enquiry has often involved the analysis of empirical patterns in 'social connections' in one form or another. For example, classical social mobility research using survey data focusses upon the social connection between parents and their children (e.g. Lipset and Bendix 1959; Glass 1954; Sorokin 1927). Classical anthropological studies looked at the importance and structure of social connections in the life experiences of people, as for instance in Bott's (1957) description of 'close knit' and 'loose knit' family and friendship networks. Interest in social connections is sustained to the present, and opportunities for accessing relevant data have expanded considerably in recent years (e.g. Treiman and Ganzeboom 2000).

In our examples, 'social interactions' refer to measured instances of social connections between people. Access to data about social connections of family is reasonably easy, because numerous linguistic terms and administrative traditions serve to record information on people's family relationships. Data on friendship links is more complicated. It is well known, for example, that if people are asked about friendships with others, it is often the case that one individual's views about who their friends are is not precisely reciprocated (e.g. Moreno 1953). Indeed, neither the frequency nor depth of interaction is a critical factor in defining social connections of friendship—some friends might communicate infrequently yet still be very important to each other; in some situations, some social connections, such as between neighbours, might involve regular, supportive contact, without necessarily having very much knowledge of their wider lives.

For pragmatic reasons, we focus analytically upon the codifications of social interactions as are recorded on social survey datasets. Household survey datasets will often collect data on multiple individuals from the same household and will also record the relationships between individuals within households in a standard manner (see e.g. Hoffmeyer-Zlotnik and Warner 2014 on definitions of the household). Likewise, questionnaire surveys in sociology often ask people to think of their 'closest friends' and then to provide further data about them. In such cases, the criteria by which social connections are defined could be disputed, but the definitions used by the surveys are reasonably transparent and well grounded.

Data on social interactions between people is most useful when we also have further information on relevant properties of the individuals involved—for instance, if we know the job category of both a survey respondent and her husband. Many datasets have these qualities. Extended records are often collected from multiple individuals who are linked by social connections, for example, in household survey designs. Some surveys also ask respondents to give relevant information about other specific people (for instance, the educational level of their nonresident parents, the occupation of the person they think of as their best friend). Some surveys also collect less specific information about the distribution of a respondent's social connections—for example, the 'position generator' tool asks people to indicate whether or not they have any friends who are in a specific nominated occupation, or some other nominated social position (e.g. Lin and Dumin 1986).

Increasingly, social scientists are able to access not just deliberately collected data, such as from social surveys, but also by-product data (often 'big data') that might provide similar information depending upon its context. Examples include social interactions fostered by and recorded through the internet (cf. Christakis and Fowler 2010); the use of administrative data that allows linkage of families over a long period of time (e.g. Platt 2005); and the use of administrative data that allows us to connect together information about different people who share the same workplace or who share membership of the same voluntary organisation (in this example, constituting data on people who may be presumed to be 'acquaintances', but who might not think of themselves as friends or family).

Generally, it is much easier to obtain empirical data on social interactions that involve family. In sociology, for example, there are an abundance of surveys with data on spouses and/or parents, but there are relatively few studies that provide data on a wider variety of friends and acquaintances. In our own work, we rely mainly on existing social survey datasets that either record information on multiple family members or that record data about survey respondents plus people that the respondents themselves have nominated as their 'friends'.

Data about social connections is critical to the measurement and analysis of 'social distance'. Social distance usually refers to a representation of the gap between influential social categories (rather than individuals), measured through the analysis of patterns of social connections between their incumbents. For example, the 'social distance' between occupational categories could be established by summarising the volumes of social interactions between people from different occupations (see Chaps. 4, 5, 6, 7, and 8), and the social distance between various other influential social categories (e.g. educational qualifications, ethnic groups, and religious groups) could similarly be explored (e.g. Laumann 1973; Chap. 10). Social distance is a valuable and exciting sociological concept because it allows us to characterise social positions in a way that is rooted in one of the most important aspects of an individual's life (i.e. their social connections). If the individuals within two social categories very rarely share (voluntary and supportive) social connections, those two categories are said to be socially distant, whereas if it is common for the incumbents of two categories to have social connections, then those categories have a low social distance. There is a long history of sociological research into the social distance between socio-economic and socio-demographic categories, for instance, in evaluating whether there is evidence that social distances between occupational and educational categories are increasing or decreasing over time (e.g. Smith et al. 2014). Conclusions vary, but many comparative studies suggest that social distances are either broadly stable over time or are gradually diminishing in contemporary societies—that is, it generally becomes slightly more common over time for individuals from different categories to interact socially (e.g. Lambert et al. 2014).

## 2.3 Social Reproduction, Social Stratification, and Social Inequality

'Social reproduction' refers in general terms to patterns of persistence through time in social structures—an enduring religious divide within a nation, for example, constitutes a pattern of social reproduction. However, we are most interested in social reproduction as it relates to patterns of 'social stratification' and 'social inequality'. These related terms refer to systematic patterns of differences between individuals in the distribution of consequential resources, such as economic assets and other aspects of welfare. 'Social inequalities' arise when these differences are in some sense linked to recognisable social divisions—examples might be systematic patterns of difference by gender, age, ethnicity, or occupational circumstances (e.g. Platt 2011). The concept of 'social stratification' is usually taken as referring to a distinctive social structure that is itself defined by inequalities in resources.

Many authors argue that 'social stratification' should intrinsically be thought of as concerning not just a particular state of inequalities in the distribution of resources, but the wider social systems that serve to reproduce and maintain those inequalities over time (e.g. Bottero 2005; Kerbo 2003). Accordingly, research on social stratification is often directed to how inequalities are reproduced through time (and the puzzle of why, across societies and over time, we see relatively little change in the unequal distribution of resources). For most sociologists, social stratification is substantially 'socially constructed', meaning that it doesn't necessarily have to take the form that it does, but the form that it takes, and that is maintained through time, is substantially a function of socially consensual behaviours and institutions.

Sociology offers no shortage of explanatory accounts for social reproduction. Amongst the most well known, from a Marxist framework, the economic interests of actors (and their interactions with technological change) are said to generate systems of unequal conflict that perpetuate stratification inequalities (e.g. Wright 1997; Braverman 1974). To postmodernists, the rise of the information age widens the gap between the resourced and the disadvantaged (e.g. Beck 2000). To neo-classical sociologists and economists, rational self-interest interacts with unequal access to opportunities to perpetuate social divisions, particularly through intergenerational transmissions from parents to their children (e.g. Breen and Goldthorpe 1997).

One tradition in sociology focusses upon how the social reproduction of social stratification inequalities is influenced by the apparently voluntary behaviour of individuals in forming and maintaining enduring social connections (e.g. Bottero 2005; Stewart et al. 1980). Bourdieu's analysis in particular, and corroborating evidence of the relationship between social stratification and lifestyle and leisure choices (e.g. Bennett et al. 2009; Bourdieu 1984), demonstrates how social choices can help reproduce inequalities regardless of intentions. As Bottero (2005, p. 256) summarises: 'People—simply by liking the things and people they like—cannot help but reproduce the stratification order, regardless of what they know or think about inequality.'

The tendency for social connections to foster social reproduction is obvious if we accept two points. First, that resources that are consequential to well-being are unequally distributed (for instance, economic assets such as income, wealth and home ownership, or non-economic advantages such as in power, influence, or knowledge); second, that individuals tend, in general, to try to help those that they have social connections with and that this includes drawing in part upon their own resources. Help might involve directly sharing assets and resources with others, but it might also involve taking steps to help others develop their resources further, such as sharing specialist knowledge. A grandmother donating funds towards a deposit on her grandson's mortgage might be a typical example of the former mechanism in the contemporary UK; a lecturer advising her friend on the best university courses for his nephew might be a typical example of the latter. Another example is sometimes seen in 'deferred gratification' in some of the most privileged occupations, such as traditional professions. Training and early career stages in these areas typically do not bring heightened economic reward, and indeed often require the incumbent to endure relatively difficult circumstances for some period of time. In this scenario, social reproduction in the composition of privileged positions is fostered because those individuals who enjoy more advantaged social connections can draw upon more support in helping them through early career privations (e.g. Devine 2004). These and many other examples represent situations whereby social connections themselves become a force for social reproduction of the stratification structure.

## 2.4 The 'Social Resin'

We use the metaphor of the 'social resin' to refer to one major way in which social connections serve to reproduce and sustain social stratification and social inequality: through their symbiotic relationship with occupations. Our idea is to think of social connections as a 'resin' that works to bind important structural units (occupations) together. Key to this metaphor is the circular relationship between the social structure of inequality and the voluntary social connections that people experience in their lives. Patterns in social connections, such as homophily and endogamy in friendship and marriage, act as an agent which binds together the components of society in a reasonably tight form with limited malleability. These social connections are a 'resin' because they are not completely fixed and inflexible, but in general they are quite firmly anchored, and, once established, they contribute to defining the social inequality structure in terms of the distribution of consequential resources. Occupations are by no means the exclusive axes of social inequalities, but they are particularly important units in the distribution of resources and as markers of long-term social positions.

Our opening chapter asked you to reflect upon your own circumstances and consider how frequently other individuals within your family and friendship networks were in similar structural positions. Of course, some individuals are unusual and have diverse social connections with people who are not particularly similar to themselves. However, for most of us, most of the people with whom we have social connections are located in rather similar situations to our own—our connections usually have the same or similar jobs, with similar educational backgrounds, and similar levels of relative advantage or disadvantage in ownership of consequential economic resources. To understand why we say that this social pattern reflects a 'social resin', try now to think of this linkage (between your own structural circumstances and your friends and connections) as if it were directly causal: what if it was the circumstances of your friends and connections (including your family) that led to your own structural circumstances? For the majority of us (albeit with individual exceptions), the hypothetical structural circumstances we would be in, if we were pushed into that position by our social connections, will be a very close model to the structural positions that we actually maintain.

Our model does not portray the link between social connections and occupations as fixed and immutable. In some ways, indeed, the true significance of the 'social resin' is in the limitations of its power: we see it as a persistent but not deterministic influence upon the outcomes that individuals experience. Its malleability may well be a property that makes it socially acceptable across contemporary societies: popular contemporary ideologies of 'meritocracy' and of individual 'freedom' do not sit well with a tightly deterministic structure, but arguably they fit very well indeed with a system that supports a moderate level of social reproduction that matches quite well to individual preferences over social connections (cf. Roemer 2012).

In Fig. 2.1, we summarise the average national-level patterns of similarities between different individuals in the UK who share either social connections of household membership or occupations. The figure shows what proportion of variation amongst people's responses is common to these two levels (these values can also be interpreted as the average correlation between two different people within the same units). As a sociological pattern, the associations shown are quite substantial (and all are statistically significant). We would read this evidence as dramatic: both occupations and households have a substantial connection to a wide range of individual outcomes—from your economic circumstances, through your sources of news, to your zest for life. It is equally true

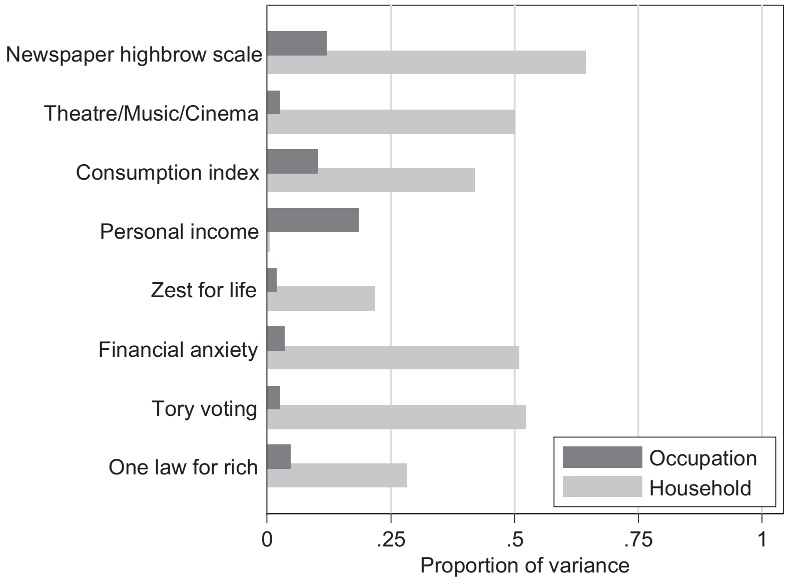



**Fig. 2.1** Relationship between selected aspects of individual and household circumstances amongst people from the same household and from the same occupation. Source: BHPS data (University of Essex 2010) using cross-sectional records from selected years subject to availability of indicators. Figure shows the intra-cluster correlation value (proportion of variance at the household/occupational level), which can be interpreted as the average correlation in the outcome between any two individuals from the same household or occupation

however that there is also a substantial 'lack of fit' shown in Fig. 2.1, whereby other things, unmeasured in this comparison, influence much of the variation. This tells us that whilst social connections and occupations are important, neither are dominant influences upon individual experiences within contemporary societies.

## 2.5 Explaining the 'Social Resin'

We highlight two arguments on why the 'social resin' connects occupations and social connections and leads to persistent inequality structures. The first concerns the roles of 'social capital' and 'cultural capital' in shaping occupational and social circumstances. A second explanation is the proposition that actors generally have a ‘preference’ for stability, and that this is a root cause of behaviours that ultimately lead to the integration of social connections and occupational structure and the persistent nature of contemporary inequality structures.

### 2.5.1 Social and Cultural Capital as Driving Forces of the Social Resin

‘Social capital’ is widely understood to refer to the stock of information resources and support that individuals can accrue from their social connections (e.g. Li 2015). Social capital can constitute advice, guidance, and knowledge that can aid in many domains of social life, such as providing help in obtaining employment positions (e.g. Granovetter 1973), the ability to access institutional information or complain effectively (Putnam et al. 1993), or the ability to engage effectively with influential professionals such as doctors (Abel 2008). Crucially, the stock of social capital that is available to different people is itself unequal; on average a person’s social capital is more favourable when they enjoy a more favourable position within the socio-economic inequality structure (e.g. Li et al. 2003).

Given the unequal distribution of social capital, it follows naturally that in many situations, the exploitation of social capital serves to retain and reinforce existing structural inequalities. For example, formal memberships of clubs, societies and associations that influence an individual’s social capital tend also to be socially structured (for instance, defined by social stratification circumstances as well as socio-demographic factors such as of ethnicity, gender and age). With regard to the link between social connections and occupational positions that we have labelled the ‘social resin’, there are obvious mechanisms whereby knowledge, information resources, expectations, and orientations can be shaped by resources of social capital that themselves are unequally distributed (e.g. Devine 2004; Willis 1977).

Research perspectives associated with Bourdieu highlight the integrated relevance of both social and cultural capital in influencing an individual’s outcomes. ‘Cultural capital’ refers broadly to the fluency that individuals have in engaging with social codes and cultural forms. Studies show that it can be influential as a marker of less clearly defined ‘soft skills', personality attributes, and capacities that also have considerable influence upon socio-economic outcomes (e.g. Zimdars et al. 2009). To Bourdieu, differential stocks of social and cultural capital tend to have the effect of perpetuating and exacerbating existing social inequalities. Influenced by this perspective, there is considerable contemporary sociological attention to mechanisms that involve both social and cultural capital in defining and reinforcing social inequality structures (e.g. Savage et al. 2015).

Links between occupations, and social and cultural capital, need not work exclusively to reproduce inequalities—there are also mechanisms when social or cultural capital linked to occupations might promote changes to positions. Granovetter (1973) provided a famous example of the benefit of 'weak ties' in obtaining new positions (our 'weak ties', those with whom we have some social connection without their necessarily being a regular or important part of our lives, can often be disproportionately influential in facilitating our access to new opportunities or important resources, such as in finding new jobs). Granovetter's example is one scenario where social connections can act to support social change rather than stability. In another example, the situational proximity of occupations can sometimes help to distribute resources, through social and cultural capital, that might not otherwise have been shared. Doctors' secretaries, for instance, might have regular communications with colleagues that allow them to tap into valuable resources that other workers, of otherwise similar circumstances, may not benefit from. These examples are still consistent with our metaphor of the 'social resin': our expectation would be that the organisation of social connections as it links to occupations helps to foster stocks of social and cultural capital that drive people apart and together in ways that reflect a largely stable, but slightly malleable, social inequality structure.

### 2.5.2 A Model of Deep-Rooted Preferences for Stability

Studies of friendship and partnership formation reveal that time and again, across societies, individuals mostly attain familiarity and repetition in their social connections (e.g. McPherson et al. 2001). Of course, individuals sometimes adapt to others *after* the formation of a social connection, but a common empirical behaviour is the selection of social connections based upon perceived similarity of circumstances (e.g. Kalmijn 1998). One plausible inference from observing homophily and endogamy is that, in general, individuals have relatively conservative tastes, a preference for stability and reproduction in 'making our way through the world' (cf. Archer 2007), which might even be psychological in origin as well as sociological. This perspective on patterns of social structure is very close to that advocated by Rytina (2010), who describes the social organisation of social relations and occupational outcomes as essentially 'sticky', suggesting that individuals have the potential to move around across the range of the social structure, but tend more usually to stay close to their origins and background. Of course, this portrayal doesn't apply to everybody, nor all of the time, but in general, empirical data is consistent with the theoretical model that individuals have deep-rooted preferences for stability and social reproduction.

The question of why individuals often behave as if they prefer stability is not one that we can fully answer, but there are many reasons to suggest that a preference for stability is a common human trait. Within sociology, a social exclusionist framework would suggest that those with wealth and resources become greedy and motivated to maintain and expand their advantages, whilst those less fortunate adjust their aspirations to not just accept, but value and express a preference for, their lesser allocation (e.g. Sayer 2016; McKenzie 2015; Friedman 2014). Separately, an orientation from neo-classical empirical sociology might highlight that stability is frequently the action most likely to maximise an individual's utility, and is hence often a rational preference (e.g. Goldthorpe 2007, c7). In biosocial sciences, some writers interested in gene polymorphisms argue that social choices have strong genetic antecedents that incline people towards dispositions that will, in terms of social structure, tend to be conservative (e.g. Blum et al. 2012).

Preferences for stability might also be evident in interpretations of subjective evaluations and political orientations. Recognising the complexities of modern societies, some authors argue that an important driving force in our orientations and preferences is simply how we develop our understanding of 'the way things should work' amidst a complex of different possibilities: we most probably think about how similar things have worked in the past, and/or how similar things have worked amongst our social contacts (e.g. Archer 2007). In parallel, it is well established in social psychology that individuals often take subjective comfort from stability and familiarity and often react with anxiety and fear towards the prospect of change (e.g. Burchell 1994)—whilst most people recognise that young children, specifically, are generally more comfortable with stability, it is plausible that all people, on average, have a similar orientation! Indeed, in an earlier generation of research, sociologists established how subjective well-being is strongly influenced by localised relative evaluations within 'reference groups' (e.g. Hyman 1967), and that individual satisfaction was often derived from a sense of stability within context. In a different example, when we think of political actions and expressions of preference that individuals make, it is striking how frequently these coalesce around a preference for stability or a resistance to change—consider, for example, how often local pressure groups are organised to resist prospective changes. We are not in a position to adjudicate between these and other explanations for stability preferences in human behaviours, but it is clear that social connections patterns do occur empirically in a way that could be consistent with them having being driven by the preference for stability.

The arguments above suggest that there may be important forces that shape life choices and social connections that are largely conservative in character. However, the occurrence of conservative patterns in social behaviours implies neither that conservatism is good nor that conservatism is overwhelming. On the latter, the point to reiterate is that average patterns are consistent with the hypothesis that people generally favour stability; nevertheless, this does not mean that all people do, nor that things never change. The observation that many individuals appear content with stability says nothing about the long-term costs or benefits from stability in circumstances. At the aggregate level, diverse and flexible societies seem to perform more favourably than socially conservative ones (e.g. Wilkinson and Pickett 2009), and there is a wealth of literature on social interactions that highlights the benefits that can accrue to individuals from having diverse and non-conservative patterns. As a quite different example, environmental challenges represent a major social priority around which conservative and stable behaviours are unlikely to provide a desirable response (e.g. Sayer 2016).

Views on intergenerational 'social mobility' make another interesting example relevant to the hypothesis of stability preferences. In many societies, 'social mobility' (the process whereby adults move into different economic circumstances to those of their parents) is a common term of political aspiration and is widely portrayed as a desirable thing, typically by linking it to support for 'meritocracy' (the process whereby economic attainment is achieved through 'merit' rather than due to parental or family support) (e.g. Payne 2012, 2017). Such pervasive attitudes might, in principle, lead to societies of rapidly changing social structure, but of course this is not the case. The underlying structures, hierarchies, and institutions of societies prove stable over many years, and across societies moderate intergenerational social reproduction is persistently observed (parent-child correlations in economic circumstances are widely estimated as around 0.3–0.6, e.g. Bernardi and Ballarino 2016, Breen 2004, Corak 2004; some authors, e.g. Clark 2014, argue that these correlation values dramatically underestimate long-term patterns of social reproduction within families). Despite expressed preferences for more social mobility, therefore, empirical patterns are more consistent with the model that children, and their parents, often aspire to social reproduction (for instance, a father hoping that his son will take over the family business). Aspirations of this nature have often been shown even when intergenerational reproduction would not be to a person's objective advantage (e.g. McKenzie 2015; Willis 1977).

Regarding social mobility, there is clearly a disconnection between high levels of ideological support for social mobility and its empirical realisation. A plausible explanation is offered by, amongst others, Saunders (2010). Saunders argues that most people, politicians included, who proffer support for social mobility do not really understand this concept in its sociological sense, but really associate it with more nebulous ideas of being treated 'fairly' and to having the capacity to 'do well' and/or to not 'loose out'. Social stability, under this account, might actually be facilitated if individuals are motivated to act to preserve their current situation or strive for small but realistic improvements on it (in the UK, the expression 'keeping up with the Jones's' refers to the urge to maintain the appearance of at least parity of circumstances with our nearest neighbours). Whilst on the face of it, widespread support for social mobility directly challenges any claim for preferences for stability and reproduction, on closer inspection we see that the ways in which people understand and realise the transmission of circumstances that is relevant to social mobility is probably closer to a conservative model than to a transformative one.

In summary, we have suggested that two mechanisms could lie behind the observed empirical properties of the 'social resin'. Conventional accounts of social capital offer one plausible explanation, but the same empirical patterns seem also to be consistent with a generic hypothesis that individuals tend generally to favour social stability. Over Chaps. 4, 5, 6, 7, 8, 9, 10, and 11, we present an empirical account of ways that the 'social resin' links occupations and social connections. This account expands upon—but is not able to comprehensively adjudicate between—these two mechanisms.

## Bibliography

Abel, T. (2008). Cultural Capital and Social Inequality in Health. *Journal of Epidemiology and Community Health, 62*(7), e13.

Archer, M. S. (2007). *Making Our Way Through the World: Human Reflexivity and Social Mobility*. Cambridge: Cambridge University Press.

Beck, U. (2000). *The Brave New World of Work*. Cambridge: Polity Press.

Bennett, T., Savage, M., Silva, E. B., Warde, A., Gayo-Cal, M., & Wright, D. (2009). *Culture, Class, Distinction*. London: Routledge.

Bernardi, F., & Ballarino, G. (Eds.). (2016). *Education, Occupation and Social Origin: A Comparative Analysis of the Transmission of Socio-Economic Inequalities*. Cheltenham: Elgar.

Blum, K., Oscar-Berman, M., Giordano, J., Downs, B. W., Simpatico, T., Han, D., & Femino, J. (2012). Neurogenetic Impairments of Brain Reward Circuitry Links to Reward Deficiency Syndrome (RDS): Potential Nutrigenomic Induced Dopaminergic Activation. *Journal of Genetic Syndromes and Gene Therapy, 3*(4), 1000e1115.

Bott, E. (1957). *Family and Social Network*. London: Tavistock Publications.

Bottero, W. (2005). *Stratification: Social Division and Inequality*. London: Routledge.

Bourdieu, P. (1984). *Distinction*. London: Routledge and Kegan Paul.

Braverman, H. (1974). *Labour and Monopoly Capital*. London: Monthly Review Press.

Breen, R. (Ed.). (2004). *Social Mobility in Europe*. Oxford: Oxford University Press.

Breen, R., & Goldthorpe, J. H. (1997). Explaining Educational Differentials: Towards a Formal Rational Action Theory. *Rationality and Society, 9*(3), 275–305.

Brynin, M., & Ermisch, J. (Eds.). (2008). *Changing Relationships*. London: Routledge.

Brynin, M., Martinez Perez, A., & Longhi, S. (2008). How Close are Couples? In M. Brynin & J. Ermisch (Eds.), *Changing Relationships* (pp. 93–108). London: Routledge.

Burchell, B. (1994). The Effects of Labour Market Position, Job Insecurity and Unemployment on Psychological Health. In D. Gallie, C. Marsh, & C. Vogler (Eds.), *Social Change and the Experience of Unemployment* (pp. 188–212). Oxford: Oxford University Press.

Christakis, N., & Fowler, J. (2010). *Connected: The Amazing Power of Social Networks and How They Shape Our Lives*. London: Harper Press.

Clark, G. (2014). *The Son Also Rises: Surnames and the History of Social Mobility*. Princeton, NJ: Princeton University Press.

Corak, M. (Ed.). (2004). *Generational Income Mobility in North America and Europe*. Cambridge: Cambridge University Press.

Curington, C. V., Lin, K.-H., & Lunquist, J. H. (2015). Treatment of Multiracial Daters in an Online Dating Website. *American Journal of Sociology, 80*(4), 764–788.

Devine, F. (2004). *Class Practices: How Parents Help Their Children Get Good Jobs*. Cambridge: Cambridge University Press.

Friedman, S. (2014). The Price of the Ticket: Rethinking the Experience of Social Mobility. *Sociology, 48*(2), 352–368.

Glass, D. V. (1954). *Social Mobility in Britain*. London: Routledge & Kegan Paul.

Goldthorpe, J. H. (2007). *On Sociology, Second Edition: Volume One—Critique and Programme*. Stanford, CA: Stanford University Press.

Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology, 78*(6), 1360–1380.

Hinde, R. A. (1997). *Relationships: A Dialectical Perspective*. Hove: Psychology Press.
Hoffmeyer-Zlotnik, J. H. P., & Warner, U. (2014). *Harmonising Demographic and Socio-Economic Variables for Cross-National Comparative Survey Research*. Berlin: Springer.
Hyman, H. H. (1967). The Value Systems of Different Classes: A Social Psychological Contribution to the Analysis of Stratification. In R. Bendix & S. M. Lipset (Eds.), *Class, Status and Power: Social Stratification in Comparative Perspective* (2nd ed., pp. 488–499). London: Routlege and Kegan Paul.
Kalmijn, M. (1998). Intermarriage and Homogamy: Causes, Patterns, Trends. *Annual Review of Sociology, 24*, 395–421.
Kalmijn, M., & Flap, H. (2001). Assortative Meeting and Mating: Unintended Consequences of Organized Settings for Partner Choices. *Social Forces, 79*(4), 1289–1312.
Kerbo, H. R. (2003). *Social Stratification and Inequality: Class Conflict in Historical, Comparative and Global Perspective* (5th ed.). London: McGraw-Hill.
Lambert, P. S., Griffiths, D., Bihagen, E., Zijdeman, R. L., Gayle, V., & Tranmer, M. (2014). *Is Britain Pulling Apart? Evidence from the Analysis of Social Distance.* Stirling: Technical Paper of the 'Is Britain Pulling Apart' Research Project, School of Applied Social Science, University of Stirling. Retrieved from www.camsis.stir.ac.uk/pullingapart.
Lampard, R. (1997). Party Political Homogamy in the United Kingdom. *European Sociological Review, 13*(*1*), 79–99.
Laumann, E. O. (1973). *Bonds of Pluralism: The Form and Substance of Urban Social Networks*. New York: Wiley.
Li, Y. (Ed.). (2015). *Handbook of Research Methods and Applications in Social Capital*. Cheltenham: Edward Elgar.
Li, Y., Savage, M., & Pickles, A. (2003). Social Capital and Social Exclusion in England and Wales (1972–1999). *British Journal of Sociology, 54*(4), 497–426.
Lin, N., & Dumin, M. (1986). Access to Occupations Through Social Ties. *Social Networks, 8*, 365–385.
Lipset, S. M., & Bendix, R. (1959). *Social Mobility in Industrial Society*. Berkeley, CA: University of California Press.
Lizardo, O. (2006). How Cultural Tastes Shape Personal Networks. *American Sociological Review, 71*(5), 778–807.
Mare, R. D. (1991). Five Decades of Educational Assortative Mating. *American Sociological Review, 56*, 15–32.

McKenzie, L. (2015). *Getting By: Estates, Class and Culture in Austerity Britain*. Bristol: Policy Press.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology, 27*, 415–444.

Moreno, J. L. (1953). *Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Sociodrama*. New York: Beacon House.

Pahl, R. (2000). *On Friendship*. Cambridge: Polity Press.

Payne, G. (2012). Labouring Under a Misapprehension? Politicians' Perceptions and the Realities of Structural Social Mobility in Britain, 1995–2010. In P. S. Lambert, R. Connelly, R. M. Blackburn, & V. Gayle (Eds.), *Social Stratification: Trends and Processes* (pp. 223–242). Aldershot: Ashgate.

Payne, G. (2017). *The New Social Mobility: How the Politicians Got it Wrong*. Bristol: Policy Press.

Pearce, A., & Gambrell, D. (2016). *Bloomberg: This Chart Shows Who Marries CEOS, Doctors, Chefs and Janitors*. Retrieved June 1, 2016, from http://www.bloomberg.com/graphics/2016-who-marries-whom/.

Penn, R. D., & Lambert, P. S. (2009). *Children of International Migrants in Europe: Comparative Perspectives*. Basingstoke: Palgrave.

Platt, L. (2005). The Intergenerational Social Mobility of Minority Ethnic Groups. *Sociology, 39*(3), 445–462.

Platt, L. (2011). *Understanding Inequalities: Stratification and Difference*. Cambridge: Polity.

Putnam, R. D., Leonardi, R., & Nanetti, R. Y. (1993). *Making Democracies Work: Civic Traditions in Modern Italy*. Princeton, NY: Princeton University Press.

Roemer, J. (2012). What Is the Justification for Studying Intergenerational Mobility of Socioeconomic Status? In J. Ermisch, M. Jantti, & T. Smeeding (Eds.), *From Parents to Children: The Intergenerational Transmission of Advantage* (pp. 482–487). New York: Russell Sage Foundation.

Rytina, S. (2010). *Sticky Struggles: Stratification as Emergent from Network Infrastructure*. Presentation to Centre on Skills, Knowledge and Organisational Performance (SKOPE), University of Oxford.

Saunders, P. (2010). *Social Mobility Myths*. London: CIVITAS: Institute for the Study of Civil Society.

Savage, M., Cunningham, N., Devine, F., Friedman, S., Laurison, D., McKenzie, L., Miles, A., Snee, H., & Wakeling, P. (2015). *Social Class in the 21st Century*. London: Pelican Books.

Sayer, A. (2016). *Why We Can't Afford the Rich*. Bristol: Policy Press.

Skvoretz, J. (2013). Diversity, Integration, and Social Ties: Attraction Versus Repulsion as Drivers of Intra- and Intergroup Relations. *American Journal of Sociology, 119*(2), 486–517.

Smith, J. A., McPherson, M., & Smith-Lovin, L. (2014). Social Distance in the United States: Sex, Race, Religion, Age and Education Homophily Among Confidants, 1985 to 2004. *American Sociological Reiwew, 79*(3), 432–456.

Sorokin, P. A. (1927). *Social Mobility*. New York: Harper and Brothers.

Stewart, A., Prandy, K., & Blackburn, R. M. (1980). *Social Stratification and Occupations*. London: Macmillan.

Stutzer, A., & Frey, B. S. (2006). Does Marriage Make People Happy, or do Happy People Get Married? *The Journal of Socio-Economics, 35*, 326–347.

Treiman, D. J., & Ganzeboom, H. B. G. (2000). The Fourth Generation of Comparative Stratification Research. In S. R. Quah & A. Sales (Eds.), *The International Handbook of Sociology* (pp. 123–150). London: Sage.

University of Essex, & Institute for Social and Economic Research. (2010, July). *British Household Panel Survey: Waves 1–18, 1991–2009 [Computer File]* (7th ed.). Colchester, Essex: UK Data Archive [distributor], SN: 5151.

Upright, C. B. (2004). Social Capital and Cultural Participation: Spousal Influences on Attendance at Arts Events. *Poetics, 32*(2), 129–143.

Wilkinson, R. G., & Pickett, K. E. (2009). *The Spirit Level: Why More Equal Societies Almost Always do Better*. London: Allen Lane.

Willis, P. (1977). *Learning to Labour: How Working Class Kids Get Working Class Jobs*. Westmead, Hants: Saxon House.

Wright, E. O. (1997). *Class Counts: Comparative Studies in Class Analysis*. Cambridge: Cambridge University Press.

Xu, X., Ji, J., & Tung, Y.-Y. (2000). Social and Political Assortative Mating in Urban China. *Journal of Family Issues, 21*(1), 47–77.

Zimdars, A., Sullivan, A., & Heath, A. (2009). Elite Higher Education Admissions in the Arts and Sciences: Is Cultural Capital the Key? *Sociology, 43*(4), 648–666.

# 3

# Measures of Social Stratification

## 3.1 Introduction

At this point we describe some of the most influential measures of social stratification in current use, particularly those based upon occupations. We refer to 'social stratification' as a structure of social inequality that is defined in terms of the distribution of, or access to, consequential resources (cf. Therborn 2013, c4; Platt 2011, c1; Blackburn 2008). In the UK in particular, many writers use 'social class' and 'social stratification' interchangeably. However, to us 'social stratification' is a more generic concept that refers to a structure of inequality which could take any mathematical form, whereas a structure of 'social class' necessarily disaggregates the members of a society into a number of distinct positions—that is, classes (e.g. Grusky and Weeden 2006).



P. Lambert, D. Griffiths, *Social Inequalities and Occupational Stratification*,
https://doi.org/10.1057/978-1-137-02253-0_3

## 3.2 Alternative Measures of Social Stratification

Social scientists have many different ways of measuring social stratification. Aggregate-level social indicators—for example, Gini coefficients, which provide a standardised summary of the overall level of inequality in a measure such as income—are often used to characterise differences in social inequality between societies as a whole (e.g. Atkinson 2005). Hereafter, however, we concentrate upon measures that seek to represent the position within the stratification structure of individuals at the micro-level.

Most measures at the micro-level use one of two functional forms. Social class schemes usually divide the population into a small number of categories, whereby the boundaries between different social classes should be reasonably clearly identified and be relatively consequential to individuals (e.g. Rose and Harrison 2010; Wright 2005). As an alternative to the multinomial cartography of a social class measure, social stratification is also commonly measured through a gradational scheme, in which individuals are located at points along a one-dimensional scale (e.g. Prandy and Blackburn 1997). It is an open and much disputed question, therefore, whether the social stratification inequalities of a particular society align more appropriately with the categories of a social class scheme or the gradations of a stratification scale.[1] Neither functional form is likely to provide the perfect representation, although gradational scales have some pragmatic conveniences, such as being easy to incorporate in a range of analytical methods, and being suited to accessible devices of communication, such as simple graphics.

Sociologists have traditionally favoured occupation-based measures of stratification position at the micro-level, but many different occupation-based measures are available (e.g. Lambert and Bihagen 2014), and many non-occupational indicators—for example, based on income, wealth, home-ownership, lifestyle, locality characteristics, or combinations of such factors—can also be considered (e.g. Hoffmeyer-Zlotnik and Warner 2014; Shaw et al. 2007). Indeed, inferring from recent national social science literatures, we have previously estimated that there could well be tens of thousands of alternative measures of social stratification position in currency in the contemporary social sciences (Lambert and Bihagen 2014). The vast volume of alternative measurement approaches is rarely acknowledged in applied research or in formal methodological discussions, which tend instead to select and work with only a single measure, or else compare only a narrow range of related alternatives (cf. Connelly et al. 2016; Wright 2005; Ganzeboom and Treiman 2003).

Why are there so many different measures? At a pragmatic level, the profusion arises because there are many different permutations in the construction of measures, and these can combine multiplicatively. It is useful, for example, to distinguish between:

1. Different 'referents' behind a measure (e.g. occupation, income, asset ownership)
2. Different ways of measuring different referents (e.g. using detailed or broad-brush data on occupations)
3. Different units of allocation (e.g. individuals, households, life courses)
4. Different recommendations over how to construct measures for a given referent (e.g. Wright 2005 for different views on measures based upon occupations)
5. Different realisations of recommended measures (e.g. using a measure in its full detail or in an abridged format)
6. Adaptations made for the local context in terms of nation or time point (such as different occupational taxonomies between countries or over time as organised by national statistics institutes)

In summary, the permutations across issues (1) to (6) mean that the total number of alternative measures of stratification might be very large indeed. Applied research studies will usually provide an account of the choices made between these options, but seldom try out and compare different plausible alternatives (cf. Lambert and Bihagen 2014).

With so many different measures available to researchers, it is not surprising to find that some degree of competition between them emerges. Sometimes writers advocate one measure in favour of others by arguing that it is theoretically more compelling—see, for instance, the advocacy of the Goldthorpe class scheme and its related measures by, amongst others, Rose and Harrison (2007, 2010). Separately (and often in combination), methodologists might argue that one measure is empirically stronger than alternatives (in the relevant context). Guveli (2006), Oesch (2006), and Hauser and Warren (1997) have advocated three respective occupation-based measures which have particularly favourable empirical characteristics, such as strong correlations to things that they are supposed to be related to.

## 3.3 Selected Influential Approaches to Measuring Social Stratification

### 3.3.1 Selected Measures of Social Stratification

Table 3.1 summarises selected measures of stratification position in a schematic way. We pick out a number of measures that are commonly used in contemporary sociology, social geography and population health research. Five of the measures listed in Table 3.1 are based upon occupations, but four are based upon other 'referents'.

The first two measures in Table 3.1 are categorical measures of social class based upon occupation. Both measures divide the occupational structure into a relatively small number of categories.

The 'Goldthorpe' class scheme (Table 3.1, row 1) refers to an influential measure developed by Goldthorpe and his colleagues (e.g. Erikson et al. 1979; Erikson and Goldthorpe 1992; Goldthorpe 1997). There are in fact many different variants to and relatives of the Goldthorpe scheme. The version that we describe, as used in the influential social mobility research programme summarised by Erikson and Goldthorpe (1992), is often used in a more detailed format with 11 categories, and can also be operationalised in versions with fewer than seven categories (e.g. Erikson and Goldthorpe 1992, pp. 38–39). The 'ESeC' and 'NSSEC' schemes ('European Socio-economic Classification' and 'National Statistics Socio-economic Classification', see Rose and Harrison 2010; Rose and Pevalin 2003) are widely regarded as appropriate contemporary versions of the scheme, and there are also several influential but 'unofficial' operationalisations, such as those published by Ganzeboom and colleagues (Ganzeboom 2016; Ganzeboom and Treiman 2003) and Leiulfsrud et al. (2005), and the deliberate modifications to the scheme that some writers have advocated (e.g. Guveli 2006).

**Table 3.1** Selected influential measures of social stratification position

| Functional form | Referent | Underlying concept of inequality | Correlation * 100 to: education; smoking; gender |
|---|---|---|---|
| 1. Goldthorpe social class (e.g. EGP-7 version; see Erikson and Goldthorpe 1992) | | | |
| Seven categories | Occupations | Employment relations and conditions | 26; 9; 21 |
| 2. Skill-based social class (e.g. UK Registrar General's Social Class scheme; see Szreter 1984) | | | |
| Five categories | Occupations | Employment skills | 31; 8; 7 |
| 3. Social interaction distance scale (e.g. CAMSIS scale for Britain; see Prandy and Lambert 2003) | | | |
| Scale | Occupations | Social interaction patterns as indicators of social reproduction | 50; 16; 5 |
| 4. Prestige scale (e.g. SIOPS; see Ganzeboom and Treiman 2003) | | | |
| Scale | Occupations | Respect from others and symbolic capital | 49; 12; 8 |
| 5. Socio-economic status scale (e.g. ISEI; see Ganzeboom and Treiman 2003) | | | |
| Scale | Occupations | Employment advantage and human and economic capital | 49; 15; 4 |
| 6. Local area-level deprivation quintile (e.g. SIMD (Scottish Index of Multiple Deprivation); see Scottish Government 2016) | | | |
| Five ordered categories | Local area characteristics | Neighbourhood characteristics | 16; 12; 1 |
| 7. Income (e.g. net personal income from all sources) | | | |
| Scale (log of income) | Personal income | Income levels (economic capital) | 36; 2; 26 |
| 8. Relative poverty (e.g. less than 50% of median disposable income; see OECD 2008) | | | |
| Dichotomous (two categories) | Household disposable income | Deprivation (economic capital) | 27; 1; 10 |
| 9. GBCS (based upon matching personal characteristics to criteria listed in Savage et al. 2013) | | | |
| Seven categories | Lifestyle, income, education, friends' occupations | Conglomeration of economic, social, and cultural circumstances | 28; 11; 8 |

Notes: Statistics based upon operationalisation of measure within UK BHPS for 2005, using sampling weights, *N* ~= 13,000, except for SIMD, which is based on analysis of Scottish Household Survey 2012, *N* ~= 10,600

The Goldthorpe scheme and its many variants are characterised by dividing occupational positions based upon their profiles of 'employment relations and conditions' ('ERC'). These include divisions between contractual arrangements (such as between self-employed and employee positions) and divisions associated with the regulation of employment for employees. The Goldthorpe measure is associated with a Weberian theorisation of social stratification, in which the ERC of occupations should be highly influential in defining the dual market and work situations of their incumbents, shaping in turn their life chances and those of their families. Operationally, the categories of the scheme are not designed to be ranked on a single ordinal scale, but there are nevertheless underlying dimensions of hierarchy that differentiate most of the social classes in a consistent way. Many validity studies have been undertaken that demonstrate that the Goldthorpe scheme can indeed be regarded as a reliable indicator of inequalities in employment relations and conditions, and as a measure of social stratification that is of considerable empirical relevance across a wide range of application areas (e.g. Rose and Harrison 2010; Rose and Pevalin 2003; Evans and Mills 1998).

Table 3.1, row 2, highlights the UK's Registrar General's Social Class ('RGSC') scheme as an example of a social class measure that is based upon an ordered ranking from lowest to highest skill level. Whilst the RGSC is unique to UK data, skill-based schemes of a similar character are available in many countries and are widely used in social research (e.g. Elias and McKnight 2001); they are particularly popular in analytical traditions, such as health inequalities research, where there is a prior expectation that inequalities related to social stratification are ordinal. Skill level itself may be defined by agreed criteria (for instance, based upon expert evaluations of the work involved, and/or data on the qualifications held by people in the job), although in practice boundaries between skill categories have sometimes been influenced by criteria that are not unambiguously related to skills, such as social judgements of prestige, or by sectoral divisions such as between manual and non-manual work (cf. Szreter 1984). Some authors argue that skill-based measures have very favourable properties as stratification schemes (e.g. Tahlin 2007). However, skill-based social class classifications are sometimes criticised for lacking a consistent sociological theorisation and/or for having inconsistent procedures of operationalisation (e.g. Crompton 1998, Szreter 1984).

Four of the measures listed in Table 3.1 define a one-dimensional gradational hierarchy of social inequality. Three of these are scales based upon occupations: scales of 'prestige' (e.g. Treiman 1977; see Table 3.1, row 4), scales of 'socio-economic status' (e.g. Ganzeboom et al. 1992; Table 3.1, row 5), and scales based on social interaction distance (e.g. Prandy 1990; Chan 2010; Table 3.1, row 3). Each scale is operationalised by assigning a score to an occupation, then allocating that score to the individual incumbents of the occupation. Prestige scales such as SIOPS define their scores through public opinion surveys that ask respondents to provide their own rankings of the relative social standing of occupations—a highly influential sociological finding from Treiman's (1977) analysis was that, by and large, people from different societies ranked the same occupations in the same relative positions. Socio-economic status scales such as ISEI define scores for occupations through the statistical analysis of databases on occupations—in the case of ISEI, by calculating a weighted average of the income and educational advantages of the incumbents of occupations. Social interaction distance scales such as CAMSIS are also based on statistical analysis of data about occupations, in this case using information about the social interactions between the incumbents of occupations—we describe the CAMSIS measure in depth over Chaps. 4, 5, and 6.

The fourth gradational measure mentioned is labelled 'income' (Table 3.1, row 7). Income is commonly measured with a linear functional form but can also be adapted to various non-linear representations (including 'social classes' based upon income categories—e.g. Gornick and Jantti 2013). By mentioning 'income' at this point, we are really alluding to a wide range of different measurement options related to economic assets, such as income from different sources, expendable income, life-course income profile, or wealth. The example described in row 7 refers to the log of total personal monthly income from all sources, but many different measures based on income are used in applied research (cf. Jenkins 2011, part 1).

Three more measures feature in Table 3.1. The measure of 'Poverty' (Table 3.1, row 8) is a binary division between being 'in poverty' and not.

The analysis uses the common criteria that a household is in relative poverty if its monthly disposable income is less than half the national median, but numerous alternative ways of identifying those in poverty are also available, including measures based upon access to resources and assets (e.g. Gordon 2006). The measure of local area deprivation (Table 3.1, row 6) reflects the popularity in social geography, and also in market research and public sector research, of relatively fine-grained area-based characterisations premised upon social inequality. The example we give is the 'Scottish Index of Multiple Deprivation', which is derived by ranking over 6000 small areas (of typically 1000 residents) in Scotland, in terms of 'relative deprivation' based upon various data covering the income, employment, health, education, geographical, crime and housing profiles of residents (e.g. Scottish Government 2016). Finally, the last measure that we mention in Table 3.1 is, like the area-based index, calculated on the basis of data about a range of different aspects of individual lives. The 'Great British Class Survey' ('GBCS') stratification measure (Table 3.1, row 9) is a social class scheme proposed by Savage et al. (2013). A class may be allocated to individuals on the basis of applying a classificatory algorithm that takes account of data on their combined profile based on measures of income, housing, education, lifestyle, and social connections. In the UK, the GBCS has attracted popular, intuitive appeal as a measure that reflects multiple dimensions of social inequality, although some have criticised it for not having consistent theoretical or empirical properties (e.g. Mills 2014), and the scheme is not readily operationalised on secondary datasets (in Table 3.1, we summarise an approximate version, which we operationalised on secondary survey data and which used comparable but not identical criteria to those recommended by Savage et al.).

A typical visualisation of the relationship between two of the measures from Table 3.1 is presented in Fig. 3.1. The figure depicts the spread of CAMSIS scores given to a sample of adults, organised by their social class categories according to the 'ESeC' scheme (Rose and Harrison 2010, 2007; ESeC is itself a variant of the Goldthorpe class scheme). The survey data used in Fig. 3.1 is from contemporary Britain and from Sweden,[2] but in other societies, a similar result would be seen: CAMSIS scores are strongly related to ESeC categories, but within each category, there is considerable heterogeneity in the CAMSIS scores given to respondents.

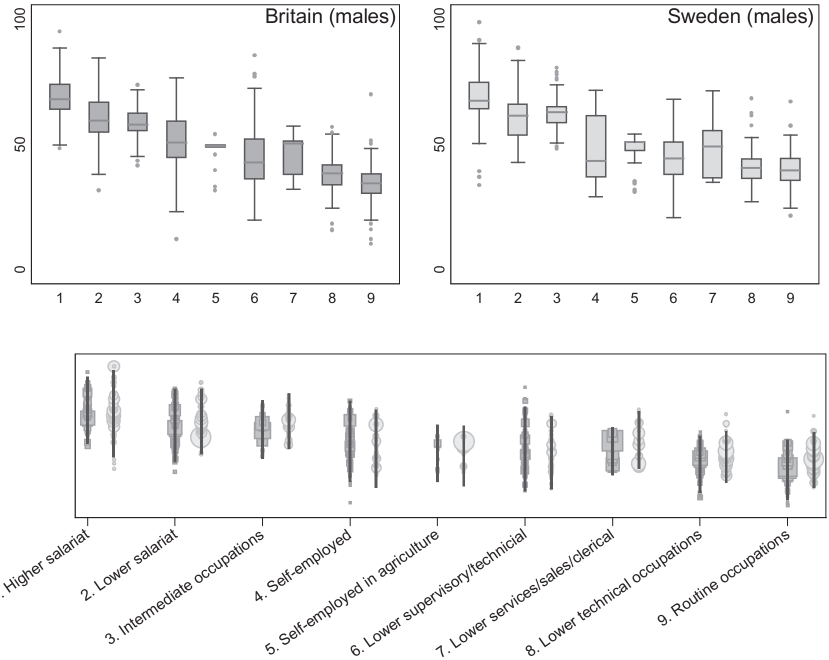



**Fig. 3.1** CAMSIS and ESEC distributions. Notes: Data for males from Britain (BHPS; see University of Essex 2010) and Sweden (Level of Living Survey, LNU; see http://www.sofi.su.se/) in 1991. The boxplots (upper panel) show the interquartile range (boxes) plus outliers. The lower panel shows actual values (weighted by number of cases) with lines showing the mean value plus or minus two standard deviations

Indeed, it is commonly the case that some categories include people with CAMSIS scores that are a long way away from their class category mean—visible, for instance, in the wide range of the boxes and whiskers in the upper panel and the spread of points and standard deviation values shown in the lower panel. Moreover, the lower panel also highlights the uneven clumping of occupations within the class categories. For example, the larger plotted points indicate that it is not uncommon for there to be quite large clusters of individuals (i.e. people working in populous occupations), who have a CAMSIS score which may be considerably lower or higher than the mean for that category.

Figure 3.1 helps to emphasise how different measures of social stratification assign individuals to categories or scores, but that in many cases the same individuals are assigned to different positions according to different measures. The disparities between the CAMSIS and ESEC measures that are shown in Fig. 3.1 raise interesting questions (and similar disparities would be seen when comparing other measures). We could conclude that the two measures genuinely capture different phenomena and that the disparities of placement should represent appropriate examples of people who are in 'contradictory' situations (see further in Sect. 3.3.2). However, it is also possible to regard the differences as substantially the product of measurement imperfections in one measure or the other—those individuals with CAMSIS scale scores that are much higher than the mean score for their occupational class, for instance, might be people whose situation is not well characterised by their allocated social class category (or people for whom their CAMSIS score inappropriately overestimates their social circumstances).

Returning to Table 3.1, we can consider the differences between measures for the same people further by looking at some information on the relative empirical qualities of the different measures when applied to samples of individuals. In general, we might expect to see a high correlation with education and with smoking behaviour (two aspects of contemporary life that are widely understood to link to position in the social stratification structure). Additionally, we might expect a low correlation with gender (an aspect of life that we might not presume should be strongly linked to stratification circumstances). As can be seen in Table 3.1, all the measures perform in a broadly comparable way, but some have slightly stronger or weaker correlations than others. Those measures that have somewhat higher correlations with both education and smoking behaviour might—arguably—be seen as better indicators of the generic concept of position in the stratification structure. By contrast, measures that have a relatively large correlation with gender might—arguably—be seen as more problematic.

Some other features of the measures summarised in Table 3.1 are worth highlighting. Firstly, the measures vary considerably in the extent to which they make use of details in the underlying 'referent'. Some of the occupation-based schemes, for example, group many different occupations together into the same categories, but the occupation-based scales are arguably a little more sensitive, because they allow a wider range of different occupations to be assigned a different score within the scale (arguably, by this reasoning, scales are generally better at reflecting finer gradations in stratification structures; however, the nuances of occupational differences are still not perfectly reflected, because the scales force the occupations to be ranked on a single dimension of difference).

Secondly, our account hitherto has neglected other important options relevant to characterising an individual's position: whether to use data solely concerning the individual's own circumstances or to make use of data about their wider household or family (such as the occupation or asset holdings of a spouse or parent) and whether to limit data to information about current circumstances or to attempt to reflect a relevant longitudinal trajectory. There are good empirical and theoretical arguments for drawing upon information about the household or family (e.g. Harkness 2013; Rose and Harrison 2010, c10 and c13) and/or for using data about the career context (e.g. Abbott 2006; Miller 1998). Nevertheless, many studies use data that is strictly about the individual and their current circumstances, such as their current job or personal income. The 'individualist' paradigm is sometimes motivated by issues of data availability—it is usually easier to obtain and exploit data about current circumstances than about other household members and/or about longitudinal trajectories. Indeed, most of our analyses below are undertaken at the individual level for similar pragmatic reasons. Some social scientists also adopt the individualist approach as an issue of ideology—it might be argued to be inappropriate to characterise current circumstances based on trajectories, and it is sometimes seen as sexist or ageist to characterise the circumstances of one person on the basis of information about a spouse or parent.

Another consequential operational choice when measuring stratification position concerns the extent to which gender is incorporated into the measurement and analysis. Table 3.1 highlights quite moderate bivariate correlations between gender and stratification measures. It is not straightforward to respond to these (e.g. Crompton and Mann 1986). Their origins lie in gender differences in the average profiles of the underlying referents that can be used to measure social stratification position.

Occupation-based measures, for example, are complicated by gender segregation in occupations; income-based measures are influenced by average differences in working hours and career continuity between men and women; measures that draw upon educational experiences are influenced by different average educational outcomes and field of study preferences of men and women; even measures that draw upon household-level characteristics can be perturbed by average demographic differences between men and women (since women tend to marry slightly younger than men, and live longer, even the average household situation of women and men is slightly different). A common response to the depth of social differences between men and women in the economics tradition is to undertake analyses separately for the two groups. In other social sciences, it is more conventional to analyse cross-gender populations, but this often raises risks of spurious interpretations, because it may be difficult to disentangle differences in experience in the stratification structure, from differences that might be related to other gendered processes. If nothing else, the correlations with gender that are shown in Table 3.1 should remind us to pay careful attention to gender when using measures of stratification position.

### 3.3.2 Multidimensionality in Measures of Stratification

Each measure listed in Table 3.1 is designed around a slightly different conceptualisation of the important features of social inequality (as summarised in the 'concepts' column of Table 3.1). In principle, this suggests that comparisons between measures can allow us to compare the theoretical processes that are related to inequality. It would be interesting to know, for example, whether health inequalities are more strongly linked to symbolic recognition than to employment relations and conditions; we might anticipate that this could be revealed by comparing empirical associations between health and the measures of prestige, and the Goldthorpe class scheme. Such comparisons are logical if social stratification is seen as a multidimensional property and if different measures emphasise different elements of stratification (e.g. Goldthorpe 2010;

Torssander and Erikson 2010; Chan and Goldthorpe 2007; Wright 2005; Marshall et al. 1988).

In a statistical sense, 'multidimensionality' might imply separate metrics which are orthogonal, or wholly independent, of each other. However, in social science application areas, concepts that are said to involve different 'dimensions' are not usually orthogonal (e.g. Platt 2011; Tomlinson et al. 2008; Sacker et al. 2001). Instead, different 'dimensions' are different concepts that can be separately measured and do not have a *necessary* correlation or conjunction (i.e. it does not follow automatically that a high position in one dimension leads to a high position in another). For example, occupational categories in the Goldthorpe class scheme, and occupational CAMSIS scores, could be said to represent two separate dimensions, even though they tend to be strongly correlated (cf. Fig. 3.1).

Some stratification measures deliberately use indicators that combine different 'dimensions' of social difference. For instance, Savage et al.'s (2013) 'multidimensional' social class measure (Table 3.1, row 9) takes account of the combined circumstances that can be defined across several different forms of social advantage (related to economic, social and cultural capital); it is also, as a by-product, correlated with other dimensions of difference (e.g. age and region). Some other academic studies advocate social stratification indicators that are based upon measures that cross-cut dimensions (e.g. Hennig and Liao 2013; Pollock 2007), and in market research, there is a long tradition of devising taxonomies that are effective for prediction and are based upon multiple elements of the social structure (e.g. Burrows and Crow 2006). In contrast, others have argued that it is unhelpful to combine such disparate elements, and that it is preferable to design separate measures that allow us to disentangle the influences of different dimensions (e.g. Mills 2014).

In either case, there are important pragmatic difficulties in trying to disentangle different dimensions of social stratification. Firstly, different measures are strongly correlated with each other, which presents a general challenge of collinearity in interpretations. Secondly, stratification measures are characterised by slight disjunctions between their empirical qualities and the underlying dimensions that they are intended to capture (e.g. Lambert and Bihagen 2014; Bihagen and Lambert 2012): whilst different measures might be designed to reflect different mechanisms of inequality, they probably don't achieve this empirically with sufficient precision to reliably disentangle the mechanisms. Our own research suggests that we should think of stratification measures as imperfect indicators of their underlying concepts, perturbed by measurement errors that might occur for many different reasons (with occupational data, an important factor is whether the occupational category is aggregated appropriately). Ultimately, we suspect that it is better to focus theoretically upon a primary underlying structural inequality—of social stratification—rather than attempt to use different stratification measures to distinguish between different mechanisms, since even when measurement instruments might seek to distinguish between different dimensions of social stratification, they might not be effective in doing so (Lambert and Bihagen 2014).

## 3.4 The Importance of Occupations

In our analysis, we use occupations as the key markers of social stratification position, on grounds of convenience, empirical relevance, and because we find it theoretically compelling to place the occupational structure at the centre of our conceptualisation of the social organisation of social stratification. Occupational data is a convenient 'referent' for a measure of social stratification because it is easy to record in standardised taxonomies that preserve relatively fine-grained detail (e.g. Ganzeboom 2010); it works well as an indicator of long-term circumstances in the stratification structure, because occupational circumstances tend, on the whole, to be more stable indicators of long-term circumstances than do other measures that can fluctuate more freely (e.g. Bernardi and Ballarino 2016); and it has strong empirical associations with things that it would be expected to be associated with in relation to social stratification (e.g. Weeden et al. 2007). The theoretical appeal partly reflects the centrality of the occupational division of labour in resource distribution (especially Wright 2005). Although occupations are not the only things that matter to the social organisation of inequality, it is clear from previous analysis that the occupational order is an important part of the allocation of unequal social positions (e.g. Rose and Harrison 2010; Wright 1997;

Marshall et al. 1988; Parkin 1972). It also reflects our own theorisation of the role of the 'social resin' (the interplay of social interaction behaviour and occupational positions) in our account of social inequality (see Chaps. 2 and 12).

Nevertheless, many social scientists are not persuaded immediately of the key importance of occupations. Most obviously, how does such a perspective help us understand the circumstances of the many people who do not hold occupations, or who work in multiple jobs, or who perhaps have part-time or temporary occupations which they do not see as very central to their lives? For example, in the UK in summer 2012, data from the Labour Force Survey (see ONS 2012) indicates that only 71% of adults aged 16–64 held a job (76% of men and 66% of women); across the whole of the UK, 29.6 million individuals held a job, from a total population, including children and the retired, estimated at 63.0 million (ONS 2011). Moreover, 27% of those with a job held it on a part-time basis of less than 30 hours per week (ONS 2012). Comparable patterns are seen across nations, through time and across societies (e.g. Scott et al. 2008).

One point is that a social system can be universally influential without having to be universal—every individual does not need to vote, say, or to carry arms to war, for a society to be a democracy, or to be at war. With regard to occupations, not everyone needs to be in employment for the social organisation of employment to be a defining feature of the organisation of social inequality (as we argue that it is). Nevertheless, almost every individual can be linked to an occupation that is socially significant to them—if not their own current occupation, perhaps the last occupation they held, or the occupation of a significant other, such as a spouse or parent. Table 3.2, for instance, shows the proportion of people in the UK's British Household Panel Survey in 2008 (University of Essex 2010) who could be linked to an occupation by various commonly used criteria. It also shows the correlation between their occupation and a selection of measures that would ordinarily be expected to be linked to social inequality. The table shows that almost everyone can be matched to an occupation—95% of adults have a record for either a current job or a previously held occupation; 99% of adults either have that or can be linked to a job held by a household sharer or, if they are aged under 30, a parent.

**Table 3.2** Coverage of cases and relative correlations when linking individuals with occupations

| | | Linear correlation * 100 with … | | | |
|---|---|---|---|---|---|
| Measure | Coverage (%) | Smoking | Health | Financial anxiety | Father's CAMSIS |
| Current job CAMSIS | 77 | 14.6 | 7.7 | 16.4 | 27.5 |
| Current or most recent job | 95 | 18.8 | 13.6 | 18.7 | 27.4 |
| Current or most recent job or job of household sharer | 98 | 19.0 | 13.6 | 18.9 | 28.3 |
| Current or most recent job, or job of household sharer, or job of parent if age <30 | 99 | 19.1 | 13.5 | 19.1 | {28.9} |
| Personal income | 95 | 9.3 | 12.5 | 20.9 | 12.7 |
| Household income | 100 | 12.4 | 14.2 | 24.1 | 16.1 |

Source: British Household Panel Survey, 9067 interviewed adults at wave 18, excl. NI, unweighted. 'Coverage' refers to percentage of all records for whom an occupation-based or income-based score can be allocated. To hold an occupation-based record, an occupational code must be linked to the case; to hold an income record, a non-zero imputed income should be available.

Table 3.2 also shows (as did Table 3.1) that as 'instruments du travail', occupation-based measures generally have favourable empirical qualities as prospective indicators of social stratification position. This is evident by the correlations shown in relation to other plausible indicators—occupation-based measures seem to perform at least as well as, and sometimes better than, alternative indicators (see also Oesch 2013; Rose and Harrison 2010). Most probably these qualities reflect the very reasons that originally drew sociologists towards occupational data—namely, the centrality of occupations to economic resource distribution, the relative stability of occupations over time, and the consistency of data recorded about occupations.

In spite of the empirical evidence, there are intuitive reasons to anticipate that recent social change has diminished the centrality of occupations to the social stratification structure. Recent decades have been interpreted by some as a transition towards a 'new capitalism' in which individuals' occupations are no longer an important social anchor. Accounts such as those provided by Beck, Bauman, and Giddens have characterised a new society where occupations diminish in their importance, to be replaced by new cleavages based upon asset holding, consumption, other individual identities, and other dimensions of social inequality. The empirical weaknesses of these accounts are well rehearsed (e.g. Doogan 2009): robust statistical evidence indicates broad stability, rather than dramatic change, in employment levels and life-course employment histories (e.g. Blossfeld et al. 2006); in the importance of employment to income, wealth, and consumption (e.g. Oesch 2013; Rose and Harrison 2010; Weeden et al. 2007; Guveli 2006); and in the role of occupations in social reproduction (e.g. Bernardi and Ballarino 2016; Breen 2004). The point here is not that social and economic change does not occur, but that there is little evidence (aside from tendentious social commentary) to suggest that occupations have lost their ability to demarcate important patterns in life courses and social experiences (e.g. Oesch 2013; Penn 2006).

Asserting that occupations maintain considerable importance is not to deny that other factors or social changes are relevant. The lengthening of educational careers, the relative growth of 'fuzzy' or 'ill-defined' jobs in place of crisply defined roles, and changes in demographic inequalities such as changes in family formation patterns and in female career aspirations have all induced changes in occupational inequalities across countries and through time (e.g. Oesch 2013). There is compelling evidence of important variations within occupations relevant to social stratification structures (e.g. Laurison and Friedman 2016). Moreover, several writers have pointed out that even if objective change in the social importance of occupations may not be so dramatic, individuals' perceptions of occupational change can be dramatic, and evolutions in subjective understanding could be of independent consequence regardless of objective circumstances (e.g. Strangleman 2012; Doogan 2009).

Part of the attraction of occupation-based measures stems from the limitations of alternative prospective measures. Measures of educational experiences reflect long-term life circumstances, but data on education is problematic because it rarely captures fine-grained differences in circumstances, and the profile of educational experiences is strongly linked to birth cohorts due to educational expansion and reforms (see Chap. 10). Two other popular alternatives are measures based upon income and those based upon localised geographical profiles. Both of these indicators are easy to communicate and have been prominent tools in a recent wave of social inequality studies in the UK in particular (e.g. Dorling 2010). However, both of these indicators are relatively more problematic in characterising individuals than are occupation-based measures. To argue this point, Table 3.3 shows details of the residents of a fictional block of flats, all of whom receive the same income (and, by definition, live in the same locality). In this depiction, the retired home owner is in a very strong financial position, whereas the single parent with mortgage repayments and employment and childcare expenses is much more vulnerable. Educationally, the child of the teacher might well benefit from more effective support through their formal education than the child of the dental assistant. However, whilst the dental assistant has the lowest expendable income, and worst financial position, they might be much better equipped to engage effectively with health services, perhaps leading to better long-term health circumstances. In such ways, measures based on income and location can hide important variations amongst people.

Our example in Table 3.3 does not demonstrate that measures of occupations could not have similar problems. Indeed, there is ample evidence of inequalities between individuals within occupational positions, and this is often cited in literatures that argue against a focus upon occupational data (e.g. Savage et al. 2015).[3] However, our example highlights that equivalent or heightened problems apply to alternative,

**Table 3.3** Indicative circumstances of residents of the same block of flats who own or are buying their home and who all receive the same post-tax individual income (£1500 per month)

| | Additional costs | Expendable income |
|---|---|---|
| Retired | No outstanding mortgage | £1500 |
| Freelance magazine writer | Outstanding mortgage (£300 p.m.) | £1200 |
| Building site labourer | Mortgage, plus £50 p.m. commuting | £1150 |
| Part-time teacher, single parent of school-aged child | Mortgage, commuting, plus £250 p.m. living expenses for child | £900 |
| Hospital-based dental assistant, single parent of preschool-aged child | Mortgage, commuting, child living expense, plus £300 p.m. childcare | £600 |

non-occupation-based methods of measuring social positions. For example, geographical and income-based profiles are strongly anchored around differences in life-course stage (e.g. Savage et al. 2013), whereas occupational measures are more stable as average indicators of circumstances, and accordingly make more reliable stratification indicators (e.g. Rose and Pevalin 2003).

A plausible response to the problems above is the position that an individual's social stratification position is best understood by a multidimensional account of multiple relevant influences—such as including occupation, education, economic assets, cultural practices, social resources, and family origins (e.g. Savage et al. 2015; Biressi and Nunn 2013, p. 1). This approach can be compelling in some contexts, but for empirically oriented studies, it does have the shortcoming that measures are neither readily replicated nor easy to assign. In addition, empirical realisations of the multidimensional approach have been consistently criticised for being too closely aligned with circumstances that arguably should not be a part of the stratification structure—most commonly age and gender (e.g. Mills 2014). At present, it remains difficult to isolate positions in a multidimensional cartography from other demographic divisions, whereas, whether by design or good fortune, the long-term career orientation of occupational codes conveniently does so for occupation-based measures.

### 3.4.1 A Closer Look: Coding and Comparing Occupation-Based Measures

One of the attractions of working with data on occupations is that commonly used titles convey quite detailed information about occupations (and, by implication, the social circumstances experienced by their incumbents). Whilst, on many other measures, survey and survey-like data can appear to collect information in a rather crude and simplifying way, this is not usually the case for occupational circumstances. The tradition probably arises because in everyday linguistics people are used to describing their exact occupation (itself a revealing insight into the importance of occupations to society).

Most empirical projects begin by recording a free text description of occupations, and this is often supplemented with responses to additional questions on aspects of the employment relationship, such as managerial responsibility and self-employment status (e.g. Hoffmeyer-Zlotnik and Warner 2014, 6.1.7; Davies and Elias 2010). Typically, an electronic coding tool is used to assign the occupational description into a fine-grained measure of 'occupational unit group' (OUG), though other strategies can be deployed (e.g. Hoffmeyer-Zlotnik and Warner 2014; ONS 2010; cf. Ganzeboom 2010). In our own analyses, the finest grained detail on occupations that we exploit are occupational unit groups, but it is worth realising that even these units (typically taxonomies of about 200–400 categories) are themselves aggregations of more complex occupational descriptions.

How is this detailed data on occupations exploited? In a few projects, analysts are interested in the circumstances of highly specific occupations or occupational-unit groups, such as studies of how health risks vary across occupations (e.g. Stansfeld et al. 2011), or what influences selection to specific jobs (e.g. Dolton and Makepeace 1993). More often, however, analysts use detailed coding simply as a means of constructing a measure of social position. For this purpose, allocation rules can be used to derive measures of social position (such as measures of 'social class' or scores on gradational stratification scales) on the basis of the occupational unit group (and, frequently, employment status information in addition). This coding activity can raise problems regarding the consistent and optimal treatment of occupational data (cf. Lambert et al. 2007). Indeed, coding detailed occupational data is a relatively challenging technical activity about which many researchers have little expertise.

Data on occupational unit groups is usually organised in a manner that allows for alternative levels of aggregation. The influential International Standard Classification of Occupations 1988 (ISCO-88—see ILO 1990) scheme makes a good example. It organises categories into a four-level hierarchy of which the most disaggregated measure is the 'unit group' (with 390 different categories), then the 'minor group' (116 different categories), the 'sub-major group' (28 different categories), and the 'major group' (ten different categories). Like many other occupational unit group schemes, ISCO-88 organises its categories through numeric indicator codes, where the number of digits in the numeric code largely corresponds to the hierarchical level of detail. For example, it features a code of 1234 for 'Advertising and public relations department managers' at the unit group level. That category is part of the minor group 123 ('Other department managers'), the sub-major group 12 ('Corporate managers'), and the major group 1 ('Legislators, senior officials, and mangers').

Although it is common practice, some have claimed that, in the social science research context, there is little pay-off to collecting very detailed occupational measures, since there may be little difference to the results that might be obtained if data is collected through a much cruder typology, such as if asking respondents to allocate their occupation to a category from a list of perhaps 10 or 20 options (e.g. Hoffmeyer-Zlotnik and Warner 2014, p. 133; Ganzeboom 2005). Williams (2013) demonstrates that more detailed occupational codes are needed to fully understand changes in wage inequality, but that more aggregate measures are robust if the interest is focussed on between occupational variations. However, other writers stress the empirical value of measuring relatively fine-grained occupational differences when studying social stratification and social reproduction (e.g. Weeden and Grusky 2012; Scase 1992, p. 34). The tradition of collecting and using very detailed occupational data is perhaps most deeply entrenched within sociology; in many other social science disciplines, only relatively simplified occupation-based measures are ordinarily used (if they are used at all).[4]

Figure 3.2 provides some evidence on the relevance of using detailed occupational data. It shows statistics that summarise the performance of a range of occupation-based measures, and other measures of social position, in explaining two indicative outcomes (to what extent a respondent agrees with the statement that 'homosexuality is wrong' and how a respondent replies when asked if they tend to feel optimistic about the future). The different occupation-based measures have been derived by using more or less detailed occupational data (indicated '1-dig', '2-dig' etc., where fewer digits means less detail).[5] Additionally, for some measures (the Goldthorpe class scheme ('EGP') and the measures of tenure, education, and income), the level of aggregation in the indicator is also varied (here, the first number indicates the number of different

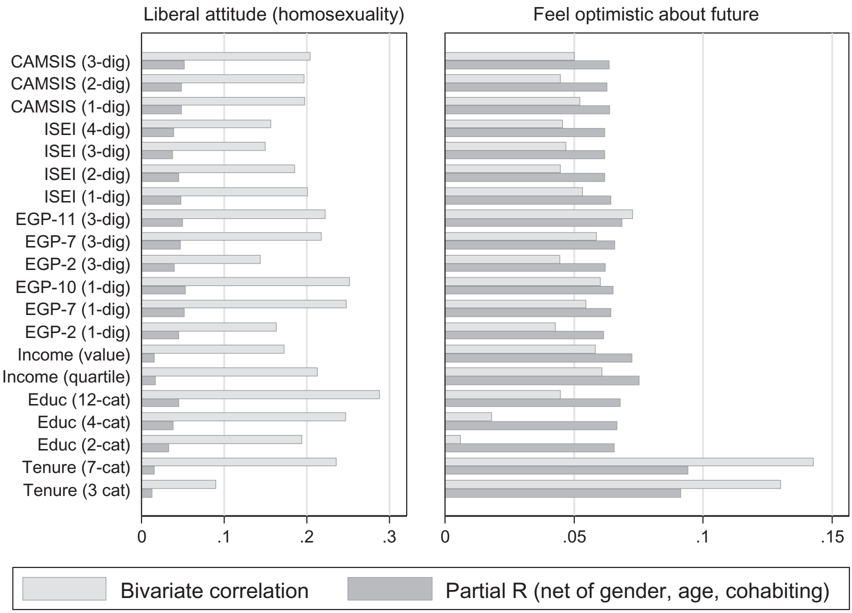



**Fig. 3.2** The relative performance of alternative measures of stratification, with variation by levels of occupational detail. Source: BHPS 2008, all adults with valid data on 'most recent occupation' in SOC-90 unit groups, *N* ~= 12,000, using sampling weights for UK-level analysis. Statistics shown are either bivariate correlation or the increment in regression model R from adding the relevant measure

categories that are distinguished, for instance, 'EGP-7' indicating a measure of the EGP scheme with seven different categories, and 'Educ (2-cat)' a measure of education with two different categories). The results in Fig. 3.2 show bivariate correlation statistics for the overall relationship between the selected outcomes and a small range of measures of social position that do not use occupations, including income, education and housing tenure, and a selection of occupation-based measures (the CAMSIS scale, the ISEI scale, and different versions of the EGP class scheme). They also show a 'partial-R' statistic, which here reflects the increment in the coefficient of determination between the model with the social position measure and the model without it, given (in both models) background controls for gender, age, and cohabitation status. We would expect these outcomes to correlate with measures of social position, so a plausible interpretation is that the bigger the correlation, the 'better' is the performance of the measure of social position.

An impression from Fig. 3.2 is that, whilst there are some differences amongst the measures, in general the choice of occupation-based measure might be of limited impact, since the correlations and partial explanations seem to be much the same regardless of whether more or less detailed occupational information was used (for instance, the performance of the measure of CAMSIS based on more detailed, three-digit occupational codes is much the same as the performance of the measure based upon two-digit codes). It is even the case that some measures based upon less detailed codes perform 'better', as, for instance, in the ISEI measure (this might perhaps reflect that the aggregate measures are less vulnerable to measurement error). Additionally, Fig. 3.2 also suggests that occupation-based measures generally have similar properties when compared to measures based on non-occupational factors such as income, education and housing tenure (although there is some variation according to the outcome measure—for instance, housing tenure is more strongly linked to 'optimism', perhaps reflecting that financial security might accompany an advantaged housing tenure in the UK).

At first sight, evidence such as Fig. 3.2 suggests that we can often afford to be sanguine in using measures of social position—pretty much any measure should be reasonably effective, and will tell pretty much the same story, whether it is based upon more fine-grained or more aggregate underlying data. Ganzeboom (2005) certainly takes this position after conducting a range of sensitivity analyses. To a certain extent, we believe this is true, and it would clearly be a convenient outcome for the wider discipline given the current situation of limited coordination. However, there are three important qualifications.

Firstly, the very presence of small differences in performance between different measures (e.g. Figs. 3.1 and 3.2) indicates that the different measures are not entirely interchangeable. In some scenarios, the differences might not be of great consequence, but in others they might. In particular, the influence of alternative parameterisations is likely to be most consequential when the relationship is itself at the margins of being statistically significant: in such situations, it could well be the case that the difference between using a more or less detailed measure (or a measure based upon one thing rather than another) could make the difference to whether a key result is estimated as statistically significant or not (e.g. Lambert and Bihagen 2014). Indeed, we can see the consequences more clearly if we focus upon sub-sections of a population. In Fig. 3.3, we summarise the overall population-level linear correlation between CAMSIS and ideological outlook for the versions of CAMSIS based on three-digit and one-digit occupational data for UK data. However, we also summarise the relationship between the detailed CAMSIS score and the outcome, within each of nine 'major groups' within the UK Standard Occupational Classification (the thin lines). For one thing, we see that the slope of the one-digit version is slightly steeper than that of the version based on three-digit occupational data—this could suggest that the averaging process masks occupational variation and overestimates the relationship. Thereafter, the major group lines are interesting because

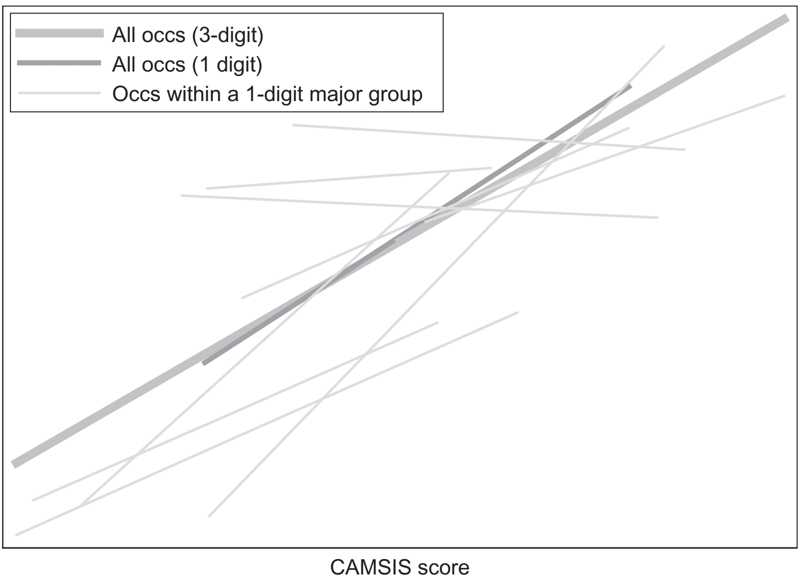



**Fig. 3.3** The influence of CAMSIS score upon liberal attitudes to homosexuality, for individuals in all occupations and for individuals within different occupational major groups. Source: BHPS, 2008, all adults with valid data on 'most recent occupation' in SOC-90 unit groups, *N* = 8395

they reveal that, within major groups, there are patterns of variation in CAMSIS scores that shape different relationships with the outcome. If we were using a one-digit version of occupational data from the start, all individuals within the same major group would be forced, by design, to have the same CAMSIS score (the within-group lines would be constrained to being horizontal). In our experience, these patterns of internal heterogeneity within aggregate categories are reasonably common (see also Prandy 1990), but of course in any given empirical study, we cannot know in advance that they will matter, unless and until we test for them. In our view, a sensible response is simply to favour the measure that exploits the most available detail, which ought, in principle, to be less likely to miss an important structural pattern.

Secondly, in many situations there is compelling evidence that there are additional empirical patterns associated with specific occupational positions, over and above those which are captured through the parameters of an occupation-based measure. Two social mechanisms might be involved—on the one hand, the influence of highly specific occupation-to-occupation differences and, on the other, the more generic influence of social structure captured by occupation-based measures. This dual model has often been recognised in studies of intergenerational mobility (e.g. Jonsson et al. 2009; Devine 2004). For example, the son of two teachers is relatively more likely to achieve an occupational position of a similarly advantaged level, thanks to the generic resources associated with his relatively privileged background; however, he is also relatively more likely to obtain a job specifically in the educational sector (or in cognate fields, such as in scientific research), because he will typically have inherited a number of skills, competencies, and propensities that are directly linked to his parents' jobs. Needless to say, detailed occupational data is required to assess such influences. In the social mobility tradition, it is common to demarcate specific 'occupational inheritance' parameters from other model parameters for this purpose (e.g. Luijkx 1994; Erikson and Goldthorpe 1992). It is possible to use log-linear models that capture a range of levels of occupational detail and evaluate the relative influence of the different levels upon social mobility patterns (cf. Weeden and Grusky 2012; Jonsson et al. 2009). 'Random effects' models ('multilevel models') can also potentially be used for this purpose—see discussion in

Chap. 9.[6] Such strategies to demarcate occupation-specific influences and more generic stratification effects could change the substantive interpretation made; conclusions drawn about the role of other explanatory factors might be altered, and control for detailed occupations will also probably lead to a change in the estimation of the overall magnitude of occupational effects.

A third reason for paying attention to more detailed occupational information is the most important with regard to the themes of this book. This is that there are certain topic areas where differences between specific occupations are of heightened interest. In particular, when exploring the social relations associated with occupations, it is immediately obvious that there could be important mechanisms of social connections that come down to the relations between specific occupational positions. We return to this issue in Chaps. 5 and 8.

## 3.5 Preparing Data on Measures of Social Stratification, Occupations, and Social Connections

How variables are constructed can have a critical influence upon results, but some of the activities associated with constructing measures of social stratification are technically complex—such as in retrieving and exploiting detailed information about occupations. To make good progress it is worth reflecting on the 'data management' or preparatory work associated with data resources (e.g. Lambert 2015; Dale 2006). A standard way of recording how data is processed is to use a software 'syntax' or 'scripting' language which provides a 'log' of the tasks undertaken (e.g. Long 2009). In recent years, motivated in part by developments in information science, new services have also been developed with the explicit intention of recording and distributing important metadata concerned with data management as it is relevant to measures of stratification (e.g. Lambert 2015; ADLS 2012; Hagenaars 2008), and researchers are encouraged more generally to construct and store documentation about their data more rigorously (e.g. Mohler et al. 2008).

One challenge for social researchers is that the exploitation of metadata on measures of stratification generally requires subject-specific expertise. Many researchers simply do not know about metadata resources that could be of relevance to their work. Resources typically originate with academic methodologists and/or national statistical institutes, who publish materials recommending certain approaches or measures (e.g. Hoffmeyer-Zlotnik and Warner 2014; Bulmer et al. 2010). Within this tradition there is a considerable literature regarding the construction of measures based upon occupations (e.g. MPC 2012; Ganzeboom 2010; Rose and Harrison 2010). As a brief 'beginners guide' to good practice in this area, we highlight six summary points about data management and using measures of social stratification:

(1) Online information is available to connect 'occupational unit group' codes with stratification measures, as well as with other descriptive data about the codes (e.g. Ganzeboom 2016; Lambert 2016). This data often takes the form of small databases or software code files that will allow automated coding from OUG categories to stratification measures.
(2) Notwithstanding (1), it is not uncommon to experience a mismatch between occupational codes on a dataset, and online information—for instance, if occupational codes are at a more aggregated level than is requested by the online coding frames, or if the coding frames also request other data linked to employment, such as on contract status. In such situations, some online sources provide recommendations for aggregation strategies, but analysts must sometimes make their own judgements (in which case it is good practice to clearly describe the procedures followed).
(3) Comparability between occupational data from different societies cannot be presumed, not least since OUG schemes are usually different between countries and may change over time. Internationally standardised taxonomies (e.g. ILO 1990) are popular options, but approaches are not universally agreed upon (e.g. Ganzeboom 2005). As mentioned previously, similar comparability problems arise when comparing male and female populations due to the worldwide phenomena of gender segregation in occupations (e.g. Charles and

Grusky 2004). There are no simple, agreed-upon solutions, except the prescription that comparability should be reflected upon and that strategies employed should be described transparently.

(4) When processing occupational data, methodologists have long argued that it is much better to use existing, published guidance to derive well-recognised occupation-based measures, rather than devising new or ad hoc measures (e.g. Connelly et al. 2016; Bechhofer 1969), or defaulting to the most readily available measure without further consideration (cf. Lambert and Bihagen 2014).

(5) Although it is often pragmatically simpler to focus only on the 'current occupation' held by an individual, in many social science datasets it is plausible that additional information on other occupations might provide better data on social stratification. Other data might cover information on job(s) previously held, multiple jobs held. or the jobs held by influential others, such as spouses or parents.

(6) Lastly, data on social connections between individuals and their occupations often requires the analyst to attach data about one individual (the 'ego') with that for another unit (the 'alter'). In the case of household surveys, this can be done by using data on the relationship between ego and alter. A few automated facilities for similar linkages exist in special scenarios (e.g. within the IPUMS-I download software; see MPC 2015). Commonly however manual data linkage activities are required that use 'match-merge' routines in statistical software (e.g. Lambert 2015).

## Notes

1. The gradational view is usually associated with a unidimensional model of stratification. Technically however we could analyse gradational inequalities in more than one dimension or in combination with categorical divisions. Gradational measures are sometimes used in models of stratification that use multiple dimensions (e.g. Bourdieu 1984) or a mixture of gradational and nominal dimensions (e.g. Jonsson et al. 2009; Rytina 2000).
2. We thank Erik Bihagen for support in accessing the Swedish data.

3. Here the 'declinist' perspective suggests that, although occupations may once have been good tools for indicating social circumstances, this is no longer the case, particularly because of growing within occupational inequalities (e.g. Dorling 2014). In our view, this is not an accurate portrayal of historical trends in within occupational inequalities—which have always been noted (cf. Routh 1980); indeed, the declinist claim is sometimes made only on the basis of contemporary patterns, without robust historical comparisons.
4. This is true even when research in other disciplines enters the traditional topic areas of sociology—see Goldthorpe (2010) for critical reflections on how studies in economics and in public health research ignore relevant data on occupations.
5. 'One-digit' refers to a version based on measurement of occupational titles with only at most ten different categories, 'two-digit', 'three-digit', and 'four-digit' as applicable refer to versions that use measures of occupations with increasingly more categories; the most extensive versions are the 'four-digit' version for ISEI (which is based upon the ISCO-88 four-digit scheme) and the 'three-digit' versions for CAMSIS and EGP (which are based upon the UK SOC90 three-digit scheme), both of which allow for around 350 different occupational unit group categories.
6. 'Fixed effects' models are also sometimes used in this way (e.g. Stansfeld et al. 2011); however, in this setting it is more problematic to differentiate between occupational-specific effects and generic effects that might be measured at the occupational level.

## Bibliography


Abbott, A. (2006). Mobility: What? When? How? In S. L. Morgan, D. B. Grusky, & G. S. Fields (Eds.), *Mobility and Inequality* (pp. 137–161). Stanford, CA: Stanford University Press.

ADLS. (2012). *P-ADLS: Portal of the Administrative Data Liaison Service*. Retrieved May 1, 2012, from http://www.adls.ac.uk/padls/.

Atkinson, A. B. (2005). Social Indicators, Policy, and Measuring Progress. In S. Svallfors (Ed.), *Analyzing Inequality* (pp. 83–107). Stanford, CA: Stanford University Press.

Bechhofer, F. (1969). Occupations. In M. Stacey (Ed.), *Comparability in Social Research* (pp. 94–122). London: Heinemann.

Bernardi, F., & Ballarino, G. (Eds.). (2016). *Education, Occupation and Social Origin: A Comparative Analysis of the Transmission of Socio-Economic Inequalities*. Cheltenham: Elgar.

Bihagen, E., & Lambert, P. S. (2012, September, 24–26). *Class and Status: One or Two Forms of Stratification?* Paper Presented at the Equalsoc Conference, Stockholm.

Biressi, A., & Nunn, H. (2013). *Class and Contemporary British Culture*. Basingstoke: Palgrave Macmillan.

Blackburn, R. M. (2008). What Is Social Inequality? *International Journal of Sociology and Social Policy, 28*(7/8), 250–259.

Blossfeld, H. P., Mills, M., & Bernardi, F. (Eds.). (2006). *Globalization, Uncertainty and Men's Careers: An International Comparison*. Cheltenham: Edward Elgar.

Bourdieu, P. (1984). *Distinction*. London: Routledge and Kegan Paul.

Breen, R. (Ed.). (2004). *Social Mobility in Europe*. Oxford: Oxford University Press.

Bulmer, M., Gibbs, J., & Hyman, L. (Eds.). (2010). *Social Measurement Through Social Surveys: An Applied Approach*. Aldershot: Ashgate.

Burrows, R., & Crow, G. (2006). Geodemographics, Software and Class. *Sociology, 40*(5), 793–812.

Chan, T. W. (Ed.). (2010). *Social Status and Cultural Consumption*. Cambridge: Cambridge University Press.

Chan, T. W., & Goldthorpe, J. H. (2007). Class and Status: The Conceptual Distinction and Its Empirical Relevance. *American Sociological Review, 72*(4), 512–532.

Charles, M., & Grusky, D. B. (2004). *Occupational Ghettos: The Worldwide Segregation of Women and Men*. Stanford, CA: Stanford University Press.

Connelly, R., Gayle, V., & Lambert, P. S. (2016). Modelling Key Variables in Social Science Research: Introduction to the Special Section. *Methodological Innovations Online*. https://doi.org/10.1177/2059799116637782.

Crompton, R. (1998). *Class and Stratification: An Introduction to Current Debates* (2nd ed.). Cambridge: Polity Press.

Crompton, R., & Mann, M. (Eds.). (1986). *Gender and Stratification*. Cambridge: Polity Press.

Dale, A. (2006). Quality Issues with Survey Research. *International Journal of Social Research Methodology, 9*(2), 143–158.

Davies, R., & Elias, P. (2010). The Application of ESeC to Three Sources of Comparative European Data. In D. Rose & E. Harrison (Eds.), *Social Class in Europe: An Introduction to the European Socio-Economic Classification* (pp. 61–86). London: Routledge.

Devine, F. (2004). *Class Practices: How Parents Help Their Children Get Good Jobs*. Cambridge: Cambridge University Press.
Dolton, P. J., & Makepeace, G. H. (1993). Female Labour Force Participation and the Choice of Occupation: The Supply of Teachers. *European Economic Review, 37*(7), 1393–1411.
Doogan, K. (2009). *New Capitalism? The Transformation of Work*. Cambridge: Polity.
Dorling, D. (2010). *Injustice: Why Social Inequality Persists*. Bristol: The Policy Press.
Dorling, D. (2014). Thinking About Class. *Sociology, 48*(3), 452–462.
Elias, P., & McKnight, A. (2001). Skill Measurement in Official Statistics: Recent Developments in the UK and the Rest of Europe. *Oxford Economic Papers, New Series, 53*(3), 508–540.
Erikson, R., & Goldthorpe, J. H. (1992). *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford: Clarendon Press.
Erikson, R., Goldthorpe, J. H., & Portocarero, L. (1979). Intergenerational Class Mobility in Three Western European Societies: England, France and Sweden. *British Journal of Sociology, 30*, 415–440.
Evans, G., & Mills, C. (1998). Identifying Class Structure: A Latent Class Analysis of the Criterion-Related and Construct Validity of the Goldthorpe Class Schema. *European Sociological Review, 14*(1), 87–106.
Ganzeboom, H. B. G. (2005). On the Cost of Being Crude: A Comparison of Detailed and Coarse Occupational Coding. In J. H. P. Hoffmeyer-Zlotnik & J. Harkness (Eds.), *Methodological Aspects in Cross-National Research* (pp. 241–257). Mannheim: ZUMA, Nachrichten Spezial.
Ganzeboom, H. B. G. (2010). *Do's and Don'ts of Occupation Coding with an Extension to ISCO-08*. Working Paper of the Department of Social Research Methodology. Amsterdam: Free University of Amsterdam.
Ganzeboom, H. B. G. (2016). *Tools for Deriving Occupational Status Measures from ISCO-88*. Retrieved March 1, 2016, from http://www.harryganzeboom.nl/isco88/index.htm.
Ganzeboom, H. B. G., de Graaf, P. M., & Treiman, D. J. (1992). A Standard International Socio-Economic Index of Occupational Status. *Social Science Research, 21*, 1–56.
Ganzeboom, H. B. G., & Treiman, D. J. (2003). Three Internationally Standardised Measures for Comparative Research on Occupational Status. In J. H. P. Hoffmeyer-Zlotnik & C. Wolf (Eds.), *Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables* (pp. 159–193). New York: Kluwer Academic Press.

Goldthorpe, J. H. (1997). The "Goldthorpe" Class Schema: Some Observations on Conceptual and Operational Issues in Relation to the ESRC Review of Government Social Classifications. In D. Rose & K. O'Reilly (Eds.), *Constructing Classes* (pp. 40–48). Colchester: University of Essex.

Goldthorpe, J. H. (2010). Analysing Social Inequality: A Critique of Two Recent Contributions from Economics and Epidemiology. *European Sociological Review, 26*(6), 731–744.

Gordon, D. (2006). The Concept and Measurement of Poverty. In C. Pantazis, D. Gordon, & R. Levitas (Eds.), *Poverty and Social Exclusion in Britain: The Millenium Survey* (pp. 29–69). Bristol: The Policy Press.

Gornick, J., & Jantti, M. (Eds.). (2013). *Income Inequality: Economic Disparities and the Middle Class in Affluent Countries*. Stanford, CA: Stanford University Press.

Grusky, D. B., & Weeden, K. A. (2006). Does the Sociological Approach to Studying Social Mobility Have a Future? In S. L. Morgan, D. B. Grusky, & G. S. Fields (Eds.), *Mobility and Inequality* (pp. 85–108). Stanford, CA: Stanford University Press.

Guveli, A. (2006). *New Social Classes Within the Service Class in the Netherlands and Britain: Adjusting the EGP Class Schema for the Technocrats and the Social and Cultural Specialists*. Nijmegen: Radboud University Nijmegen.

Hagenaars, J. A. (2008). *COMPSOC: Exploiting, Documenting and Enriching Comparative Data from Large-Scale Surveys in the Social Sciences*. Tilburg University, Investeringen NWO-Middelgroot.

Harkness, S. (2013). Women's Employment and Household Income Inequality. In J. C. Gornick & M. Jantti (Eds.), *Income Inequality: Economic Disparities and the Middle Class in Affluent Countries* (pp. 207–233). Stanford, CA: Stanford University Press.

Hauser, R. M., & Warren, J. R. (1997). Socioeconomic Indexes for Occupations: A Review, Update, and Critique. *Sociological Methodology, 27*, 177–298.

Hennig, C., & Liao, T. F. (2013). How to Find an Appropriate Clustering for Mixed Type Variables with Application to Socio-Economic Stratification. *Journal of the Royal Statistical Society Series C, 62*(3), 309–369.

Hoffmeyer-Zlotnik, J. H. P., & Warner, U. (2014). *Harmonising Demographic and Socio-Economic Variables for Cross-National Comparative Survey Research*. Berlin: Springer.

ILO. (1990). *ISCO-88: International Standard Classification of Occupations*. New York: International Labour Office.

Jenkins, S. P. (2011). *Changing Fortunes: Income Mobility and Poverty Dynamics in Britain*. Oxford: Oxford University Press.

Jonsson, J. O., Grusky, D. B., Di Carlo, M., Pollak, R., & Brinton, M. C. (2009). Microclass Mobility: Social Reproduction in Four Countries. *American Journal of Sociology, 114*(4), 977–1036.

Lambert, P. S. (2015). Advances in Data Management for Social Survey Research. In P. Halfpenny & R. Procter (Eds.), *Innovations in Digital Research Methods* (pp. 105–122). London: Sage.

Lambert, P. S. (2016). *CAMSIS Project: Files for Distribution Covering Occupational Unit Codes and Translations*. Retrieved December 1, 2016, from http://www.camsis.stir.ac.uk/occunits/distribution.html.

Lambert, P. S., & Bihagen, E. (2014). Using Occupation-Based Social Classifications. *Work, Employment and Society, 28*(3), 481–494.

Lambert, P. S., Tan, K. L. L., Turner, K. J., Gayle, V., Prandy, K., & Sinnott, R. O. (2007). Data Curation Standards and Social Science Occupational Information Resources. *International Journal of Digital Curation, 2*(1), 73–91.

Laurison, D., & Friedman, S. (2016). The Class Pay Gap in Higher Professional and Managerial Occupations. *American Sociological Reivew, 81*(4), 668–695.

Leiulfsrud, H., Bison, I., & Jensberg, H. (2005). *Social Class in Europe: European Social Survey 2002/3*. Trondheim: NTNU Samfunnsforskning/NTNU Social Research Ltd.

Long, J. S. (2009). *The Workflow of Data Analysis Using Stata*. Boca Raton: CRC Press.

Luijkx, R. (1994). *Comparative Loglinear Analyses of Social Mobility and Heterogamy*. Tilburg: Tilburg University Press.

Marshall, G., Rose, D., Newby, H., & Vogler, C. (1988). *Social Class in Modern Britain*. London: Routledge.

Miller, R. L. (1998). Worklife Mobility Typologies as Background to Current Class Position—A Research Note. *Economic and Social Review, 29*(3), 259–284.

Mills, C. (2014). The Great British Class Fiasco: A Comment on Savage et al. *Sociology, 48*(3), 437–444.

Minnesota Population Center. (2015). *Integrated Public Use Microdata Series, International: Version 6.4 [Machine-Readable Database]*. Minneapolis, MN: University of Minnesota.

Mohler, P. P., Pennell, B.-E., & Hubbard, F. (2008). Survey Documentation: Toward Professional Knowledge Management in Sample Surveys. In E. De Leeuw, J. Hox, & D. A. Dillman (Eds.), *International Handbook of Survey Methodology* (pp. 403–420). Hove: Psychology Press.

OECD. (2008). *Growing Unequal*. Paris: Organisation for Economic Cooperation and Development.

Oesch, D. (2006). *Redrawing the Class Map: Stratification and Institutions in Britain, Germany, Sweden and Switzerland*. Basingstoke: Palgrave.

Oesch, D. (2013). *Occupational Change in Europe: How Technology and Education Transform the Job Structure*. Oxford: Oxford University Press.

ONS. (2010). *Standard Occupational Classification 2010. Volume 1: Structure and Descriptions of Unit Groups*. Basingstoke: Palgrave Macmillan (with the Office for National Statistics).

ONS. (2011). *National Population Projections, 2010-Based Projections*. London: Office for National Statistics. Retrieved from www.ons.gov.uk.

ONS. (2012, September). *Labour Market Statistics Data Tables (Excel Spreadsheets)*. London: Office for National Statistics. Retrieved from www.ons.gov.uk.

Parkin, F. (1972). *Class Inequality and Political Order: Social Stratification in Capitalist and Communist Societies*. London: MacGibbon and Kee.

Penn, R. D. (2006). *Social Change and Economic Life in Britain*. Bologna: Homeless Book.

Platt, L. (2011). *Understanding Inequalities: Stratification and Difference*. Cambridge: Polity.

Pollock, G. (2007). Holistic Trajectories: A Study of Combined Employment, Housing and Family Careers by Using Multiple-Sequence Analysis. *Journal of the Royal Statistical Society Series A, 170*(1), 167–183.

Prandy, K. (1990). The Revised Cambridge Scale of Occupations. *Sociology, 24*(4), 629–655.

Prandy, K., & Blackburn, R. M. (1997). Putting Men and Women into Classes: But Is That Where They Belong? A Comment on Evans. *Sociology, 31*(1), 143–152.

Prandy, K., & Lambert, P. S. (2003). Marriage, Social Distance and the Social Space: An Alternative Derivation and Validation of the Cambridge Scale. *Sociology, 37*(3), 397–411.

Rose, D., & Harrison, E. (2007). The European Socio-Economic Classification: A New Social Class Scheme for Comparative European Research. *European Societies, 9*(3), 459–490.

Rose, D., & Harrison, E. (Eds.). (2010). *Social Class in Europe: An Introduction to the European Socio-economic Classification*. London: Routledge.

Rose, D., & Pevalin, D. J. (Eds.). (2003). *A Researcher's Guide to the National Statistics Socio-Economic Classification*. London: Sage.

Routh, G. (1980). *Occupation and Pay in Great Britain, 1906–79*. London: Macmillan.
Rytina, S. (2000). *A Keynote of Hierarchy, an Echo of Class: The Full Texture of Mobility in England and Wales.* Paper Presented at the RC33—Fifth International Conference on Logic and Methodology, Cologne.
Sacker, A., Bartley, M., Firth, D., & Fitzpatrick, R. (2001). Dimensions of Social Inequality in the Health of Women in England: Occupational, Material and Behavioural Pathways. *Social Science and Medicine, 52*(5), 763–781.
Savage, M., Cunningham, N., Devine, F., Friedman, S., Laurison, D., McKenzie, L., Miles, A., Snee, H., & Wakeling, P. (2015). *Social Class in the 21st Century*. London: Pelican Books.
Savage, M., Devine, F., Cunningham, N., Taylor, M., Li, Y., Hjellbrekke, J., Le Roux, B., Friedman, S., & Miles, A. (2013). A New Model of Social Class: Findings from the BBC's Great British Class Survey Experiment. *Sociology, 47*(2), 219–250.
Scase, R. (1992). *Class*. Buckingham: Open University Press.
Scott, J., Dex, S., & Joshi, H. (Eds.). (2008). *Women and Employment: Changing Lives and New Challenges*. Cheltenham: Edward Elgar.
Scottish Government. (2016). *The Scottish Index of Multiple Deprivation*. Retrieved March 1, 2016, from http://www.gov.scot/Topics/Statistics/SIMD.
Shaw, M., Galobardes, B., Lawlor, D. A., Lynch, J., Wheeler, B., & Davey Smith, G. (2007). *The Handbook of Inequality and Socioeconomic Position: Concepts and Measures*. Bristol: Policy Press.
Stansfeld, S. A., Rasul, F., Head, J., & Singleton, N. (2011). Occupation and Mental Health in a National UK Survey. *Social Psychiatry and Psychiatric Epidemiology, 46*(2), 101–110.
Strangleman, T. (2012). Work Identity in Crisis? Rethinking the Problem of Attachment and Loss at Work. *Sociology, 46*(3), 411–425.
Szreter, S. R. S. (1984). The Genesis of the Registrar-General: Social Classification of Occupations. *British Journal of Sociology, 35*(4), 522–546.
Tahlin, M. (2007). Class Clues. *European Sociological Review, 23*(5), 557–572.
Therborn, G. (2013). *The Killing Fields of Inequality*. Cambridge: Polity Press.
Tomlinson, M., Walker, R., & Williams, G. (2008). Measuring Poverty in Britain as a Multi-Dimensional Concept, 1991–2003. *Journal of Social Policy, 37*(4), 597–620.
Torssander, J., & Erikson, R. (2010). Stratification and Mortality—A Comparison of Education, Class, Status, and Income. *European Sociological Review, 26*(4), 465–474.

Treiman, D. J. (1977). *Occupational Prestige in Comparative Perspective*. New York: Academic Press.

University of Essex, & Institute for Social and Economic Research. (2010, July). *British Household Panel Survey: Waves 1–18, 1991–2009 [Computer File]* (7th ed.). Colchester, Essex: UK Data Archive [distributor], SN: 5151.

Weeden, K. A., & Grusky, D. B. (2012). The Three Worlds of Inequality. *American Journal of Sociology, 117*(6), 1723–1785.

Weeden, K. A., Kim, Y.-M., Di Carlo, M., & Grusky, D. B. (2007). Social Class and Earnings Inequality. *American Behavioral Scientist, 50*(5), 702–736.

Williams, M. (2013). Occupations and British Wage Inequality, 1970s–2000s. *European Sociological Review, 29*(4), 841–857.

Wright, E. O. (1997). *Class Counts: Comparative Studies in Class Analysis*. Cambridge: Cambridge University Press.

Wright, E. O. (Ed.). (2005). *Approaches to Class Analysis*. Cambridge: Cambridge University Press.

# 4

# CAMSIS and the Analysis of Social Interaction Distance

## 4.1 The CAMSIS Approach

This chapter introduces the CAMSIS approach to 'social interaction distance' (SID) analysis (Sect. 4.1). We discuss the background to the approach in Sect. 4.2 and theories that are associated with it in Sect. 4.3. Subsequently, Chaps. 5 and 6 turn to the empirical features of CAMSIS scales and the practical aspects of their construction.

CAMSIS, for 'Cambridge Social Interaction and Stratification Scales', refers to a long-standing approach towards analysing data on the social interaction patterns exhibited by the incumbents of occupational positions that was developed by three academics based at the University of Cambridge: Sandy Stewart, Bob Blackburn, and Ken Prandy (cf. Prandy 1990; Stewart et al. 1973, 1980). The original work of the 'Cambridge group' used data from the UK, and the main product of their study, a scale giving scores to occupations that reflected their social interaction patterns, became widely known as the 'Cambridge scale'. Nevertheless, the methodology involved could be applied more widely, and the name CAMSIS came to be used when similar scales were gradually constructed for other societies (e.g. Lambert et al. 2013; de Luca et al. 2010; Prandy



P. Lambert, D. Griffiths, *Social Inequalities and Occupational Stratification*,
https://doi.org/10.1057/978-1-137-02253-0_4

and Lambert 2003; Bergman et al. 2002; Prandy and Jones 2001). The ongoing CAMSIS project continues to work on generating scales based on the social interaction distance between occupations for a range of societies.

The current activities of the CAMSIS project are organised around a website from where CAMSIS scales from many different countries and time periods are disseminated, along with supplementary information about the methodology and its characteristics (see www.camsis.stir.ac.uk). From that website, it is possible to download data files that link occupational units to appropriate CAMSIS scores. Each downloadable file is associated with a different CAMSIS 'version'. A version is defined according to the country with which it is associated, the time period of the data on which the CAMSIS analysis was performed, and the occupational units on which the measure is based.

At the heart of the CAMSIS approach lie two principles. The first is that it is possible and useful to use individual patterns of social connections as a way of finding out about wider structural relations. The principle actually applies more generally than to the application area of the CAMSIS approach (occupations). For example, using an approach which is quite close to that of the CAMSIS methodology, data on social connections was recently used to generate an accurate geographical map of the world simply through statistical summaries of the relative frequency of pairs of 'facebook friends' (see Butler 2010).

The second principle of the CAMSIS approach is the tradition of analysing the social structure in terms of occupations. Although occupations are not necessarily the only appropriate indicators which could be used, occupations are felt to be especially influential and consistent features of the social structure (e.g. Prandy 1990, and see Sect. 3.4). Accordingly, the CAMSIS methodology is designed to reveal a structural pattern to the social interactions between occupations—providing information about social stratification as it is related to occupations.

### 4.1.1 Illustrative Example

CAMSIS scales are usually estimated on datasets with quite a large number of different occupational units, which in general makes their specification a little more complicated. To begin, however, we introduce the

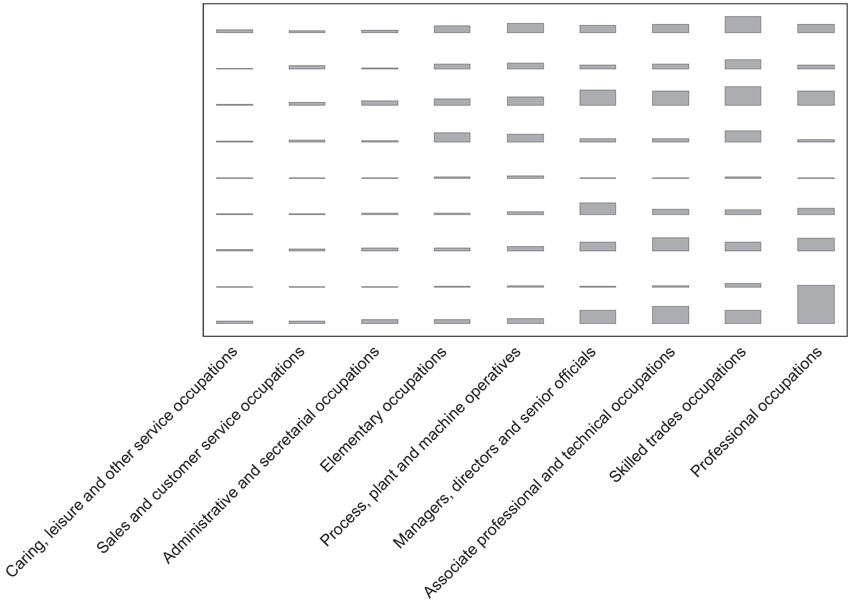



**Fig. 4.1** Representation of the distribution of occupations held by males (columns) and females (rows) for heterosexual couples. Source: 67k both-working married and cohabiting couples in combined Quarterly Labour Force Surveys, 2010–2012 (e.g. ONS 2017)

features of CAMSIS scales for a dataset with relatively few occupational categories. Figure 4.1 shows data on the nine 'major groups' of the UK's Standard Occupational Classification (SOC) 2010 (ONS 2010). For illustrative purposes, we have presented the major groups in an unorthodox order, from the least to the most populous amongst males in the UK. The plot shows a representation of the relative number of cases occurring for those married and cohabiting heterosexual couples in the UK when both individuals are currently in work. The size of each mark is proportional to the frequency of occurrences, with the lower axis showing the jobs of the males and the vertical axis that of the females using the same order to the groups (e.g. the most frequent combination is that in the lower right cell, between males and females who are both in 'professional occupations'; the third row is relatively populous, representing couples where the female is in the category 'Administrative and secretarial occupations', whilst the corresponding third column, for couples where the male is in this category, is relatively sparse). For argument's sake, imagine that there is no known hierarchy or structure to these nine categories.[1]

What structural patterns could we find out about occupations in the UK on the basis of this information about the distribution of couples? There are some calculations that we could immediately generate from the data in hand, such as the proportion of men or women in each job or, using the information on social connections, the proportion of cases where the spouse is found in the same job category (i.e. the relative 'endogamy' of the occupation). We have placed some such results within Table 4.1 (look first at columns 4 and 5). As is typical of occupational classifications, the statistics show evidence of quite strong gender segregation in occupations, as some categories are predominantly male and others are predominantly female. We also have evidence of a considerable degree of endogamy, at its highest when 43% of the working spouses of men from the Major Group of 'Professional occupations' are in the same group.

We could use additional information from the survey in order to calculate summary data about the occupations, for instance, finding the average income or educational profiles of incumbents of each occupation. There are some relevant results in columns 6 and 7 of Table 4.1, and they show quite stark differences between the major groups. Many other values might be computed in a similar manner if we had access to sufficiently rich data.[2] However, data is not always available in a consistent and standardised format. In addition, it is sometimes correlated in complicated ways to other factors (for instance, educational qualifications relate strongly to birth cohort, and income is strongly related to age and gender—so occupational profiles based on income or educational levels are also influenced by demographic compositional patterns).

As an alternative, we could characterise the structure of occupational groups using data on the frequency of social connections between occupations. Several different statistical devices can be used for this purpose. Approaches within the CAMSIS tradition use 'dimensional reduction' statistical techniques, that is, those that seek to identify underlying continuous dimensions of occupational structure that show some concord with the observed empirical patterns of social connections between the

**Table 4.1** SOC2010 major groups from the UK, with CAMSIS scores and other selected data

| | CAMSIS scores | | Other data | | | | |
|---|---|---|---|---|---|---|---|
| | Dim 1 (M) | Dim 1 (F) | Dim 2 (M) | % of all males; % of all females | % females in job; % males with partner in same group | Education score; % with degree | Mean (FT) weekly income (£) |
| 1. Caring, leisure, and other service occupations | 44.9 | 39.8 | 40.0 | 2.7; 15.2 | 85.0; 29.4 | −0.21; 11 | 246 |
| 2. Sales and customer service occupations | 38.0 | 35.0 | 49.2 | 3.8; 9.5 | 71.7; 22.2 | −0.23; 11 | 272 |
| 3. Administrative and secretarial occupations | 51.2 | 49.7 | 53.9 | 4.2; 20.0 | 82.6; 26.6 | −0.06; 17 | 304 |
| 4. Elementary occupations | 26.1 | 23.3 | 38.2 | 8.6; 9.3 | 51.8; 23.5 | −0.45; 6 | 251 |
| 5. Process, plant, and machine operatives | 30.8 | 22.7 | 46.7 | 10.7; 1.7 | 13.3; 4.6 | −0.53; 3 | 326 |
| 6. Managers, directors, and senior officials | 57.6 | 59.2 | 80.9 | 14.8; 7.9 | 34.7; 17.6 | 0.34; 35 | 593 |
| 7. Associate professional and technical occupations | 58.5 | 59.5 | 54.4 | 15.6; 12.5 | 44.6; 19.1 | 0.43; 36 | 455 |
| 8. Skilled trades occupations | 39.8 | 34.9 | 49.0 | 19.7; 2.1 | 9.5; 3.9 | −0.47; 5 | 345 |
| 9. Professional occupations | 71.2 | 69.9 | 32.1 | 19.9; 21.9 | 52.4; 42.8 | 1.15; 67 | 508 |

Source: UK Labour Force Survey (e.g. ONS 2017), subsample of both-working cohabiting couples, 2010–2012
Education score = mean value of standardised scale for highest educational qualification, based on average age of leaving education per qualification category

occupational categories. In these approaches, scores are derived for occupational units in one or more dimensions of difference, and the score given to an occupational unit helps us to establish its relative 'social distance' from any other occupation (within the appropriate dimension).

Several related statistical techniques can identify and calculate dimension scores in this way, but in the recent past, the technique of 'correspondence analysis' has most commonly been used to construct CAMSIS scale scores. Correspondence analysis is quite widely used as an exploratory statistical technique across the social sciences (e.g. Greenacre and Blasius 1994). One of its contributions is to calculate scores for categories in such a way that the scores reflect the frequency of cases falling into each cell of a relational table involving the categories. In the CAMSIS framework, this involves finding scores for occupational units which improve the prediction of the distribution of social connections between occupations. Scores can be calculated in one or more 'dimensions' of difference between the categories, but usually in the CAMSIS approach it is only the first and most statistically influential dimension that is of interest. Through such means the CAMSIS approach generates scores for occupational units that we refer to as measures of the 'social interaction distance' between occupations.[3] The epistemological position is that those scores tell us something interesting about the occupational and social structure.

We have placed the (rescaled) scores obtained from correspondence analysis of the relationship between the occupational major groups of husbands and wives in the first columns of Table 4.1, and a graphical representation of these scores is shown in Fig. 4.2. Scores in the graph are plotted for the two strongest empirical dimensions (dimension 1 accounts in this case for 71% of all the variation in social interaction patterns which can be modelled in this way). The usual approach is to interpret the scores in terms of the categories that they correspond to—dimension 1, for example, seems to indicate a gradation from groups 9 ('Professional occupations') and then 6 and 7 ('Managers, directors, and senior officials' and 'Associate professional and technical occupations'), moving through a mid-range of categories, to end with occupational groups 4 and 5 ('Elementary occupations' and 'Process, plant, and machine operatives') at the other extreme. To reiterate, the result is sug-

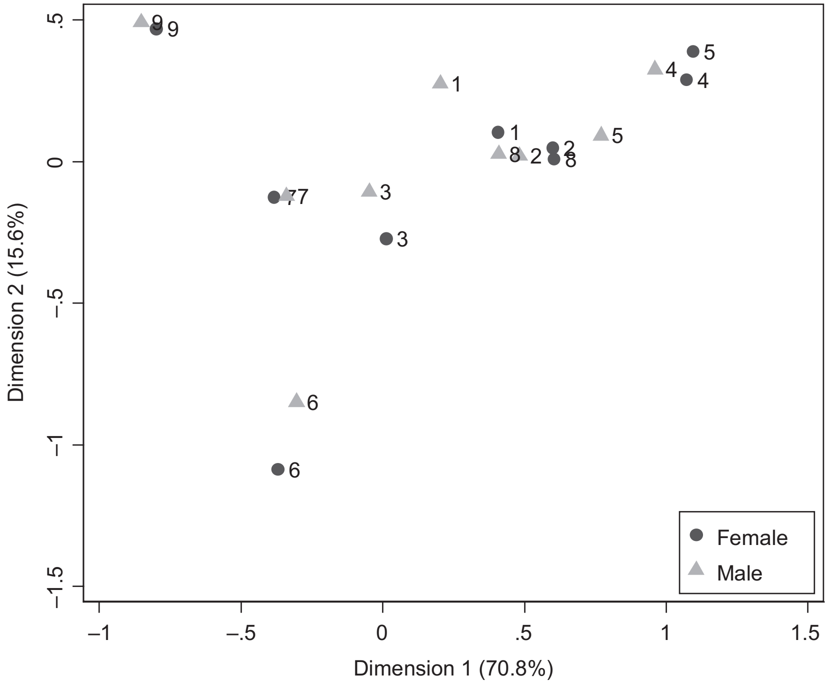



**Fig. 4.2** Correspondence analysis dimension scores for the cross-tabulation of husbands' and wives' jobs, UK, SOC2010 major groups. Source: Correspondence analysis 'biplot' (dimensions 1 and 2), data for 67k couples sampled from the UK Labour Force Survey, 2010–2012

gesting that, for example, husbands from group 9 are relatively more likely to be married to wives in group 9, or 6 and 7, but are relatively less likely to be married to wives from groups 4 and 5, and vice versa. Re-expressed, the 'social interaction distance' in dimension 1 between groups 9, 6, and 7 is low, but the distance between group 9 and groups 4 and 5 is much greater. By most interpretations, dimension 1 would seem to reflect some measure of the social advantage of occupations (note how it has a similar pattern to the scores based on income, and especially those based on education, shown in Table 4.1). Accordingly, as is typical in the CAMSIS approach, we could conclude that the major dimension of social distance between occupations in this analysis seems to be one of 'social stratification' or 'social advantage'. (In this example, the dimension scores given to husbands' occupational units and those given to wives' occupational units are more or less symmetrical, but the methodology does not impose this—see also Sect. 6.9.)

We refer to the set of scores generated through this approach as social interaction distance scales, because differences in scores between units are indicative of the relative social distance between the categories, in terms of the frequency of social interactions. Conventionally, in the CAMSIS approach, the scores on the first dimension of the correspondence analysis are taken as the CAMSIS scale and interpreted as indicators of the structure of social stratification (with the caveat that the analyst is satisfied that the first dimensional structure of social interaction distance does indeed primarily reflect a structure of 'social stratification').

The 'dimensional reduction' character of correspondence analysis and other techniques which are similar to it can be thought of intuitively as a sequential approach that starts with a question: If I were trying to capture as much of the variation in interaction patterns as possible within a single continuous dimension of difference, what structure would that dimension have?[4] Algorithms are used to find the best-fitting first dimension in this way, but the dimension will inevitably only capture a certain proportion of the interaction patterns (this proportion corresponds to the percentage listed in Fig. 4.2). Subsequently, the algorithm then asks what the next most effective explanatory dimension will be, setting aside the patterns linked to the first dimension. This analysis generates a second dimension, then a third dimension, and so forth—ultimately, we could generate as many dimensions as there are categories in the measures involved, but usually only the first few, most influential, dimensions are of interest. The attraction of this sort of calculation is that the dimensions may provide us with convenient, parsimonious characterisations of a more complex reality: it is accurate, but not parsimonious, to note that there are (say) 400 different occupational unit groups and that each has a slightly different pattern of social interaction connections to other occupations; it is parsimonious, though it incorporates some approximation, to say that a substantial component of the social interaction connections between occupations can be differentiated around a single dimension of difference between occupations which can then be described in terms of an occupational score.

Social interaction distance scale scores such as those shown in Fig. 4.2 and Table 4.1 seem to reflect a structure which is related to social stratification and inequality (i.e. dimension 1 seems to show a ranking from the most to the least advantaged groups). It is worth re-emphasising that nothing else went into the calculation of the scores, other than data on the volume of social interactions. The fact that the scores generate a persuasive description of the social structure, which corresponds quite closely to the characterisation that we would see in terms of measures such as average income or educational level, seems an impressive outcome. Indeed, a compelling motivation for using the CAMSIS approach is as an exploratory empirical tool for mapping the occupational structure in a society. Moreover, since it is reasonably easy to access data about social connections between occupations for different societies, it is appealing to use social interaction distance analysis as a tool for comparative research (Prandy and Jones 2001).

For some writers, the social interaction distance approach to exploring social inequalities is also, theoretically, a better means of understanding social and occupational structure, when compared with alternative ways of summarising social inequality. Proponents of the social interaction distance approach argue that this occurs because of the centrality of social relations to the social reproduction of an order of social stratification. If social stratification is conceptualised as the very thing which emerges from the long-term reproduction of consequential inequalities through social relations, then a measure constructed around the contours of social relations should come closest to accurately depicting social stratification itself (especially Bottero 2005a).

## 4.2 History of Social Interaction Distance Approaches

Social connections between the incumbents of occupations have long been recognised as socially significant. The first example of inference from social interaction patterns involving occupations, to social stratification or social inequality, is commonly attributed to Laumann and Guttman (1966). In their seminal paper, using a range of 55 occupational titles from the USA, Laumann and Guttman documented how an emergent statistical tool—smallest space analysis—could be used to rank the occupations recorded for male interview respondents in terms of patterns in the occupations help by people with whom the respondents held social connections. The analysis was undertaken in terms of the relationship between respondents and up to seven alters (viz. the respondent's father, father-in-law, three closest friends, and two neighbours). In findings which would prove to be remarkably consistent with those of subsequent analyses in the field, Laumann and Guttman found it useful to depict the structure of social associations in terms of broadly continuous dimensions of difference, the strongest of which was interpreted as a hierarchical structure of social inequality (labelled 'prestige'). Two other prominent dimensions to social association patterns between occupations were found, but Laumann and Gutmann, whilst speculating that these might reflect employment status or a related concept of 'situs', argued that on balance these could not be unambiguously interpreted.

A sequence of studies linked to the Cambridge group and their colleagues—what we are referring to as the CAMSIS tradition—focussed first upon social interactions in survey reports of friendship patterns, before more recently arguing that other forms of social interaction data, particularly marriage and cohabitation patterns, reveal the same structure, and thereafter predominantly using cohabitation data. In the Cambridge group's most influential studies, Stewart et al. (1973, 1980) demonstrated that a dimension of social interaction distance between jobs could be extracted from data on friendship patterns, and had the characteristics of a dimension of social stratification. The dimension was originally derived for the UK's 1970 occupational categories, and this dimension became widely known as the 'Cambridge scale'. In his 1990 paper, Prandy undertook reanalysis of similar data for 1970 and 1980 occupational categories, addressing criticisms that had been raised about the original scale regarding its focus on men, and its basis on a relatively small, geographically and occupationally selective sample (cf. Heath 1981). Prandy (1990) exploited additional survey datasets and alternative statistical models, and also used tools for allocating scores to new occupational categories that had not featured in the original Cambridge Scale analyses. His revised scale had broadly the same properties as the original scale, and he presented a series of validity analyses which demonstrated its favourable features as a measure of social stratification.

Further work in the CAMSIS tradition from the 1990s onwards has been characterised by a widening of the application area of SID analyses across countries and time periods, and increasingly by using data on marriage/cohabitation records rather than friendship patterns (e.g. Prandy and Lambert 2003; Prandy and Jones 2001; Prandy and Bottero 1998). Despite Prandy's (1990) validity analysis, some critics expressed unease about the relatively small and regionally biased sampling behind the friendship data used for the original Cambridge scale in the UK (e.g. Chan and Goldthorpe 2004; Mills and Evans 2003), but a by-product of comparative analyses in this period was evidence that social interaction distance procedures were apparently quite robust to the quality of the underlying microdata. For instance, researchers in the CAMSIS tradition argued that more and less representative samples, and data based upon various types of social interaction connections (e.g. friendship, marriage, intergenerational mobility, and intragenerational mobility), when analysed consistently, would all be likely to reveal the same underlying structure of social interaction distance between occupations. This is a consequential conclusion since it means that numerous examples of structural analysis applied to intergenerational and intragenerational data—for instance, within studies of social mobility—may also be thought of as versions of social interaction distance analysis. Moreover, as the CAMSIS project developed this position, a number of other empirical studies were undertaken which also used social interaction distance analyses of occupations in the same manner (e.g. Savage et al. 2015; de Luca et al. 2010; Chan 2010a; Chan and Goldthorpe 2004). Although conducted largely independently, these studies found the same key dimension of social interaction distance in occupations and, in doing so, further demonstrated that the SID approach to analysing occupations was fairly robust to sampling variations and differences in the form of the underlying social relationships analysed.

The earlier publications in the CAMSIS tradition (especially Prandy 1990; Stewart et al. 1973, 1980) arguably made three important contributions to research on social stratification. First, the work demonstrated the importance of social interaction patterns to studying inequalities and pioneered the use of statistical methods for drawing out the empirical relationship between social interactions and the stratification structure. Second, the work made consequential methodological contributions—documenting techniques for calculating scale scores, demonstrating the value of scaling occupational positions at a relatively fine level of detail, and demonstrating that a variety of different measures of social distance could be used to provide similar results (e.g. Prandy 1990). Lastly, and arguably of most importance, the authors developed a theoretical interpretation of the structures emergent from analysis of social interaction distance and demonstrated their empirical centrality to social processes linked to social stratification (for follow-up work, see especially Bottero 2009, 2005a, b; Prandy 2002, 2000; Prandy and Blackburn 1997).

Two methodological positions taken in the CAMSIS tradition differentiate it from most other examples of social interaction distance analysis applied to occupations. One is that 'specificity' is usually advocated, by estimating different scales for different countries, time periods, and for men and women, rather than calculating the same scale scores for the same occupations across different contexts. The majority of SID scales linked to the CAMSIS project apply to 'specific' countries and time periods. Counter-examples, however, include the international version of the CAMSIS scale ('I-CAM'), advocated by de Luca et al. (2010), which is based upon pooled cross-national survey data using internationally harmonised occupational unit groups, and the 'universal' version of the HISCAM scale for historical occupational titles (Lambert et al. 2013). We discuss the appeal of 'specificity' in Sect. 5.1.4.

A second distinctive feature of CAMSIS scales has been the commitment to using fine levels of occupational detail when possible. The scales released under the CAMSIS project have been characterised by a high volume of detail (typically around 100–500 different occupational units are scaled). Nevertheless, the relative pay-off to working with finer levels of occupational detail is open to debate. Analyses typically generate some examples where two occupational units are afforded very different scores even though, at a more aggregate level, they would have been placed in the same categories. Indeed, Rytina's work scaling intergenerational mobility patterns (1992, 2000) made a central argument that disaggregate measures of occupations were necessary to adequately depict intergenerational mobility patterns, and Bakker's inductive analysis from the Netherlands (1993) similarly found value in describing a large number of occupational categories. However, using many occupational categories leads to a great deal of work in coding and processing occupational titles, and the inclusion of more units of analysis increases the statistical uncertainty associated with occupation-specific estimates (because some categories are represented by low numbers of cases). Accordingly, many other applications of social interaction distance analysis involving occupations have used relatively few occupational units (e.g. Chan 2010b; de Luca et al. 2010).

Aside from the CAMSIS tradition, several other programmes of research using social interaction distance scales for occupations are worth commenting upon. Bakker (1993) published results of a similar correspondence analysis for data from the Netherlands. Bakker's analysis led to a characterisation of a dimension of what he interpreted as 'social status'. He noted that the scales of social status that were linked to occupations were slightly different for men and women, whereby the relative proportion of women in an occupation tended to be related to gender differences in status.

In an extended programme of analysis using American data, Levine and colleagues (Levine 1972, 1990; Levine and Spadaro 1988) used log-linear association models to characterise dimensional structures within data on both intra- and intergenerational occupational mobilities. Levine's analyses pointed to a multidimensional structure of social interaction patterns, of which the two most important dimensions reflected social stratification (labelled 'status') and other industrial and organisational structures (labelled 'class'). Like Laumann and Guttman, Levine stressed the inductive value of this approach for understanding the character of social inequalities as they relate to occupations—Levine and Spadaro (1988, p. 453) likened the approach to a medical X-ray where trace compounds are mapped as they move through the body, arguing that their analysis provided a similar trace of the 'shadow' of social movements through occupational positions.

Rytina (1992, 2000) performed a similar analysis of intergenerational mobility patterns using detailed occupational data from the UK and USA, respectively. Rytina characterised a number of factors which influenced the propensity for specific father-child occupational combinations to occur and again found that a major structuring force was a single dimensional structure to occupations which he labelled as 'hierarchy'. Rytina's analysis stressed that other factors also influenced mobility patterns, in particular precise occupational inheritance, but also other sectoral affinities in jobs (which he labelled as 'class'). His analysis stressed the centrality of a single continuous dimension to occupational differences and also argued for the importance for studying highly disaggregate occupational positions (in order to maximise the detection of inheritance and class effects).

In a series of recent studies which share many tenets with Rytina's analysis, several writers have used 'microclass' occupational units as the background for a complex log-linear model for an intergenerational mobility table involving movement between the occupational categories (e.g. Jonsson et al. 2009). These applications involve quite large numbers of occupational microclasses (typically around 100) and, within wider analyses, have reported that SID scales based upon mobility patterns capture an important component of the mobility process. In this approach, it is argued that there are several different aspects of the social structure of which a hierarchical ranking of occupations is only one (especially Grusky and Weeden 2006). In addition, an intriguing possibility is explored in Grusky and Weeden's (2006) review whereby a latent class model could first be used to define the occupational categories themselves (on the basis of their mobility patterns), and then scaling or ranking is performed (again on the basis of empirical patterns of mobility) upon the latent classes (cf. Rost 1988). This scenario also introduces the unusual example whereby the boundaries of the detailed occupational categories may themselves be negotiated within the analysis—in most other applications, the occupational units are defined *a priori* during the data construction and planning phases of analysis.

Another important group of studies using social interaction distance analyses are those described in a series of publications related to Chan and Goldthorpe's (2004, 2007) presentation of a social interaction distance scale. The original analysis for the UK was supplemented with comparable exercises in other countries as documented for six nations by Chan (2010b), for Russia by Bessudnov (2012), and for Norway by Chan et al. (2011), and the scales derived have been exploited in a number of recently influential sociological papers (e.g. Torssander and Erikson 2010; Goldthorpe 2010). In this approach, friendship and marriage patterns between occupations have been analysed through log-linear association modelling and multidimensional scaling approaches to reveal dimensions of difference between occupations. Gender segregation has been highlighted as an apparent secondary dimension in many of these studies, but the first dimension of social interaction patterns is consistently reported as an hierarchical measure of inequality that has been interpreted as a distinctive measure of 'social status'.

Driven forward particularly by Chan (2010a), the research in this programme developed new software routines and conventions of calculation, including the construction of standard errors to describe uncertainty about scale scores. In most examples, the scales associated with this approach have been calculated on a relatively low number of occupational units, with equal scores for men and women. Both of these decisions may generally be expected to lead to statistical results with greater reliability, but with possible costs in the validity of the information retrieved. The UK scale, for example, has a relatively strong relationship to gender, and the broad categories of the scheme involve combining some occupations that are very different in their gender profile.

Whilst the studies described above focus on those which have used an exploratory approach to summarising a fairly large number of occupations, several other studies have estimated dimension structures for social mobility and social interaction patterns in a similar manner, but involving a narrower range of occupations and without the inductive principles associated with Levine, Lauman and Guttman, Rytina, Chan, and the CAMSIS tradition. Savage et al. (2015, p. 138ff) performed a similar analysis on a circumscribed range of occupations that were listed in a 'position generator' tool. Separately, many studies have accommodated occupational scores within log-linear models for occupational tables. In these applications, row and column scores are estimated for an occupational table, typically based upon inter- or intragenerational mobility (e.g. Wong 2010; Luijkx 1994; McDonald 1972; Blau and Duncan 1967; Centers 1949). Here the scales are regarded as a means of more fully accounting for mobility, homogamy or homophily, rather than as a direct attempt to find out about the occupational structure itself. Separately, there have also been a small number of analyses that have applied the strategies of the SID tradition to non-occupational categories, for instance, in the analyses of marriages between ethnic and religious groups (Prandy 1979; Laumann 1973) and of social connections between housing tenure categories (Prandy 1979) and educational qualifications (Lambert 2012). We elaborate on some of these in Chaps. 10 (on education) and 11 (in discussing ethnicity); however, in broad summary, applications of SID techniques to non-occupational units have not hitherto been as revealing as might be anticipated, apparently because there are too many overlapping socio-demographic characteristics associated with non-occupational categories.

## 4.3 Theorising CAMSIS Scales

It is clear that SID analyses reveal structural patterns in social inequalities between occupations, but why should this be so and what might they tell us about the society being studied? The first point to note is that more than one-dimensional structure of difference can be identified. Later, in Chap. 11, we discuss interpretations of other 'subsidiary' dimensions to the SID solution, but here we will concentrate only on the empirically most important dimension. That dimension is usually given a post hoc interpretation based upon its properties: it correlates strongly with other socio-economic measures of occupations and their qualities and has variously been described as 'stratification', 'generalised advantage', 'social status', 'socio-economic status' (SES), and 'class'. The dimension scores shown in Fig. 2.2 and Table 2.1, for instance, are typical of the scores given to occupations in the first dimension when using SID methods.

### 4.3.1 Reflecting Social Reproduction

Ultimately we believe that the structure revealed through a SID analysis of occupations is a depiction of an order of socially reproduced consequential inequalities. The suggestion is that the structures revealed represent the 'trace of social reproduction', a structure linked to social stability and that reveals contemporary inequalities. Our view is much the same as the position developed over a number of publications by authors linked to the Cambridge Scale and the CAMSIS project, where the term 'social stratification' is generally taken as a best available shorthand description for the social phenomena that the first dimension of a SID scale represents (e.g. Bottero et al. 2009; Bottero and Prandy 2003; Stewart et al. 1980).

'Social stratification' is a convenient label for two reasons. First, pragmatically, the term is not generally linked with more specific concepts of inequality in the way that other viable labels may be (cf. 'class', 'status', 'prestige'). The label 'social stratification' is arguably more neutral, and is often understood as referring in a generic way to a structure of consequential social inequality, without a more specific implication about its character. Second, the term 'social stratification' is often associated with structures of inequality that show some persistence and reproduction through time (e.g. Bottero 2005a; Kerbo 2003). Indeed, the connection between 'social stratification' and 'social reproduction' is one that authors from the CAMSIS tradition have increasingly made central to their explanation of why SID analyses reveal the structures that they do (especially Bottero 2009, 2005a; Prandy 2002). The argument is that the social connections that people make are important mechanisms through which people attain stability and reproduction in their circumstances. When a structural pattern pervades those social connections, it logically reflects the structure of social reproduction. This implies in turn that the structure of SID scales reflects the empirical structure of social reproduction as related to consequential inequalities—a reasonable label for which is the structure of social stratification.

In recent years, several articles by Bottero have perhaps given the most compelling explications of what the structure revealed by SID is capturing (2005a, b, 2009; Bottero et al. 2009; Bottero and Prandy 2003). Bottero argues that the social distance between occupational units is defined by an 'interaction space' that constitutes a social space of inequality. The interaction space is arranged through the combination of multiple individual-level negotiations that are influenced by individuals' awareness of their own circumstances in the stratification structure, their understanding of the circumstances of the people with whom they might interact, and by the impact of wider structural constraints upon their interactions. For these reasons, people disproportionately interact in patterns of stability and reproduction—they typically make connections with people in similar situations, and they adapt to and develop a preference for the familiar and stable. Of course not all individuals are constrained to stability all of the time, but a SID analysis nevertheless teases out the patterns connected to the disproportional tendency towards social reproduction in the interaction space.

This account has many resonances with the argument often associated with Weber, that patterns of intergenerational and/or intragenerational social mobility could be thought of as defining the stratification (or 'social class') structure (e.g. Toubol and Larsen 2017; Breiger 1981; Weber 1968[1922]). The social interaction distance approach has often been applied to inter- and intragenerational social mobility data, and has revealed similar structures of inequality as are evident from the analysis of social interactions of friendship and marriage (e.g. Lambert et al. 2013; Rytina 2000; Levine and Spadaro 1988).[5] We would argue that mobility data can reasonably be presented as information on a social connection (cf. Chap. 2), and that a dimensional structure revealed through SID analysis of mobility data should indeed be conceptualised as the same structure as would be revealed through the analysis of friendship or partnership connections (i.e. it is demonstrably the same structure empirically; our argument is that it is also the same structure conceptually).

SID scales might also better reflect how social positions can be linked to longitudinal trajectories in occupational locations (e.g. Bottero 2005b; Stewart et al. 1980). In many instances, standard career trajectories move people from one occupational unit to another during the normal working life. An engineer might progress over 30 years from being an 'assistant operative', to an 'operative', then a 'supervisor', then a 'local manager'; a person working in a university setting might successively be a 'teaching assistant', 'research assistant', 'lecturer', and 'professor'. In such examples, individuals might experience considerable changes in their objective circumstances as their career progresses (such as in terms of disposable income or home-ownership status). However, their wider social circumstances—the community in which they live, the holidays they take, and the friends and family that they connect with—will probably not change very much.

If we wish to characterise the occupational position, we can either focus only upon the current objective circumstances of its incumbents or we could take some account of typical career contexts. Of course, career progressions do not divide occupations perfectly (for instance, a category of academic research assistants will include many who will go on to become professors, but also many who will not). Nevertheless, there is a case for characterising an occupation according to the profile of its incumbents in a way that is informed by information about average career trajectories. Arguably, measures based upon social interaction patterns do this neatly, because they profile the occupation based only upon the social interaction patterns of its incumbents (e.g. Stewart et al. 1980[6]). By contrast, most other occupation-based measures use methods that can only characterise occupations according to the objective economic circumstances of their current incumbents, meaning that when individuals make standard career transitions, they often change, accordingly, in terms of their occupation-based social class or stratification scale score. The choice here is a conceptual one of whether or not we would regard jobs which involve similar people, but at different stages of their career, as occupying largely the same social positions (in spite of average differences in economic circumstances). Profiling on the basis of social interaction patterns will do this to a considerable extent, whereas alternative popular occupation-based measures do not.

Empirical assessments of the correlations between CAMSIS measures and other things are also consistent with its interpretation as a measure of stratification that reflects the 'trace of social reproduction'. SID scales exhibit moderate correlations to other outcomes that are expected to relate to social inequality, and are comparable in scale to those of other measures of stratification (e.g. de Luca et al. 2010; Prandy 1990, 1998, 1999, 2000). Compared against other occupation-based schemes, SID scales are generally agreed to have stronger correlations to measures of education, lifestyle, health and economic security than are other occupation-based measures, but they typically have somewhat weaker correlations to income and to other immediate features of economic circumstances (e.g. Lambert and Bihagen 2014; Chan and Goldthorpe 2007). The latter difference is sometimes used to argue that SID scales represent a more specific concept than the general structure of social stratification (see Sect. 4.3.3). However, the pattern also makes sense when SID scales are thought of as reflecting long-term positions in the inequality structure—because qualities such as lifestyle and cultural orientations might be seen as intrinsically more strongly correlated to socially embedded inequalities. By contrast, measures of current conditions could be said to reflect correlated epiphenomena (such as income) that change in standardised ways through the life course.

### 4.3.2 Interpreting SID Scales From a Social Networks Perspective

The CAMSIS approach, of identifying structure from interaction patterns, is very much in keeping with the tradition of 'structural analysis' linked to social network analysis (SNA) (e.g. Wellman and Berkowitz 1988). Moreover, statistical tools such as correspondence analysis when applied to data on social connections, which are central to the CAMSIS approach, are sometimes claimed to be intrinsically SNA techniques (e.g. Wasserman and Faust 1994). Historically, however, there have been relatively few exchanges between academic research literatures on social interaction distance analysis of the relationships between occupations and those on approaches to social network analysis (but cf. Griffiths and Lambert 2012; Levine and Spadaro 1988). The limited cross-fertilisation may well have arisen simply because most projects that have used SID have origins in sociological evaluations of occupation-based measures, rather than being conceived of as explorations of network structures. It might also reflect differences in the character of the social connections that the two traditions focus on—in network analysis, researchers are often interested in more distant social connections such as in charting the benefits of 'weak ties' (e.g. Granovetter 1973); for SID projects, by contrast, the focus is usually on close personal ties such as of friendship or marriage. In later sections (Chaps. 7 and 8), we discuss how the same social interaction data that underlies a SID analysis can be alternatively summarised using approaches that are associated with SNA literatures. First, however, it is useful to reflect on how the CAMSIS tradition of SID analysis might be interpreted if approached for the traditional frameworks of research on social networks.

At its root, the social network analysis 'paradigm' explores the form and relative importance of network connections to social outcomes. From this framework, aspects of the social network analysis tradition might provide additional insight into the main results of a SID analysis (viz. the dimension of social stratification as it maps onto occupations). As one example, the concept of latent ties refers to connections that people potentially have but do not currently operate (Haythornthwaite 2002)—for instance, links through mutual friends that could be called upon if needed. Latent ties often operate through occupations, when, for example, the social connections of an individual's work colleagues might provide advice, support, or guidance. One reason that the first dimension of the SID solution represents an important social inequality structure, therefore, might be that it represents the realisation of resources through latent ties—that is, the most advantaged occupations might be those characterised with social connections that offer the most valuable social resources via realised, but also via latent, ties.

The connection from ties to social inequalities prompts comparisons with those literatures on social inequalities that have most influenced social network theorisations. Bourdieu's (1984) attention to the benefits of social and cultural capital, for example, suggest a mechanism from social ties to consequential inequalities in the distribution of resources. In the social network analysis paradigm, this is important because it illustrates the importance of social connections in shaping social inequality. Recent attempts in the UK have sought to create a social class scheme which incorporates cultural, social and economic resources (Savage et al. 2013, 2015). Whilst the measurement itself is not readily operationalised upon existing datasets (Mills 2014), CAMSIS scales, arguably, provide a comparable (and more readily operationalised) measurement instrument for reflecting people's networks, because they aggregate information at the occupational level which is influenced by the profiles of ties that are connected to the occupations.[7] By this reasoning, CAMSIS scales might reflect a stratification structure precisely because they proxy social and cultural capital inequalities.

The SNA tradition has also generated a separate measurement tool that has many overlaps to SID studies. 'Position generators' are designed to estimate the social capital an individual possesses by asking participants whether they know someone from a range of occupations and circumstances—thereafter summarising the range of positions known and advantages linked to those connections (e.g. Verhaeghe et al. 2012; Van der Gaag et al. 2008). Arguably these tools reflect an individualised version of the data from which SID scales are generated and could be expected to lead to similar results (Griffiths and Lambert 2015; Savage et al. 2015, p. 138ff). These tools are designed to assess the range of resources that individuals might benefit from; their overlap with other forms of SID analysis further supports an interpretation of SID scales as qualities that reflect social capital distributions (Griffiths and Lambert 2015).

In summary, the social network analysis framework might be expected to portray the first SID dimension as a reflection of the relative distribution of social resources (social and cultural capitals) which are on average held by individuals within the relevant occupations. Our own view is that this interpretation is also consistent with a wider theorisation of the social structure of reproduction associated with social interactions and occupations—what we labelled the 'social resin'. That is, an emphasis on the reproduction of stratification systems through social connections fits consistently with patterns of unequal access to social resources.

### 4.3.3 What the Principal SID Dimension Is Not

Two important debates in the academic literature concern whether or not SID scales reflect other recognised concepts than those that we have highlighted. Despite some expectations or claims to the contrary, we maintain that a SID scale primarily reflects neither preferences for interaction nor an order of recognised hierarchical 'social status'.

#### 4.3.3.1 Not 'Preferences'

Individuals' expressed preferences for their social relationships involving occupations have been studied occasionally, such as by using hypothetical questions in survey studies (e.g. Laumann 1966; Laumann and Senter 1976). A typical wording asks for agreement that 'I believe I would like to have a [*carpenter*] as … my son-in-law: {strongly agree/agree /undecided/

disagree /strongly disagree}' (Laumann and Senter 1976, p. 1315). From such data, 'social distance scores' for occupations have been constructed by profiling the responses according to occupations. Attitudes reveal an aspirational component, whereby respondents generally favour connections to more socially advantaged occupations (Laumann and Senter 1976 argue that this reveals a 'competitive status consciousness'). However, attitudes are also tempered by circumstances: increasing openness is shown to less-advantaged jobs by those who are themselves the incumbents of less-advantaged jobs, which suggests that the ranking to occupational categories is also shaped by realistic preferences (e.g. Laumann and Senter 1976). Such studies reveal that a gradational order of social preferences for interaction is very similar empirically to the gradational order revealed through behavioural social interactions.

If the gradational ranking of preferences, and that of actual behaviours, is of much the same character, might it be preferable to collect and analyse data on preferences, rather than data on realised social interactions between occupations? This might not enhance the efficiency and reliability of data, since it is relatively easy to obtain data on actual social interactions between occupations, whereas to collect data on preferences for interactions may require a lengthy survey, and set a demanding cognitive task that could be prone to bias (e.g. Coxon and Jones 1978). We can also anticipate measurement error if the connections between occupations based upon expressed preferences for interactions are unrealistic, or show misunderstanding of occupational titles. For such reasons, some of the fine-grained differences between the rankings of occupations in SID measures may not be identical to those from a preference-based evaluation. One example are skilled manual jobs, which may be ranked lower in the behavioural context (i.e. using SID), when compared to their placement based upon expressed preferences. These could reflect genuine differences between popular beliefs about occupational circumstances (captured through preference expressions) and the underlying social positions of occupations (captured by SID analyses). Accordingly, it seems likely that the structure of preferences for social connections reflects a mixture of influences that include aspects of measurement error and popular myths about occupations, resulting in small departures from the structure defined by social interaction patterns alone.

#### 4.3.3.2 Not 'Status'

A second and widely held interpretation of SID scores is as measures of status. The word 'status' itself often has a different meaning in different literatures. Here we discuss 'status' in its classical sociological sense, as reflecting relations of perceived social superiority and inferiority, or agreed social honour (e.g. Chan and Goldthorpe 2007). 'Status' is also sometimes used as a generic reference to a social position—for example, a measure of SES might refer to any structure of inequality defined by socio-economic circumstances. 'SES scales' in sociology are often understood as measures that calculate average socio-economic conditions (e.g. incomes, educational qualifications, wage progression) for each occupation.

Chan and colleagues (Chan 2010a; Chan and Goldthorpe 2004, 2007) have interpreted the main dimension from a SID analysis using a classical definition of status as agreed social honour. This position has a neat logic to it: Chan and Goldthorpe argue that social connections intrinsically reflect mutual understanding of status, social honour, and deference, and thus by definition, a measure calculated using social interaction data must reflect the structure defined by status, social honour, and deference. Chan and Goldthorpe validate this position through empirical evidence—the SID measure is seen to have properties which would be consistent with a measure of status, such as giving high rankings to professional jobs and those that require extended qualifications, and low rankings to most forms of manual work (Chan and Goldthorpe 2004, 2007). Additionally, the measure is shown to be empirically distinguishable from an occupation-based measure of social class in a manner that is consistent with the theory: the SID dimension correlates more strongly with things that are expected to relate to status, such as cultural consumption patterns, and the class measure correlates more strongly with things that are expected to be more linked to class, such as risk of unemployment (e.g. Chan and Goldthorpe 2007). Chan and Goldthorpe's argument has been influential, but we nevertheless draw a different conclusion. On close inspection, theoretical reasoning and empirical evidence both suggest that the dimensional structure revealed by SID scores is something different to that of honorific status. Moreover, as has been argued by several other contributors to the CAMSIS tradition of SID analysis, it is much less feasible to disentangle measures of 'class' from those based on SID, than is portrayed in the account given above.

The operational problem of disentangling SID and class measures has been highlighted in several contexts. Kraaykamp et al. (2010) review the correlation between class and SID measures across a number of datasets for the Netherlands and conclude that the two measures are so strongly correlated that is it not feasible to disentangle the two influences through empirical analysis. Making use of a wide range of sensitivity analyses, Bihagen and Lambert (2012) argue that the empirical differences in the properties of purported measures of status (based upon SID) and of class are not sufficiently consistent to support the hypothesis that the measures reliably disentangle the two concepts. The argument used by Bihagen and Lambert was that if the measures of class and SID are effectively measuring what they are said to measure, then the differences between them for the same people (which can be calculated as residuals in a regression framework) should be characterised by differences on these concepts (i.e. it should be apparent that the highest residuals reflect cases with high status and low class, or vice versa). When those residuals were calculated and analysed, however, Bihagen and Lambert reported that their values were largely unrelated to the theoretical differences between class and status and instead arise from a mixture of other factors, some unexplained, and others apparently things which should not, theoretically, be central to the difference between the concepts of class and status (viz. gender and employment contract status). This evaluation suggests that the empirical differences between measures of class and SID emerge more as a product of the functional form of the two measures[8] and the operational procedures used in constructing them, rather than the theoretical approaches used to inspire them (see also Lambert and Bihagen 2014).

It is perhaps one thing to argue that a SID scale is not particularly effective as a distinctive measure of status, and another to say that SID scales should not be thought of as status at all: the empirical studies above demonstrate the first point, but not the second. There are however several theoretical reasons why a social interaction distance scale might not be directly attributed to status in its classical sense, that is, to agreed social honour. Firstly, social honour itself is likely to be contextually relevant in ways that occupation-based measures are insensitive to. For example, a village policeman, a sports star within a small town, an intimidating gangster, or a works foreman might all enjoy relatively esteemed social positions characterised by the deference of others. However, individuals in such positions are not ordinarily near the top of occupation-based social interaction distance scales. In part this occurs because the reasons for their esteem might not be captured by their occupations. More generally, however, we would claim that the longer-term social relations held by people in these positions may not be as privileged as their localised social honour might suggest—implying that their social interactions are not a good measure of their social status.

Secondly, the empirical structure depicted by a social interaction distance scale frequently generates a number of occupations which are given rankings that are not consistent with a hierarchy based upon social honour. Across published CAMSIS scales, it is not difficult to identify occupations which seem to have a higher or lower relative SID score than we might have expected them to hold in terms of social honour. Examples of the first phenomena are typically jobs which have a reasonably high degree of economic security and long-term economic reward, such as administrative work, accountancy, public sector employment, routine educational work, and skilled engineering work. Such jobs generally score relatively high on SID scales, yet might not necessarily be thought of as of comparably high in 'social status'. Examples of the reverse include jobs which typically have relatively lower economic security or relatively less favourable working conditions despite overt honorific advantage, such as private sector managerial posts, sporting and athletics positions, or positions of political authority. Such occupations typically seem not to have quite such high SID scores than we might have expected if the SID scores did measure status (the UK CAMSIS score for national politician, for instance, is lower than that for 'Medical doctor'—as indeed the secure long-term salary of the position may well be). In such cases, we would argue that there are elements of social differentiation which influence SID scale positions more strongly than they would if SID scales were measures of status. Typically, it seems that differences, such as of long-term economic security, educational advantage, and work task autonomy, impact more substantially upon SID scales than would be plausible if SID scores primarily represented honorific status. These anomalies might arise because such differences correctly reflect the longer-term structure of social reproduction of advantage, but not of social honour.

Another way to compare SID scores with other measures of status is to compare their rankings with rankings of 'prestige' for the same occupations. Although it risks confusing the issue further, concepts of 'prestige', like 'status', are also typically regarded as ultimately reflecting 'social honour' (e.g. Treiman 1977), so a useful comparison can be made between occupation-based scores of prestige and those derived from social interaction distance analysis. Treiman (1977) comprehensively demonstrated the broad consistency of occupation-based prestige structures across societies. Figure 4.3 summarises the relation between CAMSIS scores for occupations, and their prestige scores using Trieman's 'Standard International Occupational Prestige Scale' (SIOPS) for data for the USA in 1990.[9] Although we see general patterns of similarity, there are also

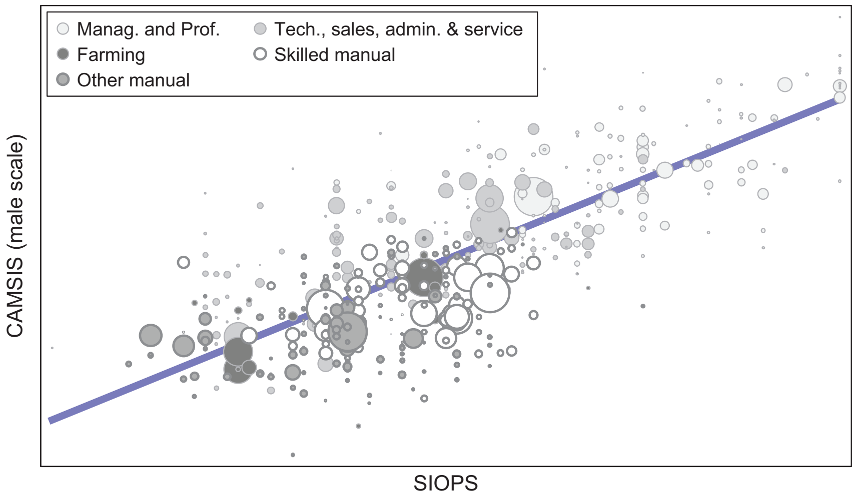



**Fig. 4.3** The relationship between CAMSIS scores and SIOPS for data for the USA, 1990. Source: Units refer to occupations. Size of markers is proportional to the number of men and of women with the occupation in the IPUMS-I 1990 public release sample (MPC 2015)

some systematic examples of occupations with higher or lower relative prestige scores than SID scores. As it happens, those occupations that tend to score relatively higher on prestige than on SID scales are often the same ones that we would, theoretically, expect to be higher on a status scale. Examples include skilled manual and supervisory occupations that may carry social honour but not necessarily long-term economic reward (e.g. the 'skilled manual' occupations in Fig. 4.3). Likewise, those occupations evaluated to a higher relative position on SID than on prestige include many well-rewarded but apparently dull positions (e.g. the administrative and services positions in Fig. 4.3). These were highlighted above as examples of occupations whose incumbents might not be expected to have particularly high social status in its classical sense, but who do enjoy relative social privileges. Similar disparities between prestige measures and SID measures have been noted in other analyses (e.g. Zijdeman and Lambert 2010; De Luca et al. 2010). In our eyes, these results suggest that a prestige measure such as SIOPS is actually a better measure of the classical concept of social status than is a SID-based measure; by corollary, a SID-based measure should not be portrayed as entirely a measure of honorific status.

A last compelling reason for separating SID scales from the concept of honorific social status is that we could anticipate that the construction of social relationships is itself about more than just status in its classical sense. People form and maintain social connections for many different reasons. It is easy to imagine that a sense of status-based appropriate behaviour (or aspiration) is an influence. However, it is less convincing to anticipate that honorific status has any greater role than many other social and psychological forces that shape social connections. For example, individuals are also influenced by their sense of what is 'normal' for their circumstances and a sense of comfort in familiarity and stability (e.g. Archer 2007). They are also driven by socially structured heterogeneities, such as the influence of age cohort in values towards religion or to immigrant origin communities. In light of these numerous influences, the reasoning that SID measures capture honorific status because honorific status drives social interactions seems unconvincing.

## Notes

1. In practice, the UK's SOC major group scheme is substantially hierarchical, since conceptions of 'skill' and of prestige are institutionalised into its taxonomy (e.g. Szreter 1984).
2. For an impressive collection of information of this character on UK occupations, see McKnight and Elias 1997; for online resources incorporating data about the average circumstances of people in different occupations for the USA and Europe, see the projects 'O*NET' (2008) and Wageindicator (2013).
3. Another nice way to describe such scores is a term coined by Laumann and Guttmann in 1966, that such scores depict the 'relative associational contiguity' of occupations.
4. This intuitive framing is quite a close match to the technical procedures that are involved. In the case of correspondence analysis, the scores constructed are derived by solving the matrix of social interaction data for its 'eigenvectors'. Eigenvectors identify vectors that capture the most substantial proportion of the matrix as is possible through a single vector (see Weller and Romney 1990).
5. Although small differences often arise because data on mobility between occupations intrinsically incorporates some influence of time period in the structure, in a way that does not apply to data on other social interactions between occupations.
6. Stewart et al. (1980) discussed a relevant example which was common at the time in the UK, namely, the propensity for the job of 'clerk' to indicate, for males at least, the early stages of a privileged white-collar career.
7. Savage et al. (2015, p. 436) in fact refer to CAMSIS scales as the 'Cambridge social contact scale'.
8. In Chan and Goldthorpe's (2007) comparison, it is presumed *a priori* that status is a continuous dimension and class is categorical. This analysis is inspired by the Weberian distinction of class and status, yet an alternative reading of Weber suggests rather that status is likely to be characterised by categorical divisions and class by continuity (Bihagen and Lambert 2012).
9. The SIOPS scores are linked to ISCO88 units which are in turn linked to US 1990 occupational unit group codes using two separate macros published by Ganzeboom (2016). The CAMSIS scores are the scales for men and for women derived for the US 1990 occupational codes as downloadable from the CAMSIS project website.

## Bibliography


Archer, M. S. (2007). *Making Our Way Through the World: Human Reflexivity and Social Mobility*. Cambridge: Cambridge University Press.

Bakker, B. F. M. (1993). A New Measure of Social Status for Men and Women: The Social Distance Scale. *Netherlands Journal of Social Sciences, 29*, 113–129.

Bergman, M. M., Lambert, P. S., Prandy, K., & Joye, D. (2002). Theorisation, Construction and Validation of a Social Stratification Scale: Cambridge Social Interaction and Stratification Scale (CAMSIS) for Switzerland. *Swiss Journal of Sociology, 28*(1), 441–460.

Bessudnov, A. (2012). A Relational Occupational Scale for Russia. In P. S. Lambert, R. Connelly, R. M. Blackburn, & V. Gayle (Eds.), *Social Stratification: Trends and Processes* (pp. 53–65). Aldershot: Ashgate.

Bihagen, E., & Lambert, P. S. (2012, September, 24–26). *Class and Status: One or Two Forms of Stratification?* Paper Presented at the Equalsoc Conference, Stockholm.

Blau, P. M., & Duncan, O. D. (1967). *The American Occupational Structure*. New York: Wiley.

Bottero, W. (2005a). *Stratification: Social Division and Inequality*. London: Routledge.

Bottero, W. (2005b). Interaction Distance and the Social Meaning of Occupations. *Sociological Review, 53*(2), 56–72.

Bottero, W. (2009). Relationality and Social Interaction. *British Journal of Sociology, 60*(2), 399–420.

Bottero, W., Lambert, P. S., Prandy, K., & McTaggart, S. (2009). Occupational Structures: The Stratification Space of Social Interaction. In K. Robson & C. Sanders (Eds.), *Quantifying Theory: Pierre Bourdieu* (pp. 141–150). Amsterdam: Springer.

Bottero, W., & Prandy, K. (2003). Social Interaction Distance and Stratification. *British Journal of Sociology, 54*(2), 177–197.

Bourdieu, P. (1984). *Distinction*. London: Routledge and Kegan Paul.

Breiger, R. L. (1981). The Social Class Structure of Occupational Mobility. *American Journal of Sociology, 87*(3), 578–611.

Butler, P. (2010). *Visualizing Friendships*. Retrieved December 1, 2012, from http://www.facebook.com/notes/facebook-engineering/visualizing-friendships/469716398919.

Centers, R. (1949). Marital Selection and Occupational Strata. *American Journal of Sociology, 54*(6), 530–535.

Chan, T. W. (Ed.). (2010a). *Social Status and Cultural Consumption*. Cambridge: Cambridge University Press.
Chan, T. W. (2010b). The Social Status Scale: Its Construction and Properties. In T. W. Chan (Ed.), *Social Status and Cultural Consumption* (pp. 28–56). Cambridge: Cambridge University Press.
Chan, T. W., Birkelund, G. E., Aas, A. K., & Wiborg, O. (2011). Social Status in Norway. *European Sociological Review, 27*(4), 451–468.
Chan, T. W., & Goldthorpe, J. H. (2004). Is There a Status Order in Contemporary British Society? *European Sociological Review, 20*(5), 383–401.
Chan, T. W., & Goldthorpe, J. H. (2007). Class and Status: The Conceptual Distinction and Its Empirical Relevance. *American Sociological Review, 72*(4), 512–532.
Coxon, A. P. M., & Jones, C. L. (1978). *The Images of Occupational Prestige: A Study in Social Cognition*. London: Macmillan Press.
de Luca, D., Meraviglia, C., & Ganzeboom, H. B. G. (2010, April 14). *ICAM: The International CAMSIS Scale*. Hafia: First Presented at ISA RC28, Hafia. Software Retrieved from http://www.harryganzeboom.nl.
Ganzeboom, H. B. G. (2016). *Tools for Deriving Occupational Status Measures from ISCO-88*. Retrieved March 1, 2016, from http://www.harryganzeboom.nl/isco88/index.htm.
Goldthorpe, J. H. (2010). Analysing Social Inequality: A Critique of Two Recent Contributions from Economics and Epidemiology. *European Sociological Review, 26*(6), 731–744.
Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology, 78*(6), 1360–1380.
Greenacre, M., & Blasius, J. (1994). *Correspondence Analysis in the Social Sciences: Recent Developments and Applications*. London: Academic Press.
Griffiths, D., & Lambert, P. S. (2012). Dimensions and Boundaries: Comparative Analysis of Occupational Structures Using Social Network and Social Interaction Distance Analysis. *Sociological Research Online, 17*(2), 5.
Griffiths, D., & Lambert, P. S. (2015). Social Capital and the Social Relations of Occupational Structure. In Y. Li (Ed.), *Handbook of Research Methods and Applications in Social Capital* (pp. 204–224). London: Sage.
Grusky, D. B., & Weeden, K. A. (2006). Does the Sociological Approach to Studying Social Mobility Have a Future? In S. L. Morgan, D. B. Grusky, & G. S. Fields (Eds.), *Mobility and Inequality* (pp. 85–108). Stanford, CA: Stanford University Press.

Haythornthwaite, C. (2002). Strong, Weak, and Latent Ties and the Impact of New Media. *The Information Society, 18*(5), 385–401.

Heath, A. F. (1981). Book Review: Social Stratification and Occupations. *Sociology, 15*(4), 620–623.

Jonsson, J. O., Grusky, D. B., Di Carlo, M., Pollak, R., & Brinton, M. C. (2009). Microclass Mobility: Social Reproduction in Four Countries. *American Journal of Sociology, 114*(4), 977–1036.

Kerbo, H. R. (2003). *Social Stratification and Inequality: Class Conflict in Historical, Comparative and Global Perspective* (5th ed.). London: McGraw-Hill.

Kraaykamp, G., van Eijck, K., & Utlee, W. (2010). Status, Class and Culture in the Netherlands. In T. W. Chan (Ed.), *Social Status and Cultural Consumption*. Cambridge: Cambridge University Press.

Lambert, P. S. (2012). *Comparative Scaling of Educational Categories by Homogamy—Analysis of UK Data from the BHPS*. Technical Paper 2012-1 of the DAMES Node, Data Management Through e-Social Science. Stirling: University of Stirling. Retrieved from www.dames.org.uk/publications.html.

Lambert, P. S., & Bihagen, E. (2014). Using Occupation-Based Social Classifications. *Work, Employment and Society, 28*(3), 481–494.

Lambert, P. S., Zijdeman, R. L., Maas, I., van Leeuwen, M. H. D., & Prandy, K. (2013). The Construction of HISCAM: A Stratification Scale Based on Social Interactions for Historical Research. *Historical Methods, 46*(2), 77–89.

Laumann, E. O. (1966). *Prestige and Association in an Urban Community*. Indianapolis, IN: Bobbs-Merrill.

Laumann, E. O. (1973). *Bonds of Pluralism: The Form and Substance of Urban Social Networks*. New York: Wiley.

Laumann, E. O., & Guttman, L. (1966). The Relative Associational Contiguity of Occupations in an Urban Setting. *American Sociological Review, 31*, 169–178.

Laumann, E. O., & Senter, R. (1976). Subjective Social Distance, Occupational Stratification, and Forms of Status and Class Consciousness: A Cross-National Replication and Extension. *American Journal of Sociology, 81*(6), 1304–1338.

Levine, J. H. (1972). A Two-Parameter Model of Interaction in Father-Son Status Mobility. *Behavioral Science, 18*, 455–465.

Levine, J. H. (1990). Measuring Occupational Stratification Using Loglinear Distance Models. In R. L. Breiger (Ed.), *Social Mobility and Social Structure*. Cambridge: Cambridge University Press.

Levine, J. H., & Spadaro, J. (1988). Occupational Mobility: A Structural Model. In B. Wellman & S. D. Berkowitz (Eds.), *Social Structures: A Network Approach* (pp. 452–475). Cambridge: Cambridge University Press.

Luijkx, R. (1994). *Comparative Loglinear Analyses of Social Mobility and Heterogamy*. Tilburg: Tilburg University Press.

McDonald, K. I. (1972). MDSCAL and Distances Between Socio-Economic Groups. In K. Hope (Ed.), *The Analysis of Social Mobility: Methods and Approaches* (pp. 211–234). London: Clarendon.

McKnight, A., & Elias, P. (1997). A Database of Information on Unit Groups of the Standard Occupational Classification. In D. Rose & K. O'Reilly (Eds.), *Constructing Classes* (pp. 116–145). Colchester: University of Essex.

Mills, C. (2014). The Great British Class Fiasco: A Comment on Savage et al. *Sociology, 48*(3), 437–444.

Mills, C., & Evans, G. (2003). Employment Relations, Employment Conditions and the NS-SEC. In D. Rose & D. J. Pevalin (Eds.), *A Researchers Guide to the National Statistics Socio-economic Classification* (pp. 77–106). London: Sage.

Minnesota Population Center. (2015). *Integrated Public Use Microdata Series, International: Version 6.4 [Machine-Readable Database]*. Minneapolis, MN: University of Minnesota.

O*NET. (2008). *O*NET Online: Occupational Information Network*. Retrieved December 1, 2008, from http://online.onetcenter.org/.

ONS. (2010). *Standard Occupational Classification 2010. Volume 1: Structure and Descriptions of Unit Groups*. Basingstoke: Palgrave Macmillan (with the Office for National Statistics).

ONS. (2017). *Labour Force Survey—User Guidance*. Retrieved March 1, 2017, from https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyuserguidance.

Prandy, K. (1979). Ethnic Discrimination in Employment and Housing. *Ethnic and Racial Studies, 2*(1), 66–79.

Prandy, K. (1990). The Revised Cambridge Scale of Occupations. *Sociology, 24*(4), 629–655.

Prandy, K. (1998). Class and Continuity in Social Reproduction: An Empirical Investigation. *Sociological Review, 46*(2), 340–364.

Prandy, K. (1999). Class, Stratification and Inequalities in Health: A Comparison of the Registrar-General's Social Classes and the Cambridge Scale. *Sociology of Health & Illness, 21*(4), 466–484.

Prandy, K. (2000). Class, the Stratification Order and Party Identification. *British Journal of Political Science, 30*, 237–258.

Prandy, K. (2002). Ideal Types, Stereotypes and Classes. *British Journal of Sociology, 53*(4), 583–601.

Prandy, K., & Blackburn, R. M. (1997). Putting Men and Women into Classes: But Is That Where They Belong? A Comment on Evans. *Sociology, 31*(1), 143–152.

Prandy, K., & Bottero, W. (1998). The Use of Marriage Data to Measure the Social Order in Nineteenth-Century Britain. *Sociological Research Online, 3*(1), U43–U54.

Prandy, K., & Jones, F. L. (2001). An International Comparative Analysis of Marriage Patterns and Social Stratification. *International Journal of Sociology and Social Policy, 21*, 165–183.

Prandy, K., & Lambert, P. S. (2003). Marriage, Social Distance and the Social Space: An Alternative Derivation and Validation of the Cambridge Scale. *Sociology, 37*(3), 397–411.

Rost, J. (1988). Rating Scale Analysis with Latent Class Models. *Psychometrika, 53*, 327–348.

Rytina, S. (1992). Scaling the Intergenerational Continuity of Occupation: Is Occupational Inheritance Ascriptive After All? *American Journal of Sociology, 97*(6), 1658–1688.

Rytina, S. (2000). *A Keynote of Hierarchy, an Echo of Class: The Full Texture of Mobility in England and Wales.* Paper Presented at the RC33—Fifth International Conference on Logic and Methodology, Cologne.

Savage, M., Cunningham, N., Devine, F., Friedman, S., Laurison, D., McKenzie, L., Miles, A., Snee, H., & Wakeling, P. (2015). *Social Class in the 21st Century*. London: Pelican Books.

Savage, M., Devine, F., Cunningham, N., Taylor, M., Li, Y., Hjellbrekke, J., Le Roux, B., Friedman, S., & Miles, A. (2013). A New Model of Social Class: Findings from the BBC's Great British Class Survey Experiment. *Sociology, 47*(2), 219–250.

Stewart, A., Prandy, K., & Blackburn, R. M. (1973). Measuring the Class Structure. *Nature, 245*, 415–417.

Stewart, A., Prandy, K., & Blackburn, R. M. (1980). *Social Stratification and Occupations*. London: Macmillan.

Szreter, S. R. S. (1984). The Genesis of the Registrar-General: Social Classification of Occupations. *British Journal of Sociology, 35*(4), 522–546.

Torssander, J., & Erikson, R. (2010). Stratification and Mortality—A Comparison of Education, Class, Status, and Income. *European Sociological Review, 26*(4), 465–474.

Toubol, J., & Larsen, A. G. (2017). Mapping the Social Class Structure: From Occupational Mobility to Social Class Categories Using Network Analysis. *Sociology, 51*(6), 1257–1276.

Treiman, D. J. (1977). *Occupational Prestige in Comparative Perspective*. New York: Academic Press.

van der Gaag, M. P. J., Snijders, T. A. B., & Flap, H. (2008). Position Generator Measures and Their Relationship to Other Social Capital Measures. In N. Lin & B. H. Erikson (Eds.), *Social Capital: An International Research Program* (pp. 27–48). Oxford: Oxford University Press.

Verhaeghe, P.-P., Pattyn, E., Bracke, P., Verhaeghe, M., & van De Putte, B. (2012). The Association Between Network Social Capital and Self-Rated Health: Pouring Old Wine in New Bottles? *Health and Place, 18*(2), 358–365.

Wageindicator. (2013). *Wageindicator.org: World Database of Occupations (WISCO)*. Retrieved March 1, 2013, from http://www.wageindicator.org/main/researchlab/occupation-data-base.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

Weber, M. (1968[1922]). *Economy and Society.* Berkeley: University of California Press.

Wellman, B., & Berkowitz, S. D. (Eds.). (1988). *Social Structures: A Network Approach*. Cambridge: Cambridge University Press.

Weller, S. C., & Romney, A. K. (1990). *Metric Scaling: Correspondence Analysis*. London: Sage.

Wong, R. S. K. (2010). *Association Models*. Los Angeles: Sage.

Zijdeman, R. L., & Lambert, P. S. (2010). Measuring Social Structure in the Past: A Comparison of Historical Class Schemes and Occupational Stratification Scales on Dutch 19th and Early 20th Century Data. *Belgian Review of Contemporary History, 40*(1), 1111–1141.

# 5

# Evaluating CAMSIS Scales

## 5.1 The Empirical Character of CAMSIS Scales

At the time of writing, scales for 34 countries are published on the CAMSIS project webpages. Each scale is associated with a particular time period (that at which the social interaction data, on which the scale was calculated, was collected) and a particular occupational unit group scheme (i.e. the taxonomy of occupational codes that have been assigned scale scores). For many countries, there are scales available for more than one time period and/or for more than one occupational unit group scheme. The majority of the CAMSIS versions are for the period 1990–2010, but there are also a number of scales for dates from earlier in the twentieth century and a few that are designed to cover the nineteenth century or earlier.

In all cases, the CAMSIS scales take a linear functional form. Figure 5.1 represents the spread of numeric values of the scale for a selection of versions across a range of different societies. The scales have been divided into twenty equal intervals, and the plot shows the volume of the male population within each interval in each society. Broadly, in each society, we see clustering of individuals towards the middle range, and smaller



P. Lambert, D. Griffiths, *Social Inequalities and Occupational Stratification*,
https://doi.org/10.1057/978-1-137-02253-0_5

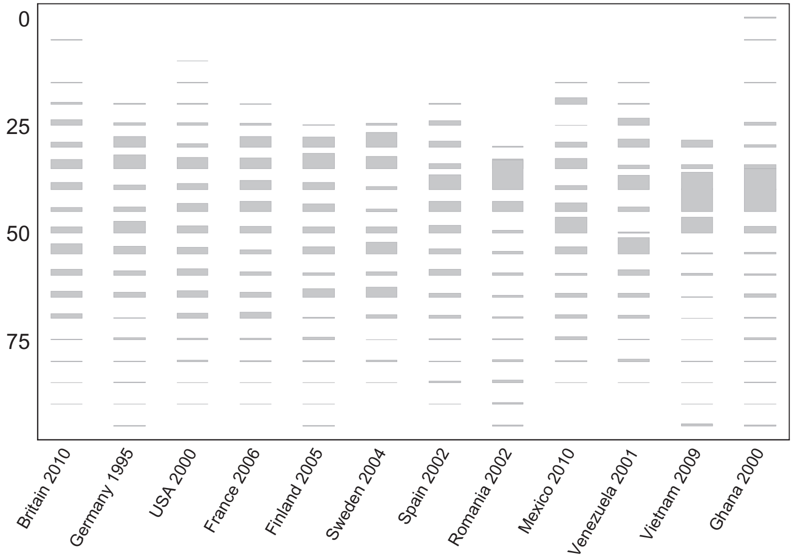



**Fig. 5.1** Banding of the CAMSIS scores for males across 12 versions. Source: CAMSIS scales from www.camsis.stir.ac.uk, linked to microdata from IPUMS-I (MPC 2015), except for Britain (Labour Force Survey; see ONS 2017), Germany (1995 Microcensus extract), Finland (Finnish Census Panel; see Statistics Finland 1996), Sweden (aggregate outputs from MONA system; see Statistics Sweden 2016), and Spain (2002 Labour Force Survey extract)

numbers of cases at the higher and lower ends, and we see that the distribution across each society is similar but not identical.

To give an indication of the sort of values which are typical of a CAMSIS scale, Table 5.1 shows a selection of occupations, alongside the CAMSIS scores which are assigned to them, for the various versions of the CAMSIS scale summarised in Fig. 5.1. Note that in all societies, CAMSIS scores are standardised to have a mean of 50 and a standard deviation of 15 in a nationally representative population of the corresponding gender, and they are 'cropped' to the range 1–99.[1] A score around 80 or 90, for instance, represents an occupation with an unusually high score (i.e. its incumbents tend to have interactions with others in occupations with high scores); a score around 10 or 20 indicates an occupation with a particularly low score (i.e. its incumbents tend to exhibit social interaction patterns with others with relatively low scores).

**Table 5.1** CAMSIS scale scores for selected occupations, across versions

| | GB | DE | US | FR | FI | SE | ES | RO | MX | VE | VN | GH |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *10* | *95* | *00* | *04* | *05* | *04* | *02* | *02* | *10* | *01* | *09* | *00* |
| | *Male scale* | | | | | | | | | | | |
| Medical doctor | 73 | 90 | 81 | 77 | 85 | 82 | 86 | 99 | 78 | 80 | 99 | 65[a] |
| University professor | 80 | 99 | 82 | 85 | 93 | 88 | 86 | 99 | 82 | 86 | 99 | 96 |
| Electrician (self-employed) | 51 | 51 | 62 | 38 | 46 | 40 | 50 | 45 | 55 | 44 | 58 | 63 |
| Nurse | 44 | 49 | 58 | 62 | 54 | 63 | 70 | 76 | 68 | 59 | 72 | 46[a] |
| Typist | 50 | 52 | 45 | 55 | 51 | 55 | 58 | 76 | 65 | 57 | 68 | 67 |
| Police officer | 61 | 48 | 53 | 62 | 55 | 46 | 50 | 50 | 54 | 49 | 49 | 50 |
| Bus driver | 30 | 33 | 40 | 39 | 35 | 34 | 46 | 45 | 52 | 44 | 53 | 44 |
| Vehicle mechanic | 38 | 40 | 48 | 38 | 48 | 34 | 55 | 45 | 54 | 43 | 49 | 56 |
| Waiter | 34 | 43 | 35 | 40 | 38 | 41 | 40 | 49 | 51 | 38 | 57 | 54 |
| Farm labourer | 22 | 27 | 25 | 35 | 37 | 32 | 25 | 35 | 26 | 19 | 46 | 32 |
| | *Female scale* | | | | | | | | | | | |
| Medical doctor | 76 | 78 | 82 | 79 | 88 | 84 | 76 | 93 | 77 | 82 | 99 | 82[a] |
| University professor | 82 | 78 | 79 | 81 | 87 | 85 | 79 | 91 | 85 | 86 | 99 | 84 |
| Electrician (self-employed) | 29 | 48 | 58 | 42 | 41 | 33 | 42 | 44 | 66 | 47 | 61 | 70 |
| Nurse | 53 | 51 | 60 | 59 | 69 | 68 | 69 | 56 | 60 | 74 | 95 | 67[a] |
| Typist | 53 | 63 | 47 | 47 | 50 | 56 | 63 | 54 | 68 | 56 | 86 | 96 |
| Police officer | 48 | 52 | 49 | 63 | 54 | 43 | 53 | 50 | 57 | 48 | 63 | 73 |
| Bus driver | 32 | 50 | 34 | 36 | 37 | 36 | 38 | 44 | 58 | 60 | 66 | 70 |
| Vehicle mechanic | 29 | 29 | 55 | 38 | 37 | 35 | 42 | 46 | 64 | 41 | 62 | 61 |
| Waiter | 29 | 35 | 34 | 34 | 35 | 37 | 44 | 45 | 50 | 33 | 57 | 62 |
| Farm labourer | 51 | 11 | 17 | 30 | 41 | 30 | 23 | 38 | 34 | 25 | 43 | 34 |

Notes: Using the CAMSIS versions for relevant countries/time periods. The score given represents the score that would be allocated to that occupation; depending upon the national occupational unit group taxonomy, the job described may or may not be a distinctive category.

[a]Category combines doctors/nurses; values shown are for other selected groups (dentists. service workers).

Table 5.1 shows illustrative examples, and we would encourage an interested reader to download and inspect the full range of scale values for these and other CAMSIS versions.

### 5.1.1 The Consistent Pattern of Gradational Inequality

Three empirical features are usually obvious when examining CAMSIS scales—and are evident from Table 5.1. These are that the order of occupations—as defined by the structure of social interactions between the incumbents of occupations—is related to social stratification, that it is substantially the same order across different versions, and that it is gradational.

At the higher end of CAMSIS scales, across different versions, are consistently found occupations such as medical doctors, university professors and business professionals, which are characterised by high levels of skill and educational requirements, high or very high incomes, and high levels of economic security. Slightly lower but still in the upper quartiles are found jobs which tend to be advantaged in some but not necessarily all of these features (e.g. school teachers, company managers, engineers). In the mid-range can be found jobs which sometimes feature favourable income and skill levels, but for which other somewhat less-privileged characteristics are readily identified (examples often include highly skilled manual jobs, public sector non-manual jobs which are filled by those with intermediate levels of education and are not usually rewarded with high pay, and small business jobs, including farm ownership, which often combine favourable economic rewards with very demanding working conditions). In the lowest quartile of CAMSIS scales can consistently be found occupations characterised by relatively low pay, lack of academic educational requirements, and menial and demanding tasks, such as factory workers, and manual and farm labourers. The general consistency across countries and time periods of the social structure depicted by a social interaction distance scale for occupations is worth emphasising. There are variations in scale scores for the same jobs from version to version (e.g. Table 5.1), but looked at on the whole, we can report that whenever a CAMSIS scale is estimated, solely using information on the social interaction patterns of the incumbents of jobs, much the same order of social inequality of occupations emerges with impressive regularity.

Figure 5.2 serves to support the interpretation of CAMSIS scales as a representation of gradational inequality (for another similar validation, see Fig. 3.1 in Chap. 3). Figure 5.2 presents the relationship between CAMSIS scores (for contemporary Britain and Sweden) and the ISEI scale of socio-economic status (based broadly on average income and educational levels per job; see Ganzeboom et al. 1992). It is clear that CAMSIS and ISEI scores are quite strongly (but not perfectly) correlated,

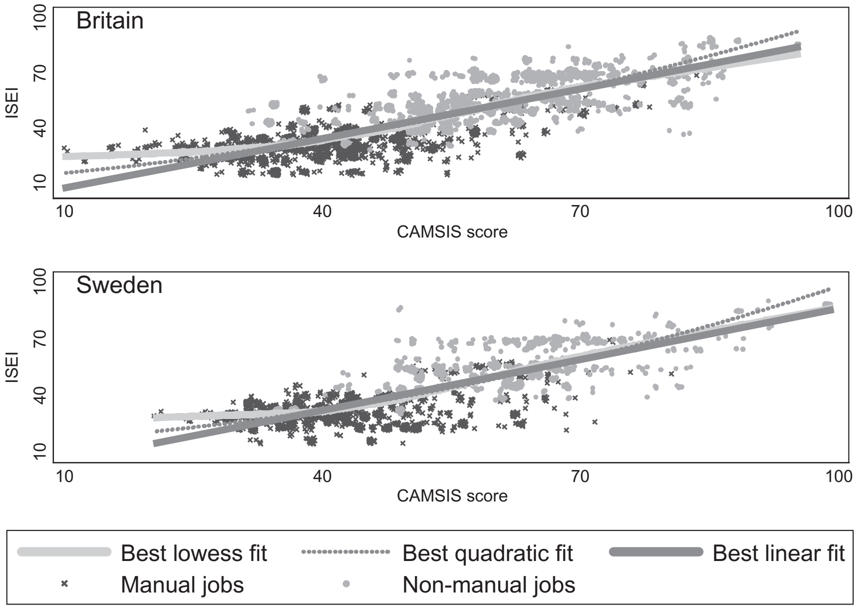



**Fig. 5.2** CAMSIS and ISEI distributions. Source: Data for males from Britain (BHPS, see University of Essex 2010) and Sweden (Level of Living Survey, LNU, see http://www.sofi.su.se/) in 1991

and the functional form of the relationship seems to be broadly linear, apparently reflecting a gradational inequality structure.

As measures that capture gradational inequality, CAMSIS scores offer one candidate measure of social stratification. We and others have published further results on the relationships between SID scales and other occupation-based socio-economic measures (e.g. Lambert and Bihagen 2014; Bessudnov 2012; Chan 2010a; Lambert et al. 2008; Mills 2007; Prandy 1990). We give some more examples in Sect. 5.2, but our claim at this point is that CAMSIS measures perform quite favourably as measurement options. Our view is that in theoretical terms the scales offer a useful way to understand social inequality (see Sect. 4.3), and, empirically, the CAMSIS scales are usually amongst the most 'powerful' occupation-based measures, since they are typically characterised by the combination of high explanatory power alongside favourable parsimony (e.g. Lambert and Bihagen 2014).

### 5.1.2 Manual and Non-manual Occupations

Figure 5.2 also shows that CAMSIS (and ISEI) scores are strongly related to a manual/non-manual division—higher scores are consistently given to jobs which don't require significant manual exertion (see also Chan and Goldthorpe 2004). Indeed, CAMSIS scales often include amongst the least-advantaged occupations many positions, usually involving manual work, which enjoy certain features often felt to be relatively attractive, such as supervisory roles or lower levels of managerial authority, task autonomy, relatively advanced vocational skills, or moderate rather than low incomes (typical examples might include skilled drivers, works foremen and supervisors, and skilled plant operatives). By contrast, there are some jobs, often non-manual in character, which may seem to be worse on such criteria, but are often given slightly higher CAMSIS scale scores (e.g. waiters, shop assistants, bar staff).

Some of these patterns are evident in Table 5.1. In most countries, jobs such as nursing and secretarial work are not given especially high wages (e.g. Charles and Grusky 2004) or social prestige (e.g. Treiman 1977), but their CAMSIS scores usually place them above the median; equally, skilled manual roles such as electrician or mechanic often enjoy moderately favourable pay or working conditions (e.g. Penn 1985), but the CAMSIS scores shown in Table 5.1 arguably place them lower than we might have guessed. Indeed, the patterns of Table 5.1 seem to suggest slight disjunctures between economic and social boundaries. We need to bear in mind that the scale is calculated by profiling the social connections held by people in the relevant occupations and is wholly independent of objective characteristics (e.g. skill or pay). For whatever reason, the pattern of CAMSIS scale scores suggests that social connections between occupations do not differentiate as strongly between the least-advantaged jobs and other manual jobs that are more intermediate—or even relatively privileged—in terms of their skill, autonomy, or pay. At the same time, CAMSIS scales do typically differentiate more substantially between manual and non-manual occupations and by educational levels. Analysing data on marriage patterns in the UK, Penn (1985) made a similar observation, that the 'economic structuration of the manual working class around the axis of skill' does not extend to social relationships

(1985, pp. 185–186). Indeed, compared to some other representations of stratification position, the pattern is that practical skills and workplace authority seem relatively less important to placement in CAMSIS scales, whereas educational backgrounds and forms of work seem more influential. In terms of theoretical interpretation, one possibility is that the long-term order of social reproduction of stratification inequalities is better captured by social than by economic indicators—perhaps because the latter often incorporate transitory circumstances that can diverge from patterns in long-term outcomes.

### 5.1.3 Distributional Characteristics

Figure 5.3 and Table 5.2 highlight some other distributional features of the CAMSIS scores that are worth noting. The light-shaded histograms summarise the distribution of CAMSIS scores given to individuals across a combined sample from all of the 12 societies being described. The

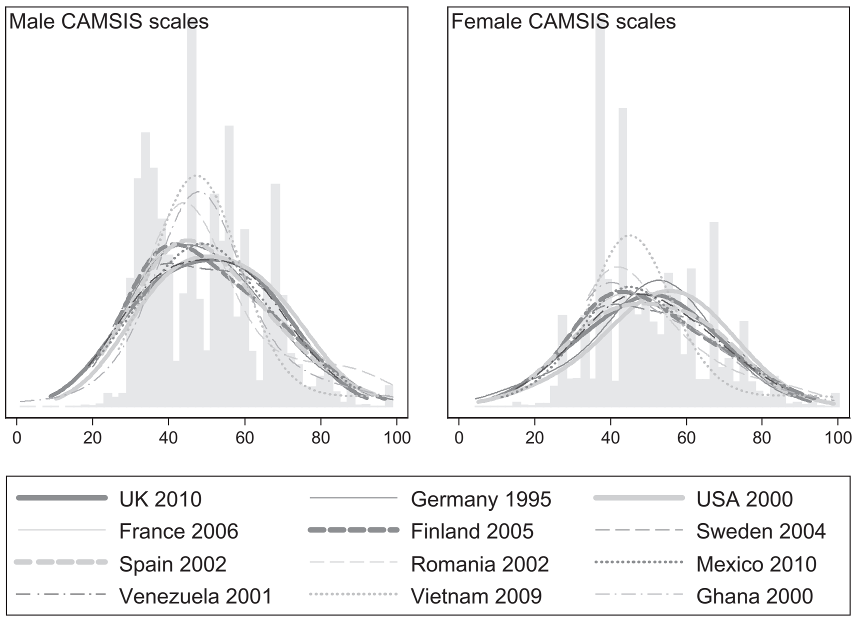



**Fig. 5.3** Distribution of CAMSIS scores across 12 countries. Source: As Fig. 5.1

**Table 5.2** Selected distributional statistics on CAMSIS scale scores across societies

| | Range | | Mean; SD (cf. 50; 15) | | | Skew; Kurtosis | | | # Categories | |
|---|---|---|---|---|---|---|---|---|---|---|
| | m | f | m | f | ma | m | f | ma | m | f |
| Britain 2010 | 9–92 | 9–93 | | | +1 | 0; 2.1 | 0; 2.5 | −0.1; 2.4 | 445 | 279 |
| Germany 1995 | 20–99 | 4–88 | | | +1; −1 | 0.7; 3.0 | −0.5; 3.2 | 0.6; 3.6 | 172 | 139 |
| USA 2000 | 10–93 | 5–99 | +2 | +3 | +4 | 0.1; 2.2 | −0.4; 3.0 | −0.1; 2.4 | 707 | 650 |
| France 2006 | 22–92 | 22–99 | | | +1; −1 | 0.5; 2.2 | 0.5; 2.5 | 0.4; 2.5 | 475 | 447 |
| Finland 2005 | 26–97 | 24–94 | | | +3; −1 | 0.6; 2.5 | 0.6; 2.5 | 0.5; 2.6 | 265 | 235 |
| Sweden 2004 | 25–88 | 22–85 | | | +3; −1 | 0.3; 2.0 | 0.3; 2.0 | −0.1; 2.2 | 111 | 107 |
| Spain 2002 | 22–95 | 19–87 | | | +1; −1 | 0.6; 2.8 | 0; 2.1 | 0.5; 2.9 | 175 | 123 |
| Romania 2002 | 31–99 | 34–97 | +3; +1 | | +3; +1 | 1.4; 3.6 | 1.3; 3.7 | 1.2; 3.5 | 370 | 356 |
| Mexico 2010 | 17–89 | 11–92 | | | +2; −1 | 0.2; 2.5 | 0.3; 2.7 | 0; 2.5 | 387 | 303 |
| Venezuela 2001 | 19–86 | 15–87 | | | +3 | 0.2; 2.2 | 0.2; 2.5 | 0; 2.2 | 88 | 71 |
| Vietnam 2009 | 27–99 | 34–99 | −1; −3 | −1; −3 | 0; −4 | 2.2; 9.5 | 2.6; 9.2 | 2.4; 10.8 | 103 | 105 |
| Ghana 2000 | 1–99 | 34–99 | | | | 0.6; 5.9 | 1.0; 3.3 | 0; 5.9 | 70 | 68 |

Source: As Fig. 5.1. Values shown are for the sub-population of adults in work who are cohabiting. ‘m’ refers to the male scale for males; ‘f’ refers to the female scale for females; ‘ma’ refers to the male scale attributed to all individuals. Values in the ‘mean/sd’ column are relative to the values 50 and 15, the conventional standardisation for the population-level mean and standard deviation (blank cells indicate zero difference).

spread itself is of limited importance since in this example the volume of cases used for each society varies dramatically. The histogram's bars do however remind us of the 'clumpy' nature of the occupational distribution: we see peaks at particular values that typically reflect the score given for one specific and highly populous occupation. Superimposed on the histogram in Fig. 5.3 are lines representing the distributional structure of the 12 separate scales. In all cases, a gradational, bell-shaped distribution is a plausible descriptor of the spread of scores through the population. Although the methods used to generate CAMSIS scores are somewhat predisposed to finding this particular distributional shape, nevertheless it does not follow automatically, and of course this is a different functional form to the uniform or discontinuous structures which are implied by adopting a 'social class' measure or other categorising tool. That the order of CAMSIS scales, as empirically derived, has a gradational functional form, has been taken as evidence of a broadly gradational order to social stratification itself (e.g. Griffiths and Lambert 2012; Bottero and Prandy 2003).[2]

Aside from the consistent bell shape, from version to version there are slight differences in the distributions of the scales, apparent in Fig. 5.3 and Table 5.2. Despite the arithmetic standardisation, there are several small variations in the overall mean, the standard deviation, and the range for the versions highlighted in Table 5.2. For information, these departures are usually related to a difference between the occupations held by the population of all adults, and of cohabiting adults. For users of CAMSIS scales, small variations in the overall means are not usually consequential, but they do demonstrate that we should ordinarily report the overall mean and standard deviation for the population or sample under analysis. Further distributional variations can be seen in Fig. 5.3 in small differences in the degree of skew (i.e. the extent to which the distributional curve is somewhat asymmetric) and kurtosis (whereby the distributional curve is somewhat more or less peaked). Greater positive skew and positive kurtosis both tend to occur in societies with larger agricultural and/or manufacturing sectors; it may be that this reflects greater polarisation between the most advantaged occupations and the (bulk of the) rest of the populations in these societies. Finally, the last columns of Table 5.2 show the number of occupational units, which varies considerably from version to version. This reflects variations in the underlying occupational taxonomies rather than the subsequent CAMSIS analysis, but the number of units can be consequential to subsequent statistical properties.

### 5.1.4 Patterns of Specificity

Some of the national differences in the distribution of CAMSIS scale scores that we see above are almost inevitable whenever an occupation-based measure is used. Particular jobs will have relatively more and less incumbents from society to society, with direct implications for the distribution of positions.[3] Typically, variations in the relative size of the agricultural sector have greatest impact upon the overall distribution of CAMSIS scores. However, there are other influential differences, such as in the proportions of manual and non-manual occupations in a society, and the relative concentrations of men and women in different occupations, which will affect the statistical properties of the occupational units.

Artefactual variations can also arise due to the way occupations are coded. An example here is that if comparable occupations are coded in non-comparable ways by two different taxonomies, it is unlikely that the derived CAMSIS scores for the same occupations will be as similar as ought, in principle, to be expected. For example, some nations distinguish the jobs of 'elementary' (or 'primary') school teacher, and 'high' (or 'secondary') school teachers on official taxonomies, but others do not. Needless to say, we might expect different CAMSIS scores for elementary teachers in a scale when their position was calculated as part of the aggregate category of all teachers, compared to one where the score applied only to their specialism.[4]

However, it is possible that differences in the distribution of CAMSIS scores from one society to another reflect genuine and meaningful variations in the very structure of social inequality as related to occupations (e.g. Lambert et al. 2008). As one example, it could be that in one nation the role of being a 'police officer' is relatively more privileged than in another, for various plausible reasons—perhaps concerning the relative size of the police service; or the way in which the police service is organised at a national or local level; or the relative distribution of other jobs within the society in comparison to police work. In this situation, the 'specificity' of CAMSIS scales makes them an appropriate and insightful way to understand social stratification. By comparison, most alternative occupation-based social classifications are derived in a manner which allocates the same occupations to the same position in the scheme regardless of national context. Indeed, an attractive feature of the SID approach is that the methodology can be used to generate occupational scales in different societies, with a view to undertaking a comparative exploration of the structure of occupations between societies.

The tradition of 'specificity' in the CAMSIS project was adopted largely on theoretical grounds, but formal criteria can be used to evaluate it. On some occasions, data might be available from a number of societies using the same occupational taxonomy, such as an ISCO classification (e.g. ILO 1990). If it can also be presumed that the same occupations are measured and coded to the scheme in a consistent way in each society, we can set up statistical comparisons between the scores given for the different units. This can be done after the scale scores have been estimated, for instance, asking if job 1234 is given a higher score in one country than another. Alternatively, a formal comparison can be made during the scale estimation process, by comparing statistical models which do and do not allow for differences in scale scores across societies. Hitherto, formal statistical comparisons have consistently shown compelling evidence of small but significant differences between the scales across societies—models which allow for scores to be different are a better fit to the data than those that do not, although the magnitude of difference is not, substantively, particularly large (e.g. Lambert et al. 2008, 2013).

However, the underlying taxonomy of occupations is often different from one society to another. For example, most of the CAMSIS scales available on the CAMSIS website, and for the collection of SID scales provided by Chan (2010a), use different taxonomies. Moreover, even if a coding frame is consistent at face value, it should not necessarily be presumed that these same occupational activities have been reliably labelled between societies—we might see cases when 'the same words are used to describe different jobs in different countries' (Treiman 1977, p. 48). Non-comparability of occupations can reflect linguistic, cultural and socio-economic differences in the uses of job titles. The occupation of 'nursing', for example, encompasses an uneven spread of positions between nations, influenced by variations in national language, culture and training traditions. In the UK, the job description and coding of 'manager' is used more liberally than in many other nations, leading to difficulties of comparability when occupational taxonomies incorporate managerial status (e.g. Elias and Birch 1994). In these cases, when the underlying taxonomy is not consistent, statistical model comparisons cannot reasonably adjudicate on the specificity of the stratification structure, but it remains possible to evaluate post hoc properties of the different measures. Lambert et al. (2008) approached this, for example, by comparing the empirical properties of specific CAMSIS scales against those of other occupation-based measures that are applied universally across societies, and argued that the specific versions provided small but significant improvements.

'Specificity' in CAMSIS scale versions brings at least two challenging features. First, it makes the scales more complicated to work with, both for practical implementations and when communicating with others. Second, it introduces a small risk that some variations between CAMSIS scores on different versions might reflect measurement and sampling errors. Standard errors statistics can be generated to ameliorate the second challenge (see Chap. 6). More generally a 'universal' (e.g. cross-national) CAMSIS measure could avert both of these problems, because as well as being 'easier' to use, the process of averaging patterns across societies might serve to 'iron out' measurement error (or provide valid 'imputations' from other contexts—for instance, if an occupation is not represented by many people in a particular society, a more accurate representation of its social position might reasonably come from information about its position in other societies). Recently a universal 'I-CAM' scale ('International CAMSIS') has been generated by de Luca et al. (2010), through analysis of data on occupations from a cross-nationally pooled survey sample. This offers a single (non-specific) CAMSIS scale with the same scores for the same occupations across societies. There are good grounds to expect that the occupational order is largely stable from society to society (especially Treiman 1977), and the I-CAM scale itself shows a high correlation (usually at least 0.9) with specific scales when implemented on the same datasets. The I-CAM scale makes a very useful and convenient tool for depicting social inequalities, but nevertheless it does not avert the possibility that interesting and meaningful variations exist in the relative circumstances of the incumbents of the same occupations from society to society.

### 5.1.5 Specific Scales for Men and Women

Another form of 'specificity' can be seen in the two separate scores for male and female jobs that are produced as standard in the CAMSIS tradition (see, for instance, Tables 5.1 and 5.2). Generating separate scales for men and women—see also Sect. 6.8 and Chap. 11—is neither an automatic nor necessary feature of the CAMSIS approach, but it has been a common convention. The logic is that strong patterns of gender segregation in occupations imply that the average social circumstances of the male incumbents of jobs may not be identical to the average circumstances of the female incumbents of the same jobs. One well-known example concerns the occupation of 'elementary school teacher'. In many societies, this job is disproportionately held by women, but the smaller numbers of males who work in it often enjoy higher probabilities of holding more senior positions, such as being a head teacher (e.g. Bradley 1989, c13). Accordingly, we might anticipate that the average social position of male elementary school teachers, relative to the distribution of other males, may be somewhat higher than the corresponding position of female primary school teachers relative to the distribution of other females. We can cater to such situations by constructing separate CAMSIS scales for male and female occupations (e.g. Prandy 1986).

Despite the theoretical case for using male and female CAMSIS scales, doing so does add complexity in the adoption of CAMSIS scales. A particular issue arises when analysing a combined population of men and women—should we use the same scale for everybody, or should the scale be contingent upon gender? Perhaps the more common approach has been to use the male scale scores for everyone, regardless of gender. This makes the interpretation reasonably straightforward: the scores reflect the level of social advantage that is on average associated with male incumbents of the occupation. It may seem counterintuitive to characterise women according to where their occupation sits within the male stratification structure, but it is nevertheless the option that we tend to favour in our own analyses. This forces all occupations into a single relative structural system, but there is a risk that the female incumbents of some occupations, for instance, those with many female and few male incumbents, are not well represented in terms of their male scale profile.

A common alternative is to assign male scale scores to males and female scale scores to females. This can be justified on theoretical grounds, since for both men and for women, the score will tell us about the relative position within the distribution of men and women, respectively. However, this interpretation is more nuanced than it may seem—for instance, it implies an overall equality between the male and female distribution of jobs, which is not likely to be the case on the basis of other indicator measures. Accordingly, as regular users of CAMSIS scales, we ourselves don't usually deploy this approach. Operationally, many other alternative strategies could also be considered when analysing a mixed-gender population though they are less commonly deployed in practice. We could, for example, calculate averages or weighted averages of the male and female scores for the same occupations (in theoretical terms, this gives much the same result as if we constructed a single scale that forced male and female scores to be equal). We could also apply some other scaling parameters, for instance, based on male and female average incomes, to the scales applied to men and women, respectively (see Chap. 11). Scientifically, we feel that any approach can be justified, but it is important to communicate clearly what has been done and what the implications might be for empirical patterns related to the scale.

### 5.1.6 Occupational SID Scales Based Upon Different Social Relationships

Occupational SID scales have been calculated on pairs of occupations defined by several different forms of social relationship, and, hitherto, comparisons between them suggest that the same dimensional structure of social stratification emerges regardless of the social relationship used (Prandy and Lambert 2003; Chan 2010b).

The most commonly used forms of social relationships are marriage/cohabitation patterns (e.g. Prandy and Lambert 2003), friendship (e.g. Stewart et al. 1980), and intergenerational relationships, typically data on the occupations of fathers and sons (e.g. Rytina 1992; Levine and Spadaro 1988). Additionally, intragenerational (job history) occupational data has been analysed to reveal the same structure (Lambert and Prandy 2002; Levine and Spadaro 1988), as has data on homosexual partnerships (Alderson et al. 2007; Chan 2010b), data on acquaintances defined as 'people that you know' (Savage et al. 2015, c4), and data on more distant family relationships, such as all pairs of adults living together in the same household and relationships involving parents-in-law rather than parents (Lambert et al. 2013). Prandy and Lambert (2003) and Chan (2010b) using UK census data, and Alderson et al. (2007) using US survey data, conducted analyses which led to them to argue that there was no substantial difference in the resulting dimension according to the type of social interaction data that is used.

Figures 5.4 and 5.5 provide some corroborating examples. Using the UK's BHPS, we can readily access data on pairs of jobs linked by at least eight forms of social connection: marriage and cohabitation, marriage only, cohabitation only, co-resident males (who are not partners), friendship, father-child, father-in-law to child, and career transitions. Figure 5.4 simply shows the distribution of combinations of people in those pairs in the one-digit categories of the standard occupational classification: the patterning is stronger and weaker for the different social connections, but the underlying relationship of homophily is always visible. In fact, so long as the relationship is there, the strength of the association is of little consequence to the underlying order that will be identified by a SID analysis. Figure 5.5 then summarises the results of applying SID analyses to the eight different sets of pairs of occupations (in this analysis, we constrained ego and alter scores for the same units to be the same). In our view, basically the same structure is revealed for every form of social relationship. There are some slight variations, such as a greater advantage estimated for 'professional' occupations in the father-in-law to son association, and greater variation in positions for some major groups than others, but these could amount to measurement error (e.g. at the one-digit level; we might not be consistently

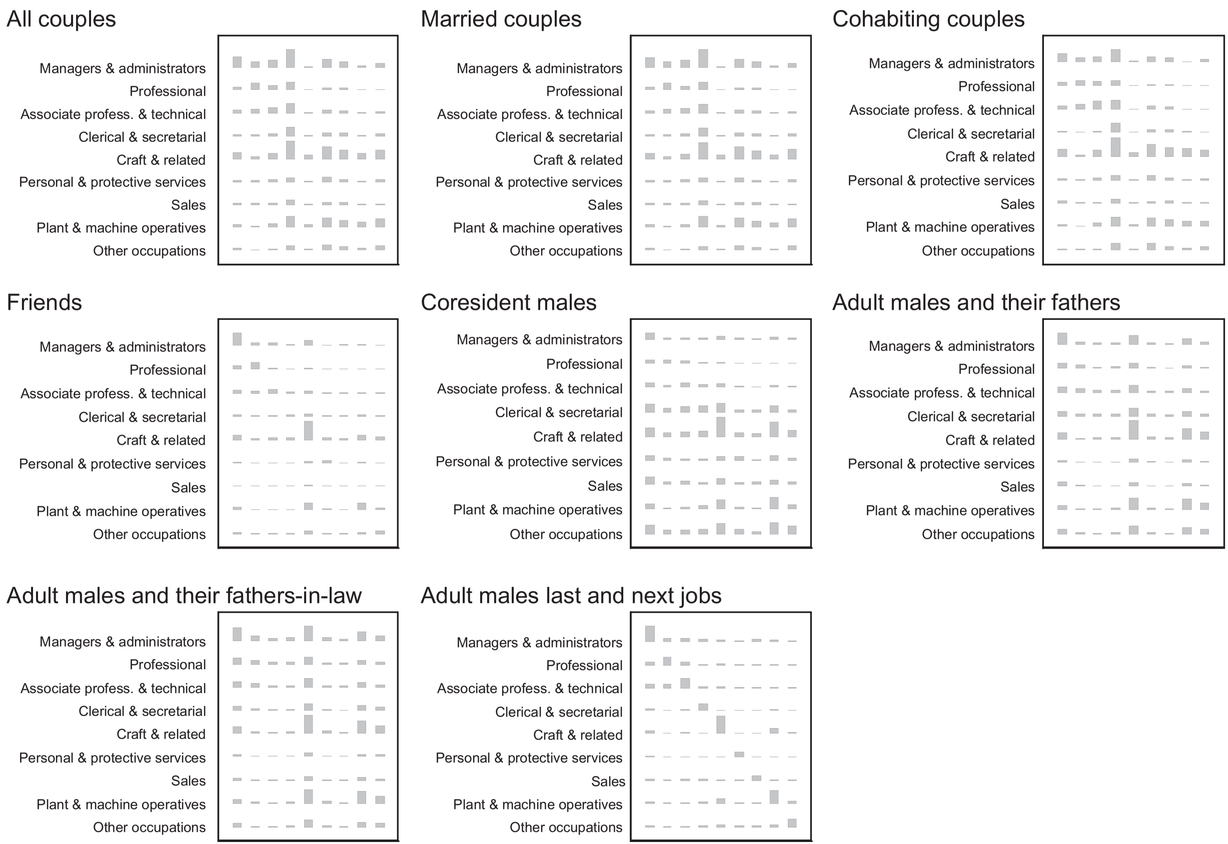



**Fig. 5.4** Comparing different types of relationship in their CAMSIS scores. Source: Analysis of data from combined waves of BHPS, 1991–2008 (University of Essex 2010)

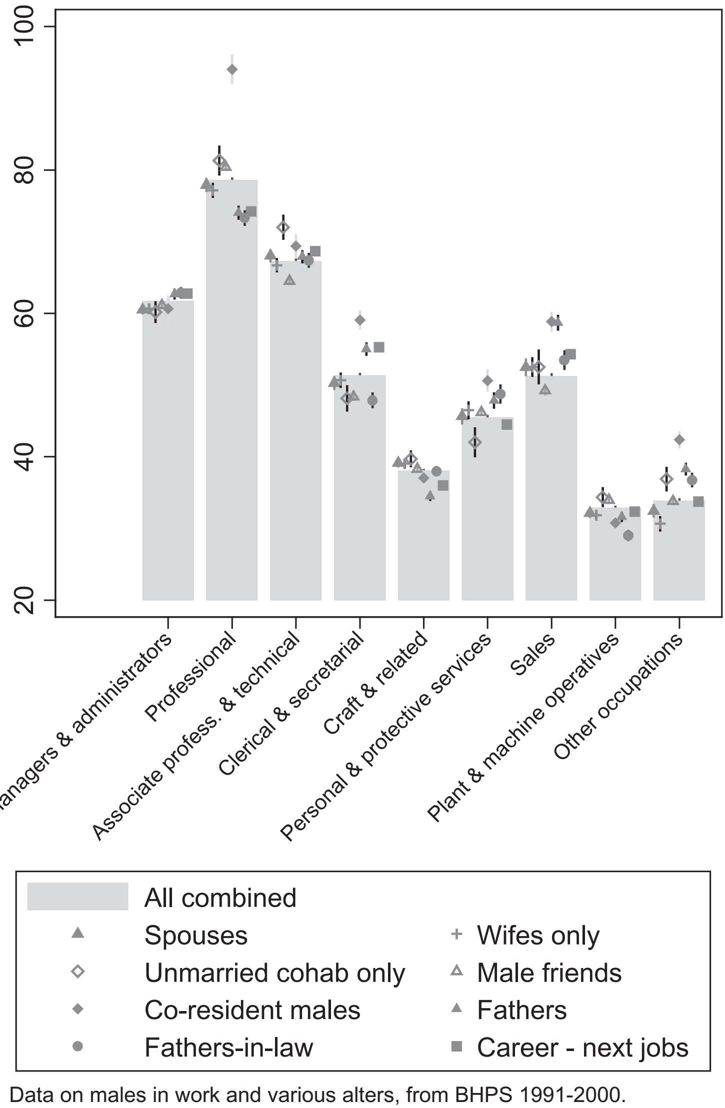



**Fig. 5.5** CAMSIS scores according to different types of data

disentangling 'subsidiary dimensional' structures that might arise in the data—see Sect. 6.5.2). However, the pattern that should really be emphasised is the consistency between these structures despite their use of different forms of social connections.

### 5.1.7 Historical CAMSIS Scales

For earlier time periods, the exploratory nature of the CAMSIS approach is particularly attractive, because in this situation we may not have easy access to large-scale data on the income or educational levels of individuals, but, through by-product data such as historical marriage registers and censuses, we do know about the jobs held by individuals and by some of their social connections such as their family. Accordingly, there have been several applications of the CAMSIS tradition to historical records of occupational data (e.g. Lambert et al. 2013; Prandy and Bottero 1998, 2000). Hitherto, attention has concentrated on the nineteenth century, when detailed occupational data is most readily available on a large scale, although there is no reason, in principle, why analysis could not cover earlier periods still.

Prandy and Bottero (1998, 2000, Bottero 2000) undertook an extended analysis for data from the UK using parent-child records of occupations as collected from a compilation of genealogical records which were collated by Prandy and Bottero into the 'Family History Study'. A rich database was formed covering around 50,000 pairs of socially connected occupations, based upon data of up to eight sequential generations. Records were derived from registers obtained at events of either birth, marriage, or death, including detailed occupational codes (and also geographical identifiers) of the adults at the relevant occasion. The occupational titles were codifed into a scheme of around 100 different units that was derived by the authors. Correspondence analysis was used to identify dimensions of inequality within the data. Two important conclusions emerged from this work. Firstly, the overall character of the occupational scale was quite similar to that of contemporary CAMSIS scales, suggesting persistence through time of the overall structure of social stratification and inequality. Secondly, Prandy and Bottero identified what seemed a turning point in occupational relations of social interaction, occurring in Britain in around 1870, when there was a (small) step change in intergenerational social mobility rates (rates increased through time, but the rate of growth was faster before 1870). To accommodate this adjustment, two separate CAMSIS scales were produced for an 'early' (1777–1866) and 'late' (1867–1913) period. This point of change is long after the early stages of the 'industrial revolution' in the UK. Accordingly, the differences between the scales were not attributed to the earlier transition from agricultural to industrial production, but linked rather to the burgeoning expansion of the division of labour generally, and of increasing bureaucratisation and when relevant professionalization of work organisations, aided by expanding transport and communication networks and growing educational provision. Modernisation, in other words, seemingly impacted upon the structure of occupational stratification inequality.

Expanding upon this analysis using more countries and a larger volume of microdata, the HIS-CAM project estimated CAMSIS scales for numerous societies and time periods between 1800 and 1938 (e.g. Lambert et al. 2013; Zijdeman and Lambert 2010). An eclectic range of data sources were used in this analysis, including census data (NAPP 2008), marriage register data (most typically featuring occupational records for the groom's job, the groom's father's job, and the groom's father-in-law's job), and other administrative data such as parish registers. Lambert et al. (2013) reported that the broad order of stratification inequality as revealed through social interaction distance analysis was similar throughout the societies and time periods concerned, but there was also evidence of variation from nation to nation and from earlier and later time periods in the relative social position of occupations. The latter variation was statistically significant, but of relatively muted consequence; in many scenarios, Lambert et al. (2013) argued that a single 'universal' HIS-CAM scale would be suitable for exploitation across countries and over time periods.

## 5.2 Validity Testing for CAMSIS Measures

### 5.2.1 Evidence of Validity

'Construct validity' usually refers to the assessment of correlations between measures of interest and qualitatively different things which might be expected to be related to the measure (or, indeed, unrelated). 'Criterion validity' usually refers to evaluating empirical associations between measures and things to which they are explicitly designed to connect with. A number of construct validity studies for the Cambridge scale and CAMSIS measures as indicators of social stratification position have been published (e.g. Prandy 1998, 1999, 2000). In these and other reviews, evidence suggests that CAMSIS measures perform as well as or better than alternative measures when predicting a range of outcomes which should be related to social stratification position at the individual level. The performance of CAMSIS scales when correlated, for example, with income, wealth, educational attainment, social mobility, unemployment risks, and poverty risks is said to be on a par with most other stratification measures (Lambert and Bihagen 2014), and CAMSIS measures are argued to be stronger predictors of outcomes covering political values and preferences, lifestyles and leisure, social reproduction, and health, than are most alternative occupation-based measures (Prandy 1998, 1999, 2000). Chan and Goldthorpe (2004, 2007) similarly explored the correlation between their SID scale and relevant outcomes, highlighting that their scale has a stronger correlation to outcomes related to lifestyle and consumption, but that social class based on employment relations was a stronger predictor of economically oriented outcomes such as income and unemployment risk.

Table 5.3 summarises some selective results which indicate the correlation or explanatory power of various occupation-based measures and selected outcomes, drawing upon a wider range of results which have been reported by Lambert and Bihagen (2014; Bihagen and Lambert 2012). Largely for convenience, we use data for Britain for 1991, but we suspect that the same patterns apply in other societies. The table gives information on scenarios where a regression model is run in which the outcome is potentially predicted by the relevant occupation-based

**Table 5.3** Validity review data for selected occupation-based measures

| | Gender | Age | Log of current income | Highest educational attainment | Smoking | Reads national broadsheet newspaper | Any future unemployment over next 10 years (with linked data) | Promotion opportunities in job |
|---|---|---|---|---|---|---|---|---|
| | L | R | R | O | O | L | L | L |
| CAMSIS | | | | | | | | |
| A | 8.1 | 5.5 | 31.1 | 32.9 | 10.5 | 44.2 | 11.3 | 13.4 |
| B | 19.1 | 18.6 | 43.3 | 33.6 | 11.1 | 45.4 | 12.7 | 40.5 |
| C | | | 9.0 | 25.4 | 6.7 | 36.7 | 6.0 | 4.4 |
| CG status | | | | | | | | |
| A | 18.3 | 3.5 | 20.1 | 30.6 | 11.4 | 40.8 | 12.2 | 16.4 |
| B | 25.2 | 18.6 | 41.3 | 32.0 | 11.7 | 43.7 | 14.0 | 39.8 |
| C | | | 9.1 | 24.5 | 7.5 | 35.2 | 6.0 | 7.2 |
| ISEI | | | | | | | | |
| A | 0.0 | 1.8 | 39.3 | 30.4 | 9.7 | 40.7 | 7.3 | 19.7 |
| B | 13.2 | 18.7 | 47.5 | 31.1 | 10.0 | 42.7 | 9.8 | 41.1 |
| C | | | 9.8 | 22.6 | 6.4 | 33.0 | 3.5 | 7.7 |
| SIOPS | | | | | | | | |
| A | 3.5 | 6.3 | 45.9 | 31.7 | 9.6 | 39.0 | 7.7 | 18.1 |
| B | 12.8 | 18.6 | 51.2 | 32.1 | 10.0 | 40.9 | 10.0 | 40.9 |
| C | | | 11.4 | 24.4 | 6.7 | 31.5 | 4.1 | 6.8 |
| NS-SEC (7 category version) | | | | | | | | |
| A | 22.5 | 16.2 | 48.8 | 33.4 | 10.3 | 41.0 | 13.1 | 37.2 |
| B | 24.4 | 19.6 | 51.6 | 33.8 | 11.0 | 42.5 | 14.6 | 42.7 |
| C | | | 13.6 | 25.1 | 5.6 | 33.3 | 8.0 | 22.0 |
| RGSC (6 category version, excluding the 'Armed forces' category) | | | | | | | | |
| A | 26.3 | 12.3 | 47.0 | 32.8 | 10.9 | 42.2 | 11.4 | 18.8 |
| B | 28.4 | 20.5 | 51.1 | 33.1 | 11.1 | 43.2 | 13.3 | 40.7 |
| C | | | 10.7 | 24.5 | 5.7 | 34.4 | 5.7 | 7.7 |

Source: Data from 6151 adults in work in the BHPS in 1991. Outcomes predicted by linear multiple regression (R), logistic regression (L) or ordered logistic regression (O). Row A shows bivariate correlation R (square root of model $r^2$ or pseudo-$r^2$) (times 100). Row B shows the same correlation value after adding employment status dummy variables. Row C shows increment in R after adding the measure to a model with controls for age, gender and gender segregation index measure, and summary statistics are used to indicate the extent to which the occupation-based measure aids prediction (the most parsimonious model is also highlighted in the table, viz. the model with the lowest BIC value).

We draw two conclusions from Table 5.3 in combination with other studies in the field. First, the correlations confirm the construct and criterion validity of CAMSIS scales as measures of social stratification. Secondly, we believe that Table 5.3 suggests that when they are considered as generic occupation-based measures of social stratification, CAMSIS scales perform favourably by comparison to most other available measures. Their associations with various outcomes are of approximately the same magnitude, and sometimes slightly higher, than are the comparator associations involving other occupation-based measures. In addition, their correlations are generally lower with things that a measure of social stratification ought not to be expected to be strongly related to, such as gender and age. By contrast, many occupation-based social class schemes have much higher correlations to gender than do CAMSIS scales,[5] and those alternative occupation-based measures that are more closely defined according to economic situations, such as ISEI, are more strongly related to age. To reiterate a point made in Chap. 4: if we believe that social stratification is a long-term structure of relatively stable social inequalities, we should not expect major differences in average social stratification position by age (since the same people often maintain similar circumstances throughout their lives), nor by gender (since most men and women live as heterosexual couples with shared social circumstances). As such, the correlations of other measures with age and gender are relatively problematic.

## 5.2.2 Disputed Validity

CAMSIS scales have in the past been criticised for lacking validation studies in support of their use as tools for understanding the stratification structure. This criticism is particularly linked to what became known as the ‘boat race’ within sociological research in the UK, where a long-standing difference of opinion over the measurement of social stratification was played out in debates between the Cambridge group, advocating the CAMSIS approach, and academics from a research programme at the University of Oxford that centred upon the analysis of social class measured through the Goldthorpe scheme (e.g. Goldthorpe 1997). In the latter paradigm, it was commonly asserted that whilst social class measure was extensively validated (e.g. Evans and Mills 1998, 2000), measures based on social interaction distance analysis lacked validation. The crux of this position was that those publications that did report correlations between SID measures and other social outcomes (e.g. Prandy 1990, 1998, 1999, 2000; Stewart et al. 1980) did not serve as demonstrations of either construct or criterion validity. This was said to be the case because it was not sufficiently clear what SID measures were supposed to represent, and therefore evidence of validation was, by definition, impossible.

The alleged lack of clarity over what is being measured through CAMSIS scales has been the source of considerable frustration through the years to advocates of the CAMSIS approach, since they would argue that its interpretation has been extensively addressed. As in the position taken in this book, CAMSIS scales have commonly been presented as indicators of the average position held by the incumbents of occupational positions within the enduring structure of social stratification. That is, CAMSIS scales measure the 'trace of social reproduction', a shorthand term for which is 'social stratification'.

Of course, CAMSIS and SID measures have also been labelled by others in slightly different ways—for example, as measures of 'generalised advantage', 'socio-economic advantage', 'prestige', 'social capital', 'status', and 'social class'. Some have portrayed them in terms of multiple concepts at once, for instance, Savage et al. (2015, p. 142) describe a SID scale with '… from top to bottom, we range down from the higher-status positions, with more power, generally requiring more education, commanding more respect, and/or earning higher incomes … [it] looks, in fact, like a clear map of social classes'. Such alternative representations have their justifications, and it is easy in this context to see why the measures may seem 'unclear'. However, it does not follow that because more than one interpretation has been suggested, a property is inherently unclear. Indeed, the appropriate interpretation of many other occupation-based social classifications is also contested.

In later publications also associated with the Oxford group, a different point was emphasised that has led to an influential alternative interpretation of SID scales for occupations: as already highlighted, after extensive analysis generating and evaluating occupational SID measures, Chan and colleagues (e.g. Chan 2010a; Chan and Goldthorpe 2004, 2007) came to argue that SID measures had construct and criterion validity as measures of 'social status' (in the classical sense of social honour or recognition). Nevertheless, in Sect. 4.3 we discussed the theoretical interpretation of CAMSIS measures and argued specifically against the view that SID scales measure status—a fuller argument is rehearsed by Bihagen and Lambert (2012), who explore the minutiae of differences between available measures of SID scales and social class measures and argue that SID scales are not consistently distinctive as measures of 'status', but can be seen more consistently as empirical representations of 'stratification'.

Empirically, we argue that the construct validity of CAMSIS as 'stratification' measures is readily demonstrated (and long has been), through evidence of associations with outcomes linked to stratification. Criterion validity as a measure of stratification can be demonstrated in terms of associations with social mobility and social reproduction patterns, where a variety of studies have demonstrated the value of CAMSIS measures as indicators of stratification positions (e.g. Rytina 1992, 2000; Prandy 1998). However, a generic problem with validity testing in this field is that most occupation-based measures have high correlations with each other, so it follows that most occupation-based measures have construct and criterion validity in terms of most relevant concepts of social inequality. CAMSIS measures, for example, also have validity, were it to be asserted, as measures of socio-economic position, prestige, educational credentials, and employment relations and conditions, insofar as they have comparable correlations with other indicators of these concepts (especially Lambert and Bihagen 2014).

### 5.2.3 Operational Issues

Numerous studies have demonstrated that when compared to categorical class schemes, CAMSIS scales are usually more parsimonious (in both substantive and statistical terms), since they are usually built into models with a single parameter for their linear association with the outcome (e.g. Lambert and Bihagen 2014). Indeed, they usually require only one parameter for their 'main effects' and one additional parameter for any further 'interaction effects'. This compares favourably to the burgeoning set of parameters that are required when using dummy variable categories for categorical class measures and their interactions. Summary statistics that capture an indication of parsimony commonly favour CAMSIS scales in such scenarios for these reasons. Such conveniences are particularly important when considering the daily practice of social research, because categorical class schemes are routinely simplified into a more aggregate level than they were originally designed for, typically for reasons of convenience, but in a manner that is likely to attenuate empirical patterns.

It is sometimes suggested that categorical occupation-based class measures are preferable to gradational scales because they are more readily adaptable to non-linear social relations. Non-linear effects are not routinely considered when using CAMSIS measures in statistical analysis, yet it is plausible to fit a curvilinear function of CAMSIS, or an effect with some other functional form. This is uncommon in practice, perhaps since it introduces more complex elements of data analysis, interpretation and communication. On a related point, it has also been noted that substantially more of the empirical structure that is associated with occupations can be captured by building in additional model parameters for employment status categories, alongside the CAMSIS scale (e.g. Prandy 1990), although this again adds a complication that some prefer to avoid. Such adaptations when using CAMSIS scales, though not well known, are plausible, and have the potential for widening the empirical value of CAMSIS scales as analytical tools.

There is an important linked debate, concerning the use of occupation-based measures, over which occupational data should be used to explore social patterns. Table 5.3, as is a common default position, evaluates the occupation-based measures allocated to individuals on the basis of the occupation that they currently hold and excludes non-working individuals from the analysis. In other situations, it may well be persuasive to exploit longitudinal data about an individual's job history in order to allocate their measure on the basis, potentially, of the last job that they held, or to use some other related criteria, such as identifying the last full-time job. Moreover, it is well established that in many situations it may be useful to allocate occupation-based measures according to the job of some other related individual, such as a spouse or parent, rather than giving primacy to the individual's current job (e.g. Davies and Elias 2010). The empirical performance of an occupation-based measure could clearly vary according to these permutations. In one sensitivity analysis, Lambert et al. (2008) reported little variation in the differences between measures according to such criteria. However, in our opinion, a further attraction of CAMSIS and SID scales as occupation-based measures is that they are naturally much better adapted to different permutations in the use of occupational data and different levels of analysis. For example, it is easy to summarise the scores from multiple relevant occupations, whether taking means (e.g. of occupations in a household), other moments (e.g. the highest occupations in a career), or other summaries of a distribution or trajectory (e.g. Prandy and Bottero 2000 summarised career trajectories through the regression gradient for occupational score by time). Similar summaries might be constructed with categorical measures, but the task is harder. Second, as scale scores based on social interaction patterns, CAMSIS and SID scales are also at an advantage since they draw upon wider social relations that cross gender and life course boundaries. Whereas socio-economic profiles based upon jobs might be tied to the social trajectories of occupational experiences (for instance, the most senior positions might be disproportionately held by older, full-time, males), the same segregations do not apply to social interaction patterns. Accordingly, the stratification structure captured by SID scales is, arguably, better connected to the social structure of inequality and better isolated from the vagaries of family and life course relations to employment participation.

In summary, whilst CAMSIS measures are correlated to other occupation-based schemes, we believe that the small empirical and operational differences between alternative instruments often point in favour of using CAMSIS measures as indicators of position in the structure of social stratification. Of course, we have presented only a brief range of empirical data such as in Table 5.3, but we have also highlighted a considerable corroborating literature. There remain some difficulties with

CAMSIS measures as inequality indicators—like all occupation-based measures, CAMSIS scales ignore within-occupational heterogeneities, which may be substantial if the occupational units on which the measure is based are relatively aggregated; in the calculation of scores, discussed in the next chapter, there are also elements of sampling error and procedural uncertainties that are not simple to resolve. Nevertheless, we would encourage readers to consider using CAMSIS measures more often, on the grounds that they tend to have relatively favourable empirical and operational properties; in our view, there are also compelling theoretical grounds for using CAMSIS scales (see Chap. 4).

## Notes

1. 'Cropping' means here that if after standardisation to mean 50 and standard deviation 15 the estimated value is below 1 or above 99, it is recorded as the lower or upper limit (i.e. 1 or 99). Arithmetic standardisation is usually implemented on the basis of the sample of nationally representative men and women that was also used for the derivation of the CAMSIS version, but this means that when the scores are linked through to a different dataset, the mean and standard deviation in the new dataset are unlikely to be exactly 50 and 15.
2. Alternative techniques that explicitly look for clustering patterns within the interaction structure might lead to a different result. Toubol and Larsen (2017) use a clustering algorithm to analyse interactions between occupations as defined by career mobility, and identify a set of categorical social classes in contemporary Denmark.
3. A point to emphasise is that the values for one occupation on a scale like CAMSIS are contingent upon the overall distribution of scores through the population. In a nation with a larger volume of jobs of a non-manual character, for example, the average scale score for non-manual jobs should be, by construction, closer to the average than in one where that sector is much smaller.
4. The boundaries between, and circumstances of, these two occupations may also vary. For instance, in some societies the label 'teacher' is only used for roles that require a university-level teaching qualification, but in others the same label describes a much more diverse set of roles in education.

5. The association between the CAMSIS scale and the 'status' scale of Chan and Goldthorpe (2004) is relevant here. The status scale is derived using almost identical methods, but with a broader level of occupational aggregation. The status scale has a higher correlation both to individual gender and to an occupation-based measure of gender segregation (see Bihagen and Lambert 2012). These patterns suggest that the separate mechanism of gender segregation is better disaggregated from a SID scale if more detailed occupational units are used.

## Bibliography

Alderson, A. S., Junisbai, A., & Heacock, I. (2007). Social Status and Cultural Consumption in the United States. *Poetics, 35*(2/3), 191–212.

Bessudnov, A. (2012). A Relational Occupational Scale for Russia. In P. S. Lambert, R. Connelly, R. M. Blackburn, & V. Gayle (Eds.), *Social Stratification: Trends and Processes* (pp. 53–65). Aldershot: Ashgate.

Bihagen, E., & Lambert, P. S. (2012, September, 24–26). *Class and Status: One or Two Forms of Stratification?* Paper Presented at the Equalsoc Conference, Stockholm.

Bottero, W. (2000). Gender and the Labour Market at the Turn of the Century: Complexity, Ambiguity and Change. *Work Employment and Society, 14*(4), 781–791.

Bottero, W., & Prandy, K. (2003). Social Interaction Distance and Stratification. *British Journal of Sociology, 54*(2), 177–197.

Bradley, H. (1989). *Men's Work, Women's Work: A Sociological History of the Sexual Division of Labour in Employment*. Cambridge: Polity Press.

Chan, T. W. (Ed.). (2010a). *Social Status and Cultural Consumption*. Cambridge: Cambridge University Press.

Chan, T. W. (2010b). The Social Status Scale: Its Construction and Properties. In T. W. Chan (Ed.), *Social Status and Cultural Consumption* (pp. 28–56). Cambridge: Cambridge University Press.

Chan, T. W., & Goldthorpe, J. H. (2004). Is There a Status Order in Contemporary British Society? *European Sociological Review, 20*(5), 383–401.

Chan, T. W., & Goldthorpe, J. H. (2007). Class and Status: The Conceptual Distinction and Its Empirical Relevance. *American Sociological Review, 72*(4), 512–532.

Charles, M., & Grusky, D. B. (2004). *Occupational Ghettos: The Worldwide Segregation of Women and Men*. Stanford, CA: Stanford University Press.

Davies, R., & Elias, P. (2010). The Application of ESeC to Three Sources of Comparative European Data. In D. Rose & E. Harrison (Eds.), *Social Class in Europe: An Introduction to the European Socio-Economic Classification* (pp. 61–86). London: Routledge.

de Luca, D., Meraviglia, C., & Ganzeboom, H. B. G. (2010, April 14). *ICAM: The International CAMSIS Scale*. Hafia: First Presented at ISA RC28, Hafia. Software Retrieved from http://www.harryganzeboom.nl.

Elias, P., & Birch, M. (1994). *Establishment of Community-Wide Occupational Statistics: ISCO-88(COM), A Guide for Users*. Warwick: Institute for Employment Research, University of Warwick.

Evans, G., & Mills, C. (1998). Identifying Class Structure: A Latent Class Analysis of the Criterion-Related and Construct Validity of the Goldthorpe Class Schema. *European Sociological Review, 14*(1), 87–106.

Evans, G., & Mills, C. (2000). In Search of the Wage-Labour/Service Contract: New Evidence on the Validity of the Goldthorpe Class Schema. *British Journal of Sociology, 51*(4), 641–661.

Ganzeboom, H. B. G., de Graaf, P. M., & Treiman, D. J. (1992). A Standard International Socio-Economic Index of Occupational Status. *Social Science Research, 21*, 1–56.

Goldthorpe, J. H. (1997). The "Goldthorpe" Class Schema: Some Observations on Conceptual and Operational Issues in Relation to the ESRC Review of Government Social Classifications. In D. Rose & K. O'Reilly (Eds.), *Constructing Classes* (pp. 40–48). Colchester: University of Essex.

Griffiths, D., & Lambert, P. S. (2012). Dimensions and Boundaries: Comparative Analysis of Occupational Structures Using Social Network and Social Interaction Distance Analysis. *Sociological Research Online, 17*(2), 5.

ILO. (1990). *ISCO-88: International Standard Classification of Occupations*. New York: International Labour Office.

Lambert, P. S., & Bihagen, E. (2014). Using Occupation-Based Social Classifications. *Work, Employment and Society, 28*(3), 481–494.

Lambert, P. S., & Prandy, K. (2002). *Intra-Generational Mobility and Social Distance: Work History Analysis and Occupational Structure*. Paper Presented at the ESRC Seminar Series: Occupational Information in Social and Economic Research: Occupational Careers.

Lambert, P. S., Tan, K. L. L., Gayle, V., Prandy, K., & Bergman, M. M. (2008). The Importance of Specificity in Occupation-Based Social Classifications. *International Journal of Sociology and Social Policy, 28*(5/6), 179–192.

Lambert, P. S., Zijdeman, R. L., Maas, I., van Leeuwen, M. H. D., & Prandy, K. (2013). The Construction of HISCAM: A Stratification Scale Based on Social Interactions for Historical Research. *Historical Methods, 46*(2), 77–89.

Levine, J. H., & Spadaro, J. (1988). Occupational Mobility: A Structural Model. In B. Wellman & S. D. Berkowitz (Eds.), *Social Structures: A Network Approach* (pp. 452–475). Cambridge: Cambridge University Press.

Mills, C. (2007). Unequal Jobs: Job Quality and job Satisfaction. In P. McGovern, S. Hill, C. Mills, & M. White (Eds.), *Market, Class and Employment* (pp. 233–282). Oxford: Oxford University Press.

Minnesota Population Center. (2015). *Integrated Public Use Microdata Series, International: Version 6.4 [Machine-Readable Database]*. Minneapolis, MN: University of Minnesota.

North Atlantic Population Project and Minnesota Population Center. (2008). *NAPP: Complete Count Microdata. NAPP Version 2.0 [Computer Files]*. Minneapolis, MN: Minnesota Population Center [distributor]. Retrieved from http://www.nappdata.org.

ONS. (2017). *Labour Force Survey—User Guidance*. Retrieved March 1, 2017, from https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyuserguidance.

Penn, R. D. (1985). *Skilled Workers in the Class Structure*. Cambridge: Cambridge University Press.

Prandy, K. (1986). Similarities of Life-Style and the Occupations of Women. In R. Crompton & M. Mann (Eds.), *Gender and Stratification* (pp. 137–153). Cambridge: Polity Press.

Prandy, K. (1990). The Revised Cambridge Scale of Occupations. *Sociology, 24*(4), 629–655.

Prandy, K. (1998). Class and Continuity in Social Reproduction: An Empirical Investigation. *Sociological Review, 46*(2), 340–364.

Prandy, K. (1999). Class, Stratification and Inequalities in Health: A Comparison of the Registrar-General's Social Classes and the Cambridge Scale. *Sociology of Health & Illness, 21*(4), 466–484.

Prandy, K. (2000). Class, the Stratification Order and Party Identification. *British Journal of Political Science, 30*, 237–258.

Prandy, K., & Bottero, W. (1998). The Use of Marriage Data to Measure the Social Order in Nineteenth-Century Britain. *Sociological Research Online, 3*(1), U43–U54.

Prandy, K., & Bottero, W. (2000). Reproduction Within and Between Generations—The Example of Nineteenth-Century Britain. *Historical Methods, 33*(1), 4–15.

Prandy, K., & Lambert, P. S. (2003). Marriage, Social Distance and the Social Space: An Alternative Derivation and Validation of the Cambridge Scale. *Sociology, 37*(3), 397–411.

Rytina, S. (1992). Scaling the Intergenerational Continuity of Occupation: Is Occupational Inheritance Ascriptive After All? *American Journal of Sociology, 97*(6), 1658–1688.

Rytina, S. (2000). *A Keynote of Hierarchy, an Echo of Class: The Full Texture of Mobility in England and Wales.* Paper Presented at the RC33—Fifth International Conference on Logic and Methodology, Cologne.

Savage, M., Cunningham, N., Devine, F., Friedman, S., Laurison, D., McKenzie, L., Miles, A., Snee, H., & Wakeling, P. (2015). *Social Class in the 21st Century*. London: Pelican Books.

Statistics Finland. (1996). *The Finnish Census Panel Data*. Helsinki: Statistics Finland.

Statistics Sweden. (2016). *MONA—Microdata Online Access*. Stockholm: Statistics Sweden. Retrieved from www.scb.se/mona-en.

Stewart, A., Prandy, K., & Blackburn, R. M. (1980). *Social Stratification and Occupations*. London: Macmillan.

Toubol, J., & Larsen, A. G. (2017). Mapping the Social Class Structure: From Occupational Mobility to Social Class Categories Using Network Analysis. *Sociology, 51*(6), 1257–1276.

Treiman, D. J. (1977). *Occupational Prestige in Comparative Perspective*. New York: Academic Press.

University of Essex, & Institute for Social and Economic Research. (2010, July). *British Household Panel Survey: Waves 1–18, 1991–2009 [Computer File]* (7th ed.). Colchester, Essex: UK Data Archive [distributor], SN: 5151.

Zijdeman, R. L., & Lambert, P. S. (2010). Measuring Social Structure in the Past: A Comparison of Historical Class Schemes and Occupational Stratification Scales on Dutch 19th and Early 20th Century Data. *Belgian Review of Contemporary History, 40*(1), 1111–1141.

# 6

# Constructing CAMSIS Scales

## 6.1 Introduction

In this chapter, we describe the construction of CAMSIS scales. We begin with a brief illustration of software code (Sect. 6.2). We then comment on the underlying data on social connections between occupations which are used (Sects. 6.3, 6.4, and 6.5), and some of the issues that emerge in relation to different ways of using data (Sects. 6.6, 6.7, and 6.8). Next, we cover the techniques of statistical analysis (Sect. 6.9). Lastly, we comment on ways of using the results (Sect. 6.10) and on options for undertaking the whole process in a more automated way (Sect. 6.11). We draw primarily upon approaches that have been used hitherto in the CAMSIS project. More materials and supplementary resources such as samples of software code can be found on the CAMSIS project webpages. We also stress that analysts have some autonomy when generating CAMSIS scales—the exact results of any particular scale estimation are likely to be contingent upon operational decisions and subjective judgements that are made during the process.



P. Lambert, D. Griffiths, *Social Inequalities and Occupational Stratification*,
https://doi.org/10.1057/978-1-137-02253-0_6

## 6.2 Software Code

Relevant software tools and statistical techniques have evolved considerably through the years. The CAMSIS project website features a number of worked examples illustrating the derivation of CAMSIS scales which make use of tools that are available in the software packages lEM, SPSS, Stata, and R, and it would be feasible to achieve the same ends with other packages alternatively. In Fig. 6.1, we include a sample of code in one of those packages, Stata, using one of the relevant statistical techniques, correspondence analysis. The example illustrates the estimation of a CAMSIS scale and its implementation in further analysis for a publicly available dataset. Whilst the illustration in Fig. 6.1 is deliberately succinct, it is nevertheless a realistic example.

If you are already familiar with Stata code and with aspects of the analysis of social connections, then the contents of Fig. 6.1 may be quite easy to follow; if you are not, then they might at first sight appear gobbledegook! Broadly, segment (i) refers to opening the original data, which was downloaded as an extract from IPUMS-I (Minnesota Population Center 2015). Segments (ii) and (iii) involve preparing data for the SID analysis, with activities that involve 'recoding' the occupational measures to avoid having sparse coverage of certain occupational units, and defining combinations of occupations that are to be excluded from the analysis (what we call 'diagonals' and 'pseudo-diagonals'). Segment (iv) implements the main statistical analysis. Finally, segments (v) and (vi) involve processing the results of the analysis to generate an output file that lists occupations and their corresponding SID scores.

On the face of it, Fig. 6.1 suggests that the construction of CAMSIS scales based upon SID analysis is a straightforward process. Omitted from the code is recognition that many of the tasks involved are not trivial to resolve, but often require difficult decisions, grounded in considerable background work in reviewing and modifying data. The stages of work and relevant issues are elaborated upon in the discussions below, and in many cases further examples and details can be found on the CAMSIS project website.

```
*****************************************************************
*****************************************************************

**** Preliminary: Specifying locations/names for data and metadata files
global path1 "C:\camsis\countries\portugal\data\2011\" /* Location of IPUMS-I downloaded dat file and do file */
global path2 "C:\camsis\countries\portugal\scales\release\" /* Location for outputs - CAMSIS index file - to be generated */
global file4 "C:\data\resources\isco\labels\isco08_labels_2.do" /* File with value labels for ISCO-08 */
global path9 "c:\temp\" /* Location for temporary file storage */
*****************************************************************

**** (i) Open source data from IPUMS-I: Portugal 2011

do $path1\ipumsi_00054.do /* downloaded from ipumsi: all Portugal 2011 sample  */
             /*  with sex, occupation of ego and their spouse  ('attach characteristics') */
tab1 occ occ_sp /* this is occupation of ego and alter, 3-digit ISCO-08, valid codes 11-962 */
keep if sex==1 & sex_sp==2  & occ >= 11 & occ <= 962 & occ_sp >= 11 & occ_sp <= 962
codebook occ occ_sp, compact /* 69k both-working heterosexual couples, 125 occ units  */
gen freq=1
collapse (sum) freq, by(occ occ_sp) /* code to convert to a 'table format' dataset (one row per occ-occ_sp combination) */
sav $path9\file1.dta, replace /* saves a temporary copy of the data file */
/* Acknowledgement:
  Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 6.4 [Machine-readable
     database]. Minneapolis: University of Minnesota, 2015. The author wishes to acknowledge the statistical
   office that provided the underlying data making this research possible: National Institute of Statistics, Portugal. */
*****************************************************************

** (ii) Recoding occupational records

use $path9\file1.dta, clear
do $file4
label values occ isco08_min
label values occ_sp isco08_min /* attaches text labels to isco08 3-digit minor groups */
numlabel _all, add
tab1 occ occ_sp  [fw=freq] /* a few male (occ) / female (occ_sp) jobs n < 50: manual recode below */
tab1 occ occ_sp if occ~= occ_sp [fw=freq] /* here we are checking only the 'non-diagonal' occ numbers */
clonevar hocc3=occ
recode hocc3 211 212=213   223 225=226  232=233 262=263 314=311 322 324=325 413=411 ///
        431=432     521=524   632 634=631    754=753     816=818    941=962
clonevar wocc3=occ_sp
recode wocc3  11 21 31=31   111=112  133=134     211 215=214 212=213 223=226 232=235 252=251 ///
       314 315=313 322 324=325   412 413=411    621 622 632=633   712 713=711 741=742  754=731 ///
     811 812 813=818 831 832 834=833     931=932
tab1 hocc3 wocc3 [fw=freq]
tab1 hocc3 wocc3 if occ~=occ_sp [fw=freq]
* These are recodes chosen subjectively and by manual inspection: designed to be plausible/reasonable for the
*  jobs concerned, and to leave the remaining categories with 30+ non-diagonal cases representing them
egen hocc3n=sum(freq), by(hocc3)
egen wocc3n=sum(freq), by(wocc3)
tab hocc3 if hocc3n <= 30 [fw=freq]
tab wocc3 if hocc3n <= 30 [fw=freq] /* a robustness check: there should be no observations, after above recodes */
*****************************************************************

**(iii) Defining diagonals/pseudo-diagonals

gen diag=(occ==occ_sp) /* indicator of 'diagonality', i.e. husband and wife in same job */
tab occ_sp if ((occ >= 611 & occ <= 633) | occ==921 | occ==131) [fw=freq] /* checking on farming occs  */
tab occ if ((occ_sp >= 611 & occ_sp <= 633) | occ_sp==921 | occ_sp==131) [fw=freq]
gen farm1=((occ >= 611 & occ <= 613) | occ==631 | occ==131 | occ==921)  ///
       & ((occ_sp >= 611 & occ_sp <= 613) | occ_sp==631 | occ_sp==131 | occ_sp==921)
 /* commonly used farming pseudo-diagonal definition */
gen farm2=  (( (occ >= 611 & occ <= 613)) & occ_sp==911)    ///
  | (occ==711 & ((occ_sp >= 611 & occ_sp <= 613) | occ_sp==631))
/* extra, assymetric, and bespoke control for farming jobs in PT */
gen teach1=(occ >= 232 & occ <= 235) & (occ_sp >= 232 & occ_sp <= 235)
/* commonly used teaching pseudo-diagonal definition */
gen psd1=(diag==1 | farm1==1 | farm2==1 | teach1==1) /* will exclude if H-W pair meets any of these conditions */
tab psd1 diag [fw=freq]  /* identifies 7k occs that are 'diagonals' or else are a typical 'pseudo-diagonals' */
tab1 hocc3 wocc3 if psd1==0 [fw=freq] /* checking that still have decent numbers of cases net of exclusions */
*****************************************************************
```

**Fig. 6.1** Software code (Stata format) for an analysis using SID

```
** (iv) Running the Correspondence Analysis

ca hocc3 wocc3 if psd1==0 [fw=freq]
cabiplot /* this plot is convenient for diagnostic purposes */
* Subjective interpretation:
*     Having made the exclusions above, the first dimension scores do seem plausible as a CAMSIS scale
*     -> take these as the scale scores with no further adjustment
*****************************************************************

** (v) Re-scaling the results of the Correspondence analysis

predict hdim1, rowscore(1)
predict wdim1, colscore(1)
sav $path9\m1.dta, replace
use $path9\m1.dta, clear
keep occ hdim1
collapse (mean) mcam=hdim1, by(occ) /* ensures each occ has an hdim1 score linked to it (even if missing originally) */
sort occ
sav $path9\m2.dta, replace
use $path9\m1.dta, clear
keep occ_sp wdim1
collapse (mean) fcam=wdim1, by(occ_sp)
rename occ_sp occ
sort occ
sav $path9\m3.dta, replace
do $path1\ipumsi_00054.do /* reopens original census dataset as above , which features further variables, e.g. sex */
do $file4
label values occ isco08_min
label values occ_sp isco08_min
numlabel _all, add
tab sex /* we'll now use census data as representative sample of individuals (i.e. not both-working couples only) */
keep sex occ
sort occ
merge m:1 occ using $path9\m2.dta
drop _merge
sort occ
merge m:1 occ using $path9\m3.dta
drop _merge
summarize /* mcam and fcam are CA based scores for the individuals in the dataset */
summarize mcam if sex==1
replace mcam = 50 + 15*((mcam - r(mean)) / r(sd)) /* standarise to mean 50, sd 15  */
recode mcam -100/1=1 99/200=99  /* 'crop' values to min 1, max 99 */
summarize fcam if sex==2
replace fcam = 50 + 15*((fcam - r(mean)) / r(sd))
recode fcam -100/1=1 99/200=99
table occ, c(mean mcam n mcam mean fcam n fcam)
* At this point, we have scale scores standardised around a nationally representative
*   sample of men and women
*****************************************************************

** (vi) Exporting the results in an 'index file' of occupations and scores

collapse (mean) mcam fcam , by(occ)
replace mcam = round(mcam, 0.01)
replace fcam = round(fcam, 0.01) /* round data to two decimal points, for convenience */
keep if ~missing(mcam) | ~missing(fcam)
replace mcam = -9 if missing(mcam)
replace fcam = -9 if missing(fcam) /* gives score of -9 to occ if there were no men/women representing it in analysis */
label variable mcam "CAMSIS scale for males, Portugal, 2011"
label variable fcam "CAMSIS scale for females, Portugal, 2011"
rename occ isco08_3
label variable isco08_3 "ISCO-08 minor groups"
keep isco08_3 mcam fcam
saveold $path2\pt2011isco08_3.dta, replace
outsheet using $path2\pt2011isco08_3.dat, nonames nolabel replace
export excel using $path2\pt2011isco08_3.xls , firstrow(varlabels) replace
* (these last 3 lines save out  the 'index file' in three different formats - these files are also at www.camsis.stir.ac.uk)

*****************************************************************
******************************************************************
```

**Fig. 6.1** Continued

## 6.3 Sources of Data

At least at the time of writing, it is quite easy for social scientists to get access to large volumes of secondary data about individuals, their occupations, and the occupations of other people with a social connection to them—namely, the elements of data which are required to undertake social interaction distance analysis on occupations. Many of the contemporary CAMSIS scales, for example, are based upon census data accessed for free from IPUMS-I (Minnesota Population Center 2015), and most other SID projects are based upon alternative freely available large-scale secondary survey datasets (cf. de Luca et al. 2010). Indeed, the use of secondary data means that many SID projects are favourably transparent and replicable to others. As an example, the code in the first segment of Fig. 6.1 begins by opening up a secondary survey data file that was obtained for free from IPUMS-I—if you are motivated to do so, whilst reading this chapter you could download the same data and try to implement the same analysis.[1]

Various social connections held by the 'incumbents of occupations' could be studied. Connections of friendship, marriage, or cohabitation are most commonly used, but SID analyses on occupations using data on relationships of acquaintanceship, co-residence, shared family membership, intergenerational relationships (parent-child data), and intragenerational records (career data) have also been explored (see Sect. 5.1.6). Whilst the main outcomes of analysis are likely to be the same (Sect. 5.1.6), the type of social connection can have operational implications—marriage records, for instance, are much more readily available than most other forms of data on social connections, and access to greater volumes of data influences further analytical decisions, for instance, concerning the level of occupational detail at which the analysis is undertaken (see Sect. 6.5).

In general, we benefit from an abundance of data resources that could be used for SID analyses on occupations, and most perspectives on developments in data resources suggest that such opportunities should expand through time (e.g. Playford et al. 2016). Nevertheless, there are scenarios where relevant data is not so easily available. To apply SID analysis to historical datasets, analysts have often used by-product datasets, such as administrative records, which required considerable preparatory activities by relevant historians (cf. Lambert et al. 2013; Prandy and Bottero 1998). In some contemporary societies, suitable survey data is available in principle, but its use in practice is quite restricted—for example, substantial fees might be required to access the data, or there might be policies that set restrictive conditions on reuse of the data. In addition, some nations don't collect (or collect, but don't distribute) data with extended detail on occupations—data on occupations is sometimes thought to raise statistical disclosures risks and is sometimes deliberately removed from secondary datasets.

## 6.4 Data Preparation

The main unit of analysis in a SID approach is a record of a social connection. That is, a dataset is developed in which each row contains information on an individual ('ego'), plus linked information (including the occupation) on another individual ('alter'), who is socially connected to the ego. In some scenarios, secondary data originates in this format, for example, in the case of sample surveys where the main respondent is asked to describe their spouse's occupation as well as their own. In addition, some data providers, for instance, IPUMS-I, have services that prepare data in this format on behalf of analysts. Our illustrative example, in Fig. 6.1, begins from this point, using a dataset where each row contains data on the occupations of a nominated respondent ('occ') and their cohabiting spouse ('occ_sp'). In this example, the opening sections of the Stata code select only those records where the respondent is male ('sex==1') and their spouse is female ('sex_sp==2'), which means that the data comprises records on the occupations of heterosexual couples, where one variable ('occ') indicates the male's job and another variable ('occ_sp') indicates the female's job.

Alternatively, in many secondary studies, the relevant 'egos' and 'alters' initially contribute separate records to the dataset. This is a common scenario if using household survey data, where each row of the dataset typically stores information from a different member of the household. The job of linking together information from separate (but related) records can be somewhat challenging; bespoke software code is sometimes

**Table 6.1** Depiction of the process of estimating CAMSIS scale scores

| | job units | 0 | 1 | 2 | 3 | … | 98 |
|---|---|---|---|---|---|---|---|
| Husband's job units | ***Row / column scores*** | ***88*** | ***71*** | ***66*** | ***58*** | ***…*** | ***25*** |
| | | | | *<- Cell counts ->* | | | |
| Managers) | ***80*** | 880 | 481 | 257 | … | … | 2 |
| 1 (Administrative & Finance Managers) | ***75*** | 229 | 461 | 163 | … | … | 4 |
| 2 (Farm, Construction, Education Managers) | ***64*** | 172 | 265 | 1460 | … | … | 2 |
| 3 (Engineering, Medical, Science & Services Managers) | ***70*** | … | … | … | … | … | … |
| … | ***…*** | … | … | … | … | … | … |
| … | ***…*** | … | … | … | … | … | … |
| 98 (Military occupations) | ***30*** | 14 | 40 | 23 | … | … | 96 |

Notes: Based upon IPUMS-I data for USA 2005, SOC two-digit groups. Table depicts selected frequencies for the combination of row and column categories (husbands' and wives' jobs), along with the job categories (0, 1, …, 98) and the estimated CAMSIS scale scores for both the husbands' jobs and the wives' jobs (e.g. 88, 71, 80, 75, etc.)

required to restructure the dataset into one where each record indicates a different social connection (there are some relevant examples of suitable code on the CAMSIS website, or see Longhi and Nandi 2015, c6, for a recent elaboration using Stata).

Ultimately, the SID analysis evaluates a large contingency table that denotes the volume of connections for each ego-alter occupational combination. Table 6.1 shows an illustrative example of the contingency table that is used. The first record in each cell gives the number of occurrences of each husband-wife combination—for example, there are 481 occurrences of a husband in job 0 combined with a wife in job 1. The 'row' and 'column' scores shown in Table 6.1 are the results of the statistical analysis (see Sect. 6.9). They indicate numeric scores calculated by a statistical model that seeks to maximise fit between the actual number of cases in each cell and model-based predictions—these scores, in turn, become the SID scale values.

To analyse the matrix shown in Table 6.1, we need the count of records for each possible combination. In fact, this information can actually be stored in an alternative, more efficient format. The 'matrix' format, shown in Table 6.1, contains all the required information, but is relatively difficult to work with in conventional software packages. A 'microdata' format, described above, in which each case in a database refers to a different ego-alter combination, also stores all relevant data, but potentially uses up a lot of rows in a dataset (e.g. the data from IPUMS-I in Fig. 6.1 begins with approximately 69,000 records, one for each male-female couple with occupational data). For reasons of efficiency, the information that is required for a SID analysis is instead typically stored in a third layout, which is often called a 'table' format. In this arrangement, each entry features a different permutation of ego and alter occupation, and a further column of data records the frequency of that combination (i.e. the cell counts in the ego-alter matrix). Table 6.2 illustrates this structure, which is generally more efficient and easy to work with. In Fig. 6.1, the line within the first section that begins 'collapse …' restructures the data from microdata to table format, converting it from a database with 69,000 different records (one per couple), to one with at most 15,625 records (i.e. one for every ego-alter occupational combination—there are 125 different occupations in this example, so there could be at most 125 * 125 = 15,625 records in the table format file).

**Table 6.2** Depiction of the 'table' format commonly used to store the data for a SID analysis

| Record | Husband's occupation | Wife's occupation | Frequency |
|---|---|---|---|
| 1 | 0 | 0 | 880 |
| 2 | 0 | 1 | 481 |
| 3 | 0 | 2 | 257 |
| … | … | … | … |
| 100 | 0 | 98 | 2 |
| 101 | 1 | 0 | 229 |
| 102 | 1 | 1 | 461 |
| 103 | 1 | 2 | 123 |
| … | … | … | … |
| … | … | … | … |
| … | … | … | … |
| 9799 | 98 | 96 | … |
| 9800 | 98 | 97 | … |
| 9801 | 98 | 98 | 96 |

Note: Here, there are 9801 permutations of husband-wife job combination so this dataset has 9801 cases. It is common practice to delete any combinations which are not represented by any cases (i.e. for which the 'Frequency' is zero)

In some datasets, there may be several ego-alter connections which could be related to each other. For example, if we have data on the occupations of Andrew, Bill, and Colin, who are friends who all attend the same social club, this represents three records of social connections (Andrew-Bill, Bill-Colin, and Andrew-Colin). Each of these contributes a new record to the dataset of pairs which could be analysed in a cross-tabulation (between 'friend 1' and 'friend 2'). In such circumstances, the stored data should include some features (e.g. a shared identifier code) that records the mutual connections shared by this group. Analysis might ideally feature further adjustments to reflect the shared origins of the three linked pairs.[2]

## 6.5 Units of Occupational Measurement

### 6.5.1 Disaggregating by Employment Status

The SID approach uses occupations as a 'marker' of a person's situation, and it is plausible that other aspects of information could contribute to defining that marker alongside occupations. CAMSIS studies often use additional information on aspects of 'employment status' in this way.[3] Data on employment status concerns the contractual and organisational circumstances under which a person works, such as if they are self-employed or an employee, and whether or not they manage or supervise other employees. This data is widely recorded and commonly available in a standardised format (e.g. Elias 2000). Moreover, employment status is used in many other occupation-based measures, such as in occupation-based social class schemes (e.g. Rose and Harrison 2010). The option here with regard to a SID analysis is whether to treat an occupational category (e.g. 'Plumber') as the analytical unit itself or whether to disaggregate that category by information on employment status and allow for estimation of different scores for each unit (e.g. 'Plumber (self-employed with employees)', 'Plumber (self-employed, no employees)', 'Plumber (employee)'). The rationale is that employment status divisions might indicate important and consistent divisions within occupations. Theoretically, the relative importance of employment status divisions might fluctuate over time and between countries, and it is plausible that divisions by employment status might be important within some occupations but not within others.

Operationally, analysis proceeds by defining units according to their occupational and employment status combination, then undertaking the SID analysis on the social connections between these units. This adds to the complexity of the scale estimation process, since it may involve considerably expanding the number of categories—for instance, considering Table 6.2, we could add rows for every permutation of ego-alter employment status within each occupational combination. This disaggregation typically raises difficult analytical decisions with regard to 'pseudo-diagonals' (see discussion in Sect. 6.7) and 'sparsity' (see Sect. 6.6). Ordinarily, after generating CAMSIS scale scores for 'occupation-by-status' units, it is also desirable to specify a category of occupation combined with 'unknown employment status' and assign a scale score to it.[4] This has the important convenience of allowing users to allocate CAMSIS scale scores to records or datasets when the occupational title is known but the employment status is not. The CAMSIS project website features illustrations of software code that define 'occupation-by-status' categories and use those in each stage of the scale estimation process. However, because it is relatively more complicated to do so, our illustrative example in Fig. 6.1 avoids this and illustrates a CAMSIS scale that uses occupational titles only.

Despite the extra effort, there is a fair argument that if employment status data is available, then it is sensible to use it. Theoretically, the difference could matter. Empirically, validity analyses suggest that SID measures which incorporate disaggregation by employment status have slightly stronger empirical associations with other measures that they are expected to correlate with (e.g. Prandy and Lambert 2003). For such reasons, the standard approach in the CAMSIS tradition has been, when reasonably possible, to use employment status information in addition to occupational title information. In practice however, employment status disaggregations have been used in some SID analyses, and not others. They are used in the bulk, but not all, of the contemporary CAMSIS scales distributed from that project's website. When a CAMSIS scale has been calculated without using employment status information (described on the CAMSIS website as being at the 'title-only' level), this is often because the source data provided insufficient volumes of cases to allow for statistically reliable disaggregations. In addition, there are also some SID applications where employment status distinctions are not systematically included, but are featured *de facto* because some of the relevant occupational categories are themselves defined by employment status (e.g. Levine and Spadaro 1988).

### 6.5.2 Revisiting Occupational Details

We have already discussed (Sect. 3.5) the relative value of using detailed or 'fine-grained' data about occupations. Detailed occupational data is usually preserved on standardised occupational taxonomies, and there are various reasons to expect that quite detailed differences between occupational positions are (or, at least, could be) socially significant (especially Weeden and Grusky 2012). Accordingly, most (but not all) SID scales released through the CAMSIS project differentiate between relatively fine-grained occupational units (typically between 100 and 500 occupational positions), but other SID scales have also been constructed that use far fewer occupational units (see also Sect. 4.2). To the best of our knowledge, those applications of SID to occupations that have worked at a more aggregate level have done so largely for pragmatic reasons, concerning the minimum representation of cases per category, and would not in principle be opposed to working at a more disaggregate level (e.g. Chan 2010, chap. 2). Nevertheless, is it plausible that the important dimensions of occupational difference are perfectly well captured by 17 (e.g. Levine and Spadaro 1988) or 32 (e.g. Chan and Goldthorpe 2004) occupational units? Or is it more appropriate to use much more fine-grained occupational measures when they are available, such as in the studies linked to the CAMSIS tradition? In practical terms, an analyst using SID techniques would often have to make a decision relatively early in the process concerning which level of detail to use from the data available to them.[5]

One simple clue comes from comparing Fig. 4.1 (Chap. 4) with Fig. 6.2. Both give graphical representations of the relative frequency of

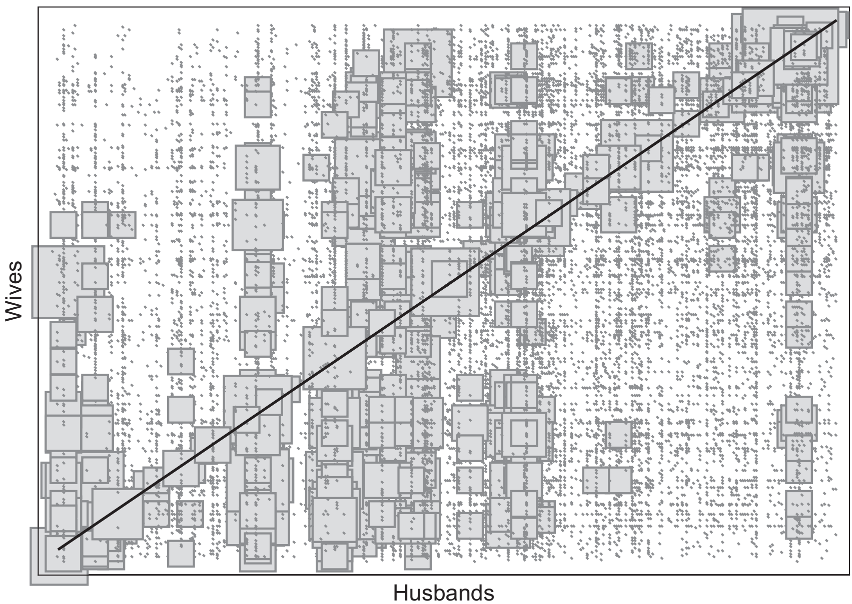



**Fig. 6.2** SOC2010 unit groups for husbands and wives, ranked by CAMSIS score. Source: UK Labour Force Surveys 2010–2012

different pairs of occupations held by cohabiting heterosexual couples in the contemporary UK. Figure 4.1 used only a small number of different occupational categories (9), and Fig. 6.2 represents the structure involving a much larger number of occupations. In both figures, the size of the plots is proportional to the number of cases, and there are obvious links between the patterns of occurrence and a structure of social stratification and inequality. Descriptively, however, Fig. 6.2 also suggests there are empirical nuances to the data that would be overlooked if we worked with more aggregated occupational units.

Figure 6.3 explores the issue more systematically. On its vertical axes, we show variations in the first three dimensions of a SID analysis of husband-wife occupational combinations, on the one hand, for an analysis conducted at ISCO-88 two-digit level (upper panel) and on the other for an analysis conducted at ISCO-88 four-digit level (lower panel). The points on the graph represent occupational units at the respective level of aggregation, with size proportional to the number of cases; points are plotted three times for the same occupations, which are vertically aligned in relation to the units' ISEI scores. As a general point, we can see in this example that the differences in the properties of the different measures are not extreme. In each example, regardless of the level of occupational detail, the first dimension correlates quite well ($r = 0.8$) with the ISEI socio-economic index, and there are seemingly comparable differences between units in the other dimensions. Although not shown, we have undertaken similar comparisons for many other datasets, and these patterns (for Romania) are very typical of those seen across other countries and time periods. de Luca et al. (2010) similarly explore the differences in this first dimension of a SID analysis when it is calculated at more and less aggregate levels and come to the conclusion that the differences are of little practical importance and so working at more aggregate levels is a sensible response.

However, it is also apparent from Fig. 6.3 that the first dimension is not identical in the more and less aggregated solutions, and moreover that this might have consequences for some interpretations. For example, the relative order of the two large farming groups (61, farmers; 92, farm labourers) seems counter-intuitive in the first dimension of the two-digit solution, but not at the four-digit level (indeed, the position of these

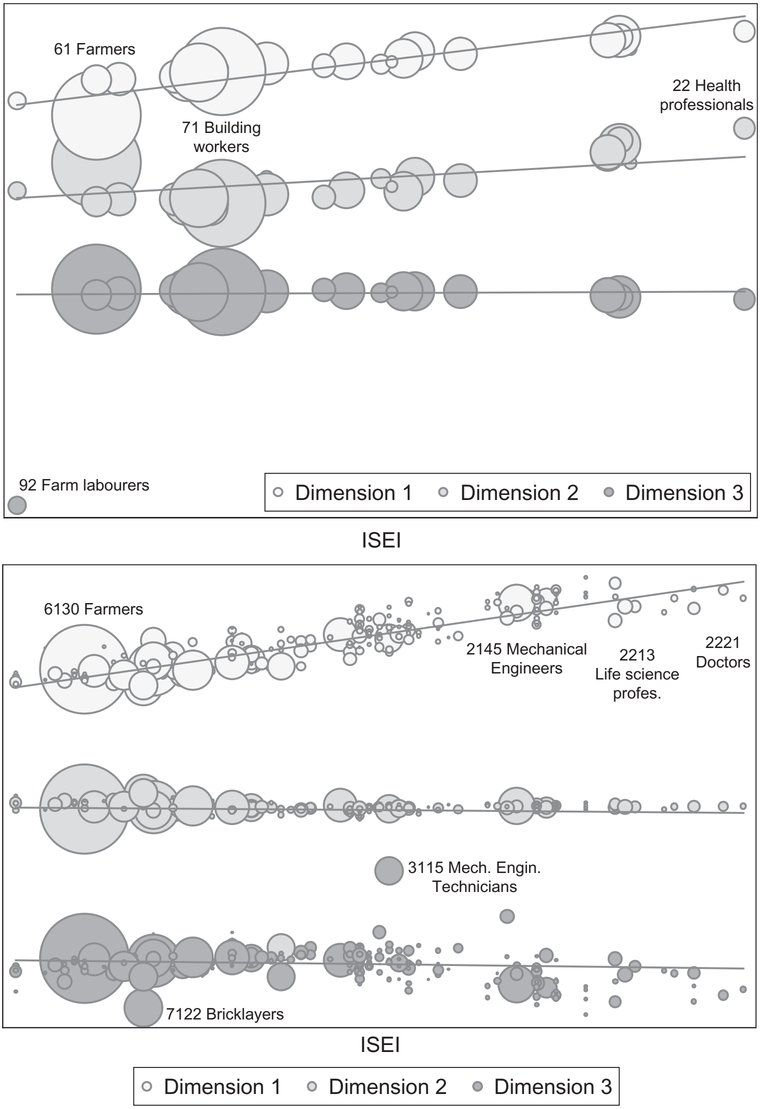



**Fig. 6.3** First three dimensions of the social interaction distance solution for husband-wife occupational combinations, Romania 2002. Notes: Upper panel shows the solution with two-digit occupational detail. Lower panel shows the solution with four-digit occupational detail. Horizontal axes are ISEI scores for occupations; vertical axes are scores in the respective dimension for the SID solution (rescaled to aid display). Points show scores for male occupations only, with the size of points proportional to the number of males in the occupation as used within the analysis (males living with a co-resident female spouse, and both partners must have valid occupational codes)

groups seems to be influential to both the first and second dimensions at the two-digit level, in a way that is less pronounced at the four-digit level). Equally, at the top end of the scale, the two-digit solution has much less differentiation of values than the four-digit solution, and indeed some occupations towards the higher values of the dimension are allocated substantially different placements between the two solutions.

Furthermore, the solution with more detailed occupational categories seems more plausible since it reveals differentiation in positions that would not be recorded when using more aggregate measures. In this example, we suspect that the scale values are particularly shaped by the impact of the largest categories, such as farming occupations, which have different profiles of connections to other occupations in the two solutions (because those occupations themselves are recorded with different levels of aggregation). It is also apparent that the second and third dimensions have somewhat different properties at the more and less aggregated level. For example, in the four-digit solution, the second dimension is orthogonal to ISEI, but there is a slight correlation in the two-digit solution; more generally, the most influential occupations in determining these subsidiary dimensional structures (or, more accurately, the most influential combinations of husbands and wives) are not the same, but reflect different influences at the different levels.[6] Moreover, since it is likely that highly specific occupational combinations play a different role in the SID solution depending upon the level of aggregation, it is also possible that the key first-dimensional solution, in a more aggregated version, might be less likely to achieve a clear separation between the influence of other separate mechanisms behind social interaction (see Sect. 6.8 and Chap. 11). For instance, in contemporary societies with high levels of occupational gender segregation, it is well known that gender segregation is only fully recognised when occupations are differentiated at a relatively detailed level (e.g. Charles and Grusky 2004), so accordingly the influence of gender segregation upon social interaction patterns might filter through to the first-dimensional solution unless a fine level of disaggregation is possible.

We argue therefore that in most situations the SID solution with more disaggregation of categories is likely to be preferable, even though it is unlikely to be hugely different from that obtained at a more aggregate level. This principle is empirically vindicated, and there are good theoretical reasons for following it. Nevertheless, the level of detail of occupations might reasonably be balanced against other priorities. Aside from the extra time and complications raised by using more disaggregate measures, there is also a very reasonable concern about the statistical reliability of scale scores that are estimated for high numbers of occupational units (since by definition the volume of cases per unit is decreased—see further in Sect. 6.6). For such reasons a decision to restrict analysis to a relatively broad-grained level of occupational detail is defensible, although for us not entirely convincing.

## 6.6 Sparse Representation of Occupations

If you have accessed the Portuguese data that is used in the software code example of Fig. 6.1, you might already have noticed that several occupations in that file are represented by only a few cases. The sample is a relatively large one, of over 69,000 'both working' male-female couples, and the occupational scheme features 125 different occupational categories. This suggests that the data could feature up to 15,625 possible husband-wife occupational combinations.[7] Hypothetically, this equates to an average of 500 records per occupation and an 'average cell size' of around 4.4. However, because occupational units are highly skewed in their social distribution, we can expect that not all of the 15,265 cells will have occurrences in them, and we can also expect that whilst many occupations will be represented by far more than 500 incumbents, others will be represented by far fewer. Indeed, in the data as a whole, only 6428 out of the possible 15,625 husband-wife combinations feature any cases at all. Amongst the men, the occupation represented by the most respondents has 5513 cases (builders), but there are 15 occupations represented by less than 30 males, including one occupation (nursing assistants) that is represented by only four. The skew amongst the women is even more pronounced, including one occupation with just one incumbent (train drivers), 27 occupations with less than 30, whilst the largest occupation (cleaners/domestics) is represented by more than 7000 incumbents.

Such 'clumping' in occupational categories is a standard occurrence across societies, and this often introduces challenges of how to treat those that are sparsely represented (even when very large datasets are available for analysis). The SID analysis seeks to make inferences about the social interaction profile of the occupations in question. If an occupation is not represented by any cases at all within a dataset, we cannot reasonably make inferences about it; if it is represented by some cases, but only a small number, we face a challenge of sampling representation. One response, from a statistical perspective, is to introduce uncertainty statistics such as standard errors for the relevant scale scores. In the SID tradition, uncertainty statistics can be calculated (see Sect. 6.9), but they have not hitherto been used as a matter of routine, and their implications are ambiguous.

A second response to sparsity is routinely implemented in examples of a SID analysis of occupational units. It involves, in anticipation of issues of statistical reliability, the decision to merge those occupational units which are represented by only small numbers of cases with others (thus reducing the total number of units, and of cells in the cross-tabulation). In the CAMSIS tradition, the convention has been to perform analysis upon as many different occupational titles as the data will allow, but to specify a minimum number of cases per occupational unit—the number 30 is the most commonly used—and to merge occupations with other categories if there are fewer than 30 individuals representing them. The threshold of 30 has emerged, through empirical experience within the CAMSIS tradition, as a plausibly effective minimum standard that nevertheless sustains quite fine-grained occupational comparisons; it is not however justified by a specific power analysis or statistical evaluation, and smaller or larger minimum thresholds could plausibly be substituted.

When merging sparse units, it makes sense to try to reallocate the occupations into other units that have similar qualities. Decisions about merging occupational units can be made in a largely *a priori* way. This approach might exploit hierarchical aggregations in occupational units, and can potentially be done in an automated or semi-automated way (see also Sect. 6.11). A typical example might be to set a rule that says that if there are, say, fewer than 30 cases representing an occupation (e.g. occupation '723'), then it should be merged with the largest other occupation from its surrounding minor group (e.g. '72'), or from its surrounding major group (e.g. '7') if necessary.[8] Historically, however, most existing CAMSIS versions used an alternative approach, whereby the analyst makes responsive decisions over how best to combine sparse categories when they occur. For example, in Fig. 6.1, the lines of code in segment (ii) illustrate the implementation of bespoke recoding decisions which are undertaken after inspection of the data and involve separate decisions for the male and female occupational variables. An advantage of this approach is the ability to make bespoke judgements about suitable recodes that don't necessarily rely upon administrative taxonomies, such as connecting jobs that have similar characteristics but are coded in different areas of the taxonomy.

It is useful to highlight that when the 'ego' and 'alter' populations are structurally different (such as when egos represent males and alters represent females), it is naturally the case that the recoded occupations involve different categories for egos and alters, and the SID analysis will take place on a non-square matrix. This is standard practice in the CAMSIS tradition—in Fig. 6.1, for instance, the units of 'hocc3' and 'wocc3' on which the SID analysis is applied do not have the same categories. However, if there is no structural difference between egos and alters (e.g. if they represent the jobs of 'friend 1' and 'friend 2'), it is more compelling to recode both measures by the same criteria, and continue to work in a square matrix.

Further complications can arise when units are defined by the combination of occupational taxonomy and employment status categories. In this scenario, it is plausible to prioritise occupational similarities, or employment status similarities, or take an intermediate position. The question here is whether it is better to recode, say, 'self-employed bus drivers', with other bus drivers, or with other people who are self-employed but in different occupations. Most previous CAMSIS studies have prioritised employment status similarities within occupational sectoral boundaries—meaning that the self-employed bus drivers might be merged with, say, self-employed truck drivers, but not with self-employed restaurateurs. The judgement here is that the employment status category, within sectoral boundaries, is the relatively stronger indicator of the incumbent's social circumstances.

More generally, and also recognising the cognitive and operational difficulties of using a great many occupational units, SID analysts will often disregard some elements of occupational differences in some other strategic way. In this manner, typically by taking account of some other external data about occupational structures, analysts might combine together similar occupational titles, disregard some or all divisions of employment status, and/or work at a more aggregate level in terms of occupational units than is necessarily required. Several of the published CAMSIS scales feature some strategic action that limits the range of occupational units used in the analysis.

Merging some occupational categories is a common feature in most CAMSIS scale calculations, but the implications can become rather messy. When scores are published at a later stage (see also Sect. 6.11), it is appropriate to distribute them to all of the original occupational units, but this can make it easy to overlook that a scale value for one occupation might have been derived for a combined category: for instance, when reviewing a CAMSIS score for 'Midwife', say, we might actually be looking at a score that was calculated on a category that combined 'Nurses', 'Midwives' and 'Other specialist nurses'. Metadata is usually attached to convey this information (see Sect. 6.11), but it is easily overlooked. In our view, the very fact that the users are often interested in looking up specific scores for occupations provides good grounds for working at a disaggregate level, with as few recodes as is feasible.

Clearly, the decisions that might be made in response to the sparse representation of some occupations are substantially a matter of judgement, and it is worth reiterating that the construction of SID scales for occupations is not an entirely neutral process—each version will be influenced by the decisions and strategies deployed by the analyst who generated it. Additionally, the variety of plausible responses to sparsity should also make obvious the benefits of adequately documenting any adjustments that are made. Ideally, an audit trail should be provided for a SID scale that allows others to identify what actions were taken—at the CAMSIS website, for example, many versions are accompanied by supplementary documentation files in the form of data and software code that provides this information for others.

## 6.7 Excluding 'Diagonals' and 'Pseudo-Diagonals'

Statistical methods such as correspondence analysis are used to identify distinctive orthogonal dimensions to the social interaction structure. The main and subsidiary dimensions that are identified usually differentiate positions across a wide range of different occupations. However, there are some scenarios where highly specific relationships between occupations lead to non-standard volumes of social interactions, and in these cases, in the SID tradition, it is common practice to exclude them from further influence in the dimensional reduction calculation. When this is done, we usually refer to excluding, or setting parameters to deal with, 'diagonals' and 'pseudo-diagonals'.

The particular circumstance of a social connection between two people in exactly the same occupation is often called a 'diagonal'.[9] For several reasons, diagonal combinations of occupations are often unusually common. A husband and wife, for example, might have the same occupation because they work together on the same family business, or because they met at their workplace, or during their occupational training. Two friends might have the same occupation, for example, because they met at work, or perhaps because one let the other know about a vacancy coming up at their workplace.

However, similar mechanisms sometimes explain the disproportionate occurrence of specific connections between occupations that are not the same. A widely recognised example is that the combination of male doctors married to female nurses arises in many societies much more often than might be statistically predicted given the average circumstances in the stratification structure that are associated with these two occupations. In this case, it is easy to imagine that workplace proximity means that doctors and nurses are in regular contact with each other, leading in turn to higher-than-average partnership formations. There are many other mechanisms that can lead to specific pairs of occupations having many more (or fewer) occurrences than we might otherwise expect. For instance, in many countries a 'train driver' and a 'train guard' are different occupations, but they have above-average volumes of social connections, and it is easy to imagine that these are driven by friendship formation during the process of working together. In the SID approach, we describe these sort of instances with the terminology of 'pseudo-diagonals': such combinations are not 'diagonals' (the same occupation for both persons), but they have similar qualities to diagonals and arise because of similar social mechanisms.

Consider Table 6.3. It shows hypothetical husband-wife data for a society where social stratification influences social connections of marriage in the usual way. It is particularly common in this society for husbands and wives to be in exactly the same occupation. However, for one male occupation ('farmer'), the wives of men in this occupation have mostly been coded as a different occupation, namely, 'agricultural labourer'.

Table 6.3 shows three sets of possible scores in the first empirical dimension for the husbands' occupations: a model which uses all the

**Table 6.3** Illustrating the influence of 'diagonals' and 'pseudo-diagonals' upon SID solutions

| | *Husband's job* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| *Wife's job* | | | | | | | | |
| **1. Medical doctor** | 25 | 10 | 2 | 0 | 0 | 0 | 1 | 0 |
| **2. Teacher** | 20 | 30 | 10 | 3 | 2 | 1 | 5 | 0 |
| **3. IT assistant** | 5 | 5 | 30 | 5 | 5 | 0 | 0 | 0 |
| **4. Shop assistant** | 2 | 5 | 10 | 25 | 20 | 20 | 5 | 10 |
| **5. Taxi driver** | 0 | 0 | 5 | 0 | 20 | 0 | 5 | 0 |
| **6. Assembly line worker** | 0 | 0 | 5 | 5 | 5 | 50 | 0 | 15 |
| **7. Farmers** | 0 | 0 | 0 | 0 | 5 | 0 | 10 | 0 |
| **8. Agricultural labourers** | 0 | 0 | 0 | 0 | 0 | 5 | 50 | 10 |
| ***Dimension 1 scores…*** | | | | | | | | |
| Using all cases | 76 | 71 | 59 | 48 | 45 | 38 | 35 | 36 |
| Excluding diagonals | 56 | 61 | 53 | 64 | 59 | 46 | 19 | 62 |
| Excluding diagonals and the Farmer-Agricultural labourer pseudo-diagonal | 73 | 60 | 54 | 55 | 38 | 29 | 60 | 31 |

Note: Synthetic data. Dimension scores are the correspondence analysis scores, rescaled to mean 50, SD 15. The light-shaded cells would be considered 'diagonals', and the dark-shaded cell represents the 'pseudo-diagonal' (husband farmer, wife agricultural labourer) cases, a model which drops all the husband-wife combinations which are the same (often called 'diagonals'), and a model which also drops the specific combination of husband farmers married to wife agricultural workers (a 'pseudo-diagonal'). In this example, the last model gives the most plausible stratification dimension as its first-dimensional solution. Arguably, the dimensions reported in the first two models are partially 'polluted' by these forces of 'situs' (husbands and wives often in the same job) and administrative codification (the wives of farmers being labelled 'agricultural labourers' rather than 'farmers'). The first model, for example, puts the agricultural jobs at the bottom of the scale because the high density of diagonals separates them from other social positions. The second model is inappropriately skewed by the farmer-agricultural labourer combination, primarily differentiating, for men, between farmers and all others. Only the third model seems to identify a dimension that we find persuasive as one of 'stratification', achieved as it has been by separating out the influences of the 'diagonals' and 'pseudo-diagonals'.

In general, if any occupational combinations are treated as 'diagonals' or 'pseudo-diagonals' and excluded (or parameterised) in a SID analysis, then it can have some consequences for the statistical results and therefore for our understanding of the structure of social stratification itself. As such, an analyst's operational decisions can be of great consequence to the final results, but if the issues are simply neglected, it is quite plausible that those different social mechanisms (that drive pseudo-diagonal occurrences) are inappropriately entangled with the core SID solution. Reflecting their potential importance, the next subsections expand the theme of diagonal and pseudo-diagonal combinations.

### 6.7.1 Identifying and Implementing Diagonal and Pseudo-Diagonal Adjustments

Certain occupational combinations might be defined as 'diagonals' or 'pseudo-diagonals' *a priori* or in response to interactive evaluation of the properties of a dataset. A key (subjective) criterion is that the analyst can think of a theoretical reason why the specific occupational combination is likely to occur more often than would otherwise be expected and is satisfied that this reason is separable to the general structure of social stratification that is associated with the first dimension of the SID solution. An example is illustrated within Fig. 6.1, section (iii), when selected husband-wife occupational combinations are defined as diagonals and pseudo-diagonals, and then excluded from further analysis. One of the combinations ('farm1' in Fig. 6.1) identifies male and female couples when both work in agricultural jobs that don't have the same occupational codes (the assumption is that the couple share institutional proximity and/or joint business ventures). Another example in Fig. 6.1 ('teach1') identifies male and female couples who both work in professional teaching but have different occupational units (here the assumption is that many of these couples share workplaces and/or training institutions).

After relevant diagonal/pseudo-diagonal combinations have been identified, they are typically excluded from further analysis, but it is also possible to use a more explicit model-based solution, by fitting parameters to the relevant combinations rather than excluding them from the data. The CAMSIS website includes sample code showing examples that exclude cases, and illustrating one example of parameterising the combinations by using the lEM package.[10]

The identification of occupational combinations as diagonals or pseudo-diagonals is often ambiguous. For any given occupational taxonomy, there are often certain specific combinations for which we would expect pseudo-diagonals to occur in social connections—for instance, because the jobs themselves are routinely undertaken together. However the identification of plausible pseudo-diagonals is often contingent upon local expertise and subjective judgement. Indeed, even the concept of 'diagonals' is not always crisply defined, because the relative granularity of occupational taxonomies might mean that a diagonal as defined by the taxonomy doesn't really reflect both people holding exactly the same occupation (for instance, aggregations in the taxonomy might put different occupations together in the same unit group code). In some CAMSIS scales, analysts have also defined diagonals at a broader level of occupational aggregation than the units on which the scale is estimated (for instance, we might say that all four-digit occupational combinations that are within the same three-digit 'minor group' are treated as diagonal). This strategy has a pragmatic convenience because it will exclude diagonals and will also pick up on many pseudo-diagonals that occur within minor groups, which the analyst might not otherwise have noticed. However this approach is also potentially unsatisfactory since it risks unnecessarily excluding cases that are not in any way 'pseudo-diagonal', as well as missing out other possible pseudo-diagonals that cross different minor groups. It should also be borne in mind that there is typically an interactive relationship between issues of the sparse representation of occupations (Sect. 6.6) and those of excluding diagonals or pseudo-diagonals. The relevant number of records used in analysis should be calculated as those present after removing or parameterising the influence of diagonals and pseudo-diagonals. This implies that the strategies described in Sect. 6.6 are most appropriately implemented after consideration has been paid to the exclusion or parameterisation of cases.

In summary, whilst the definition of diagonals and pseudo-diagonals in a SID analysis can be consequential, there are no easy or widely agreed procedures for identifying cases that might be treated as either. It is sensible to consider sensitivity analysis over different options, and to document the procedures undertaken, including description of which occupational combinations were treated as diagonals and pseudo-diagonals in a particular analysis. In practice, if a highly influential pseudo-diagonal is not dealt with, this is often obvious from the initial dimension reduction solution (e.g. in societies with a large agricultural sector, we might see a first dimension that is defined entirely around agricultural sector pseudo-diagonals, which we might then code out in an iterative second stage of analysis). It is also worth being aware that the impact of coding specific occupational combinations as diagonals or pseudo-diagonals is generally greater when working with more disaggregated occupational codes. Indeed, many SID analyses that have been applied to broadly defined occupational units do not make any specific exclusions for diagonals/pseudo-diagonals. This might encourage us to ignore occupational detail and diagonal/pseudo-diagonal structures, but our own experience is that doing so risks introducing errors of inference, because the first dimensional solution will be unintentionally 'polluted' by these separate sociological mechanisms.

Figure 6.4 gives an illustration of diagonals and pseudo-diagonals in complex occupational data. It shows the cross-tabulation of occupations for husband-wife combinations (upper panel) and friendship combinations (lower panel) for data from the UK in the 1990s using occupational units at the 'microclass' level (75 different microclass categories). In the figure, the size of markers is proportional to the relative over-representation of the particular occupational combination within the data (i.e. a large square indicates that the combination of microclasses occurred more often than would be predicted if social connections were distributed randomly, and a smaller square indicates that the combination occurred relatively less often).[11] We see that generally speaking, though not always, 'diagonals' (the medium-shaded squares) occur more often than would be expected. In a few instances, we can note that the diagonal square is quite small (i.e. the combination did not occur very often)—this typically represents the influence of gender segregation in the husband-wife panel (e.g. if a husband is in a male-dominated job, it is not especially likely that his wife will be in the same job).

The off-diagonal squares that are labelled as pseudo-diagonals in Figure 6.4 (darkest squares) were selected iteratively, during the analysis process, by inspecting the data and preliminary results and identifying those combinations of occupations where there was an apparent shared workplace (examples include, for husbands and wives, the link between 'farmers' and 'farm labourers', and between 'proprietors' and 'shop assistants'; examples for friendship include 'military occupations' linked to 'security occupations', and combinations of 'tailors' and 'textile workers'). We can observe from the figure that the combinations that we identified as possible pseudo-diagonals do indeed often occur a little more frequently than might otherwise be predicted, yet this is not always the case, and there are other instances of unusually common off-diagonal combinations of occupations that we did not label as pseudo-diagonal.

One way to try to identify pseudo-diagonals systematically could be to look at patterns of correlation between over-representation in social interactions and other characteristics of jobs. Using data on marriage records from the USA, Fig. 6.5 shows the bivariate correlations between statistical over-representation of occupational combinations (as defined by two thresholds) and selected aggregate measures (of the occupations'

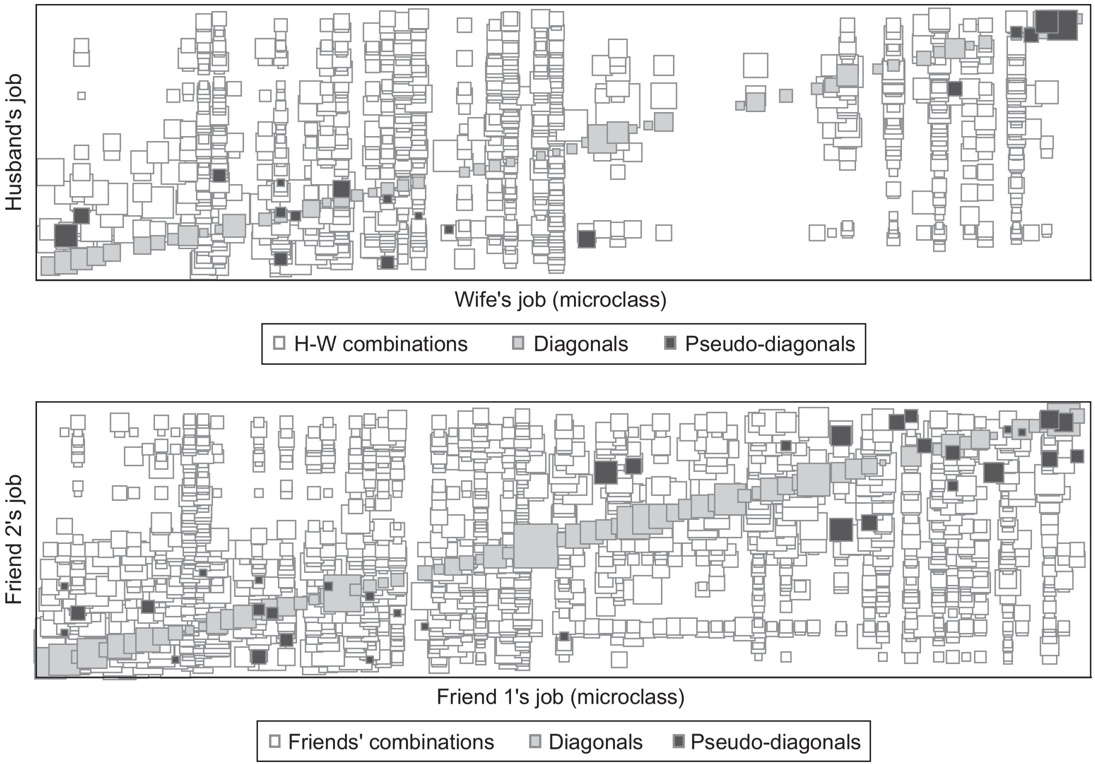



**Fig. 6.4** Indication of 'diagonals' and 'pseudo-diagonals' within data for Britain using the 'microclass' occupational classification. Notes: *N* ~= 100k pairs of occupations, obtained from records from the UK BHPS 1991–2008. Size of squares indicates relative over-representation of the combination. Squares are only shown if the combination occurred at least three times (upper panel) or seven times (lower panel) within the dataset

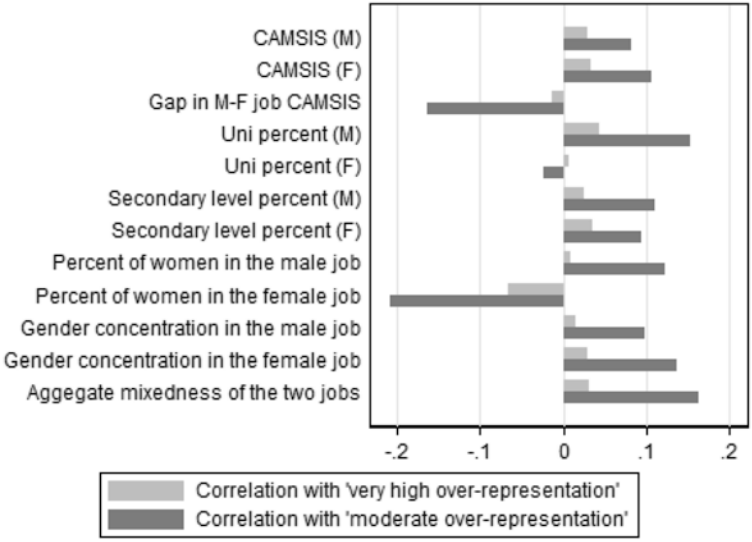



**Fig. 6.5** Correlation between aggregate-level measures and extent to which job combinations are statistically over-represented. Source: Over-representation defined as combinations that occur either ten times or three times more often than would be predicted if social connections were distributed randomly

position in the stratification structure using CAMSIS, their relative concentration of men and women, and their average educational levels).[12] Interestingly, there are only ever low correlations with any of these variables. The strongest associations, though still quite weak, are with the difference between CAMSIS scores for men and women (over-represented jobs are disproportionately likely to have smaller gaps in their CAMSIS scores) and with measures of gender concentration (over-represented combinations are less likely to feature a female job which is female dominated, but are more likely to feature a male job which is female dominated and are also more likely to occur when either or both of the jobs involved are gender balanced). However, a general message from this analysis is that broad-brush association patterns do not clearly identify pseudo-diagonal patterns.

In the social interaction distance (SID) tradition of analysis, pseudo-diagonals often influence scores in the primary and 'subsidiary dimensions' to the structure of social connections. 'Subsidiary dimensions' refer to dimensional patterns that are not interpreted as the main dimension of social stratification and its reproduction, but that seem to reflect other patterns in social connections between occupations (see also Sect. 11.2). In SID studies, specific 'pseudo-diagonal' combinations may themselves be the main drivers of a subsidiary dimensional structure. For example, pseudo-diagonals may be apparent as occupations identified at extreme positions, and/or with the greatest 'inertia', in subsidiary dimensions of the social distance space. Because of this, the dimensional analysis may sometimes serve to reveal particular pseudo-diagonal combinations that might not otherwise have been obvious. More prosaically, it might also reveal unintended problems of coding and classification.[13] When it comes to the primary dimension of the SID solution, it is common practice to exclude diagonal and pseudo-diagonal cases from the analysis, reasoning that the social drivers of these combinations are substantively different to those that shape the key structure of social stratification reproduction. However, if an analyst does not agree that a particular combination ought to be labelled as a 'pseudo-diagonal' (or if they simply don't notice the relevant combination in their data), then the occurrences in that combination will necessarily make some contribution to the primary dimensional structure.

Table 6.4 shows examples of each of these processes. On the left hand, we list the occupations at the extremes of the first four subsidiary dimensions in the SID solution, and, if relevant, other occupations that those records have a disproportionately common linkage to. It is apparent that a number of these occupations could have shared workplaces—these combinations could be interpreted as pseudo-diagonals and treated accordingly in further analysis. On the right hand, we list standardised dimension scores for the same occupations, and selected other occupations, on the male scale, for the first dimension of two different SID solutions, the first excluding pseudo-diagonals and the second including them. There are clear differences in interpretation that might arise.

Our own prescription is that during a SID analysis, we should thoroughly explore possible pseudo-diagonals within a dataset and exclude any that are identified, as this should make the first dimension a more coherent representation of social stratification inequality. However, we doubt that there can be any absolute guide to identifying and excluding

**Table 6.4** Examples of occupations at the extremes of subsidiary dimensions

| Influential occupations at extremes of the dimension | | | Score in first dimension, after excluding pseudo-diagonals (as CAMSIS) | | Score in first dimension, not excluding pseudo-diagonals | |
|---|---|---|---|---|---|---|
| | Male | Female | m | f | m | f |
| Dimension 2 | 20 Farm managers | 20 Farm managers | 51 | 45 | 45 | 37 |
| | 21 Farmers | 21 Farmers | 49 | 47 | 42 | 35 |
| | | 40 Postmasters | | 41 | | 40 |
| | 434 Animal trainers | | 51 | | 51 | |
| | 600 Farm supervisors | 600 Farm supervisors | 41 | 39 | 33 | 33 |
| | 601–605 Farm workers | 601–605 Farm workers | 28 | 26 | 19 | 12 |
| Dimension 3 | 165 Medical scientists | 165 Medical scientists | 89 | 95 | 97 | 99 |
| | 176 Physical scientists | 176 Physical scientists | 87 | 89 | 89 | 93 |
| | 182 Psychologists | 182 Psychologists | 85 | 91 | 90 | 93 |
| | | 301 Dentists | | 82 | | 91 |
| | 306 Physicians | 306 Physicians | 93 | 88 | 81 | 99 |
| Dimension 4 | 34 Lodging managers | 34 Lodging managers | 52 | 50 | 51 | 50 |
| | 430 Gaming supervisors | | 49 | | 48 | |
| | 440 Gaming workers | 440 Gaming workers | 39 | 38 | 36 | 35 |
| | | 443 Baggage porters | | 45 | | 44 |
| | 530 Hotel clerks | 530 Hotel clerks | 42 | 37 | 42 | 37 |
| *Other selected illustrative occupations* | | | | | | |
| | 62 Human resources | " | 60 | 63 | 61 | 60 |
| | 172 Chemists | " | 71 | 75 | 71 | 76 |
| | 232 Secondary teachers | " | 67 | 73 | 75 | 73 |
| | 313 Registered nurses | " | 60 | 63 | 65 | 62 |
| | 460 Childcare workers | " | 50 | 42 | 46 | 43 |
| | 651 Roofers | " | 27 | 31 | 30 | 25 |
| | 720 Auto technicians | " | 37 | 34 | 39 | 32 |
| | 814 Welding workers | " | 32 | 24 | 34 | 26 |
| | 913 Drivers | " | 36 | 36 | 37 | 31 |

Notes: Analysis of husband-wife combinations from 2000 census (accessed via IPUMS-I), *N* = 2191k for whole sample, and 2053k without pseudo-diagonals pseudo-diagonals, and there is certainly no guarantee that different analysts working on the same data will identify the same combinations (at the CAMSIS website, we typically release documentary metadata that lists any pseudo-diagonals that were excluded in a derivation, and we would encourage others using the SID approach to do the same). Because of the role of pseudo-diagonals, therefore, the SID approach necessarily has a subjective character to it: the dimensions of social structure are to some degree influenced by the way in which the analyst identifies and treats diagonal and pseudo-diagonal combinations.

### 6.7.2 Interpretations of 'Pseudo-Diagonals'

Over and above being a 'nuisance' when constructing SID scales with occupations, the mechanisms that lead particular occupational combinations to occur disproportionately often can sometimes provide interesting substantive insights. The first and most widely cited mechanism behind pseudo-diagonals is that of 'situs', or shared industrial environment or workplace contingency (e.g. Chan and Goldthorpe 2004; More and Suchner 1976; Morris and Murphy 1959). It may seem self-evident that people who share the same occupations are more likely to form new social connections (or are better equipped to assist their friends or family to gain access to the same occupation). Through either mechanism, higher volumes of social connections between people in exactly the same occupation ('diagonals') are widely observed. However, 'situs' mechanisms often operate across different occupations that share a physical environment. Well-known examples in terms of marriage and cohabitation include the disproportionate volumes of partnerships between doctors and nurses, dentists and dental technicians, airline pilots and air stewardesses, teachers and educational administrators, and farmers and farm labourers. In each case, the connection occurs more frequently than we might otherwise expect, and the most plausible explanation is that the shared workplace environment of the jobs either increased the chances of new social connections being formed, and/or increased the chances of recruitment to the job via existing social connections. Similar effects of situs are observed for data on friendship patterns rather than cohabitation—for example, links between policemen and security officers—and in intergenerational inheritance, where 'affinities' between certain occupations might reflect shared business ventures or environments—for instance, a father who is a restaurateur and his daughter who is a waitress.[14] For completeness, there are also certain patterns of social relationships between occupations that occur much less frequently than we might otherwise anticipate for reasons that are likely to reflect the absence of situs. For instance, the jobs of 'bouncers' and 'fishermen' share many characteristics and occupy similar positions in the stratification structure, but are physically located in very different environments, so there are unusually few social contacts between them.

Although the driving force behind 'situs' relationships might not particularly reflect the influence of social stratification, these patterns are relevant to our wider understanding of the nature of social organisation and social structure. They may reveal, for instance, important 'bridging' occupations or sectors that serve to connect social groups from unusually different positions (the healthcare sector is a good example, since it tends to bring people from diverse social positions into regular proximity). Connections driven by 'situs' might also reveal channels of social mobility that facilitate unusual social connections (or help consequential social divisions to endure). The link from airline pilots to air stewardesses might be seen as an avenue of social mobility, since the former are dramatically better remunerated than the latter; the absence of a link between bouncers and fishermen, on the other hand, might perpetuate ethnic segregation, or divisions between urban and rural communities.

'Situs', however, is not the only mechanism that can push together people from different occupations (or keep them apart). Perhaps the next most important factor is the gender typing of jobs. Between a half and three quarters of all occupational positions in most contemporary societies are held either by men in 'male-dominated' jobs, or by women in 'female-dominated' jobs (e.g. Jarman et al. 2012; Hakim 1998; Sokoloff 1992; Rytina and Bianchi 1984). In the SID tradition, gender segregation in occupations often emerges as a statistically important dimension in social interaction patterns (e.g. Chan and Goldthorpe 2004), but it can also account for highly specific patterns of under- or over-representation in the social connections between occupations that might best be represented by diagonal or pseudo-diagonal terms. If we consider, for example, a hospital that employs 30 male hospital porters and 50 female cleaners, we see a situation of employees being found in gender-segregated jobs, but working within wider institutions of a mixed-gender character that are likely to result in certain gender-specific patterns of social interaction. Importantly, much gender segregation in occupations occurs at a very fine level of occupational activity and can be masked by more aggregate-level occupational codes (e.g. Crompton and Sanderson 1990).

A third mechanism that can influence the volume of social connections between occupations concerns the shared educational requirements of occupations. Although educational requirements themselves vary substantially by birth cohorts, it follows naturally that different occupations, which require very similar educational backgrounds, may exhibit unusually frequent volumes of social contacts between their incumbents (perhaps, for instance, because they are old college friends). Accordingly certain patterns of occupational combinations may occur disproportionately often on the grounds of recognisably similar educational requirements (for instance, in social connections between medical doctors and veterinarians).

The geographical distribution of different occupations can be another relevant social force. Occupations, particularly those linked to manufacturing and production, are often concentrated in certain regions within a society. The implications for social connections patterns may be obvious: occupations (for instance, mining) that have very uneven regional distributions will tend to have higher-than-average levels of social connections to other occupations found in the same regions, but lower-than-average levels of social connections to others. In many countries, a similar issue applies to the concentration of immigrant and ethnic minority groups within certain areas, occupations and industries. It is very common for immigrant cohorts and their descendants to concentrate in selected localities, selected occupations, and even within specialised economic 'enclaves' (e.g. Guinea-Martin et al. 2010; Waldinger and Lichter 2003; Foner 2000). In some circumstances, this occurs to the extent that the occupation itself, within a country, comes to be dominated by incumbents from a certain ethnic, immigrant or religious group. In these situations, patterns of homophily by ethnicity, immigrant status or religion lead in turn to uneven patterns of social connections between the occupations concerned.

Some generic social mechanisms can even lead to patterns in social contacts and occupations despite not being linked directly either to social stratification or to occupations (cf. Kalmijn 1998). Examples might include the way religions, or religious institutions, encourage certain social contacts but discourage others (e.g. Shaw 2000), or the way in which family migration shapes the frequency of contact with kin and non-kin alike (e.g. Bott 1957). Whilst these generic mechanisms are not intrinsically linked to occupations, they can sometimes align with occupations in a way that generates distinctive social patterns in turn. As one example, Kalmijn (1998) discusses how 'third parties' can have a substantial influence upon the formation of social contacts that lead to cohabitation. Third parties themselves are frequently friends or relatives of one of the partners, so a logical consequence is that the propensity of friends or relatives to hold similar occupations to individuals serves in turn to increase the chances of particular connections being made between certain occupations. For example, one reason that a doctor might form a social relationship with a nurse may not be because they work at the same institution, but because the friends of doctors (who might broker a relationship) are themselves more likely to work at the same institutions as nurses.

## 6.8 Taking Account of Gender

Another important practical consideration in undertaking a SID analysis of social connections between occupations is whether gender differences in occupational distributions are deliberately incorporated into the analysis. If the social interaction data is already gendered, such as if the ego-alter pairs are defined as the male and female records for data on cohabiting couples, then most analytical techniques will by default generate separate scale scores for the occupations of men and of women. The example code within Fig. 6.1 would generate separate scores for males and females, because this data is organised as male-female social interactions (in the terminology of correspondence analysis, the male cases are the rows and the female cases the columns). Preparatory adjustments to the data such as merging sparse categories are also ordinarily undertaken separately for the male and female occupations. This is illustrated, for example, in segment (ii) of Fig. 6.1, when the 'hocc3' and 'wocc3' variables (for husbands and wives) are recoded with different criteria, meaning that the subsequent SID analysis takes place on a slightly different set of occupational units for the men and for women.

Alternatively, for gendered social connection data, a constraint can be set in the analysis to enforce equivalence of scores for the same occupation (the CAMSIS project website features some illustrative examples of ways of setting this constraint).[15] Moreover, if the social interaction data is not gendered, such as if it refers only to male records, or if it refers to records that do not imply a gender (e.g. 'friend 1' linked to 'friend 2'), then the scale scores will not directly differentiate between male and female positions (although it is likely that, in practice, the analytical results will still be influenced by the role of occupational gender segregation—for examples, see Chan 2010; Bakker 1993).

We commented on the relative attractions and drawbacks of generating separate ('specific') CAMSIS scales for men and women in Sect. 5.1.5. Researchers in the CAMSIS tradition have usually favoured separate scales when feasible (e.g. Prandy 1986); however, in some contexts CAMSIS scales have not separated male and female scores, for instance, for pragmatic reasons of lack of coverage or variation in female occupational positions (e.g. Lambert et al. 2013 for historical datasets). Scientifically, either approach is plausible, and arguably the key issue is that decisions taken in generating a relevant CAMSIS version should be adequately documented, whatever the operational choices with regard to gender.

## 6.9 Statistical Analysis of Social Interaction Data

In broad terms, social interaction distance analysis centres on summarising a cross-tabulation of the occupations of related individuals. The cross-tabulation itself is long recognised as a foundational element of social science data analysis (cf. Treiman 2009; Gilbert 1993; Clogg 1982). Our key concern is to find summary metric scores that capture patterns of differences between the different categories (commonly referred to as 'row scores' and 'column scores'). There are a number of related statistical techniques that can be used to derive category scores that serve as a quantitative representation of an underlying structure within the cross-tabulation. 'Correspondence analysis' (e.g. Greenacre and Blasius 1994) is the technique most commonly used within the CAMSIS tradition, but 'log-linear association modelling' (Wong 2010; Goodman 1979) and 'multidimensional scaling' work in almost exactly the same way as correspondence analysis in the case of a two-way table of relationships, and indeed all three have been used to derive CAMSIS scale scores and other similar social interaction distance scales (cf. Stewart et al. 1980, who use multidimensional scaling; Prandy and Lambert 2003, who use association modelling; and Lambert et al. 2013, who use correspondence analysis).[16] The CAMSIS project website features downloadable examples of code for undertaking correspondence analysis (using the software Stata and SPSS) and for using log-linear association models (using the software lEM and R). Our example of Stata code in Fig. 6.1 (Sect. 6.4) uses correspondence analysis.

Table 6.1 illustrated intuitively the way in which the row and column scores are generated as the product of a statistical model, and Table 6.5 expands upon that example. Previously, we noted that the row and column scores (in Table 6.1) were designed to give a model that better predicts the 'observed count' (the number in each cell of Table 6.1). Table 6.5 indicates how the relevant statistical models lead to model-based predicted counts for each cell. The middle row within each cell indicates the predicted values from a first model ('model 1') that uses only the row and column total frequencies (often called 'marginals') to guess the cell counts. The second row in the cell shows the predicted counts from one version of the SID model, where row and column scores are derived and used to help predict the number of cases, but no other adjustments are made (model 2). Finally, the last row in the cells shows the predicted count with a common extension to the basic SID model, whereby some extra parameters are added to the model to provide exact predictions for nominated 'diagonal' and 'pseudo-diagonal' combinations. The point here is that 'model 3', generating the values shown in the third row, is a closer fit to the actual data, at the cost of only a few extra parameters—typically, we say it is a better, more parsimonious model. This is formally demonstrated by its model fit statistics—model 3 has a smaller log-likelihood, indicating less model error than any other option, and it has a smaller 'Bayesian Information Criteria' (BIC) statistic, indicating that it could be considered more parsimonious than the other models. In summary, the row and column scores for the occupations that are generated by model 3 help us to better predict the actual distribution of husband-wife social interactions.

The key statistical contribution of a SID analysis, therefore, is the derivation of row and column scores for categories based upon social interaction patterns. Seminal applications of this style of analysis to social interactions between occupations date to the 1960s (e.g. Laumann and Guttman 1966; Laumann 1966), and there is a long tradition of deriving dimension scores in similar ways in other social science domains (e.g. Guttman 1944 on attitudinal scales). It is common to use graphical representations to summarise the category scores in a two-dimensional framework, such as with Cartesian plots that portray the results from a correspondence analysis. However, the quantitative structures of interest may be primarily one-dimensional (e.g. Stewart et al. 1980 regarding occupations; Prandy 1979 regarding immigrant groups); or they may reasonably involve more than two dimensions (e.g. Bennett et al. 2009 regarding dimensions of cultural consumption patterns).

Model 3 from Table 6.5 identified selected diagonals and pseudo-diagonals. This model fits specific parameters for each individual cell (it does this for a list of 100 row-column combinations; the list is not shown in Table 6.5 but would be supplied with the model specification code—see examples on the CAMSIS website). An almost-equivalent result can be achieved simply by dropping all cases from the relevant cells from subsequent analysis—this is probably the more common approach within the CAMSIS tradition and is illustrated in Fig. 6.1 at segment (iv) (the relevant code in Stata is 'if psd1==0'). In Table 6.5, the impact is evident in the exact prediction of cells that are on the diagonal line within the cross-tabulation (for model 3)—in this case, the model has fit an extra parameter to exactly predict each diagonal entry.

**Table 6.5** Model-based estimated values with different association models for the social interaction distance structure for husband-wife occupational connections (USA 2005, SOC 2-digit)

Cells indicate:
observed values
*expected values under Model 1*
*expected values under Model 2*
*expected values under Model 3*

| WH | 0 | 1 | 2 | 3 | … | 96 | 97 | 98 | Total |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 880 | 481 | 257 | 189 | … | 40 | 5 | 2 | 15,393 |
| | *211* | *270* | *195* | *192* | | *151* | *4* | *9* | |
| | *300* | *341* | *271* | *206* | | *38* | *1* | *10* | |
| | *880* | *403* | *195* | *142* | | *43* | *1* | *7* | |
| 1 | 229 | 461 | 163 | 118 | … | 36 | 0 | 4 | 10,963 |
| | *151* | *192* | *139* | *137* | | *108* | *3* | *7* | |
| | *203* | *238* | *185* | *147* | | *33* | *1* | *7* | |
| | *220* | *461* | *142* | *112* | | *35* | *1* | *5* | |
| … | … | … | … | … | … | … | … | … | … |
| … | … | … | … | … | … | … | … | … | … |
| 97 | 2 | 4 | 5 | 3 | … | 11 | 13 | 0 | 678 |
| | *9* | *12* | *9* | *8* | | *7* | *0* | *0* | |
| | *4* | *8* | *4* | *7* | | *15* | *0* | *0* | |
| | *4* | *7* | *4* | *8* | | *14* | *13* | *0* | |
| 98 | 14 | 40 | 23 | 20 | … | 4 | 0 | 96 | 2415 |
| | *33* | *42* | *30* | *30* | | *24* | *1* | *1* | |
| | *33* | *44* | *31* | *33* | | *17* | *0* | *2* | |
| | *27* | *39* | *25* | *32* | | *17* | *0* | *96* | |
| Total | 6500 | 8309 | 6005 | 5911 | … | 4661 | 118 | 281 | 473,210 |

| | Description (*IEM specification*) | Log-like | BIC | Degrees of freedom |
|---|---|---|---|---|
| Model (1) | Marginals only *mod {H, W}* | −3,738,397 | 7,479,356 | 9604 |
| Model (2) | (1) + first-dimensional row-column scores *mod {H, W, ass2(H,W,5e)}* | −3,699,926 | 7,404,960 | 9409 |
| Model (3) | (2) + second-dimensional row-column scores + 100 diagonals and pseudo-diagonals *mod {H, W, ass2(H,W,5e), ass2(H,W,5e), fac(HW,100)}* | −3,674,656 | 7,358,250 | 9116 |

Notes: Based upon data from USA, IPUMS-I 2005, SOC two-digit amalgamated groups (not an officially used classification)

A special feature of the analysis of cross-tabulations in the SID tradition using occupations is that in many cases the analysis involves a large and sparse cross-tabulation (e.g. a table of perhaps 300 rows and 300 columns, which can be expected to include many cells that contain no cases at all). This contrasts sharply with more common conventions in the analysis of cross-tabulations, which involve quite low numbers of rows and columns. In terms of deriving suitable row and column scores, the size of the table is not necessarily problematic, so long as each individual row and column (e.g. husband's occupation; wife's occupation) is represented by a reasonable number of cases. For example, most CAMSIS scales have typically worked to the requirement that any row or column category ought to be represented by around 30 cases as a minimum (see Sect. 6.6), noting that these should be the number of cases after any relevant diagonal or pseudo-diagonal combinations have been removed (see Sect. 6.7). Although it is sometimes anticipated that sparse tables with many empty cells will not sustain reliable results, this does not follow intrinsically from the statistical methods, and sensitivity analysis suggests that this is not the case with regard to inferences for the row and column scores (e.g. Wong 2010, pp. 30–31). In practical terms, however, the large number of rows and columns might cause problems for statistical software. For instance, calculations may become quite slow—it is not unusual for a model using a very large matrix to take several hours to converge using a standard computing facility. Moreover, a few software routines feature upper limits on the number of categories that they can accommodate, although these limits are imposed for reasons of computer memory, rather than statistical theory. Nevertheless, in our own research, we have not usually encountered sustained estimation problems when working with large, sparse cross-tabulations when using correspondence analysis in Stata and SPSS, nor using association models in lEM and R.

When estimating row and column scores for occupational categories, it is useful to conceptualise the scores as sample-based estimates, implying that standard error statistics might be associated with them. For example, an occupational unit that is represented by large numbers of cases might have small standard errors around its scale score, but a score for a unit represented by only a few cases could be expected to have large standard errors.[17] A variety of methods to estimate standard errors for row and column scores have been proposed (e.g. Wong 2010, pp. 27–30). However, at the time of writing, standard errors are not routinely calculated by many of the popular statistical software—to our knowledge, the only example from the traditions that we have described previously is the row-column association model that is incorporated within Turner and Firth's (2007) gnm library in R. Alternatively, various plausible approximations for margins of error for scores derived through correspondence analysis can be readily calculated. For instance, on the CAMSIS project website, illustrative code is given showing how approximate standard errors can be constructed by calculating, for any given ego occupation, the standard error of the mean for the scale scores for the corresponding alters of the egos which represent that occupation. Figure 6.6 illustrates an example of derived row scores for occupations in Britain using the SOC2010 scheme, with margins of error based upon this calculation.[18]

Whilst it is feasible to produce such uncertainty estimates, it is not so straightforward to know what best to do with this information. The standard errors could be used to evaluate descriptively whether the SID score estimate for one occupational category is significantly different from that of another occupation (or that of the same occupation in another society). However, many users of CAMSIS scale scores are mainly interested in the point estimate for the score, for instance, as an instrument for use in further analysis. In principle, the uncertainty about the score might still be represented in such uses—for instance, we could take multiple simulations of plausible values within the range defined by the margin of error. In practice, however, standard errors for SID scale scores are not yet widely exploited.

## 6.10 Post-Processing CAMSIS Scale Scores

The last segment of Fig. 6.1 shows some 'post-processing' taking place upon the scale scores that are derived from the SID analysis. This includes the common standardisation within the CAMSIS project, namely, setting a mean of 50, and standard deviation of 15, for a nationally representative population. This is implemented by matching the model-based

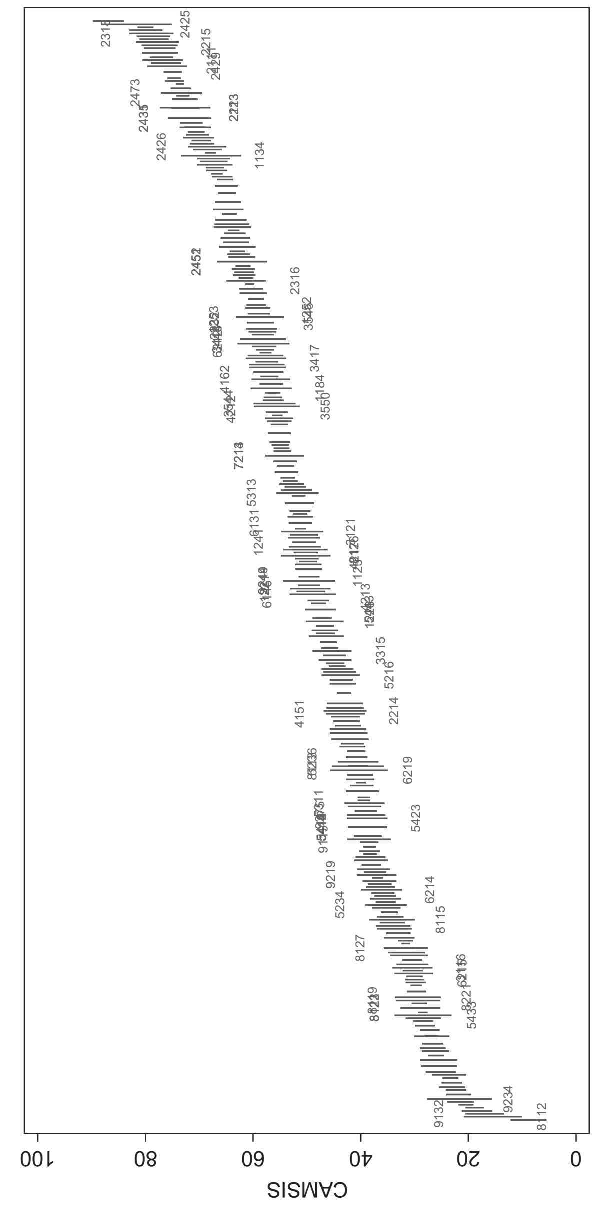



**Fig. 6.6** Standard errors for a CAMSIS scale score (UK SOC2010 categories, male scale scores). Source: Based on CAMSIS scores for SOC2010 derived from analysis of Labour Force Survey datasets 2010–2012 (see CAMSIS project website). SOC2010 unit code is listed if SE exceeds (mean + SD) of all SE's

scores back to the underlying population data and rescaling accordingly. Another important part of the post-processing activity is the construction of an 'index file' that lists occupations with their corresponding scale scores. Typically, there are two scale scores, one for the scale for men and one for the scale for women. In addition, there may be other information to convey about the scale construction, such as margins of error estimates for the scale scores, data on the number of cases used, information on which occupations were merged together in recoding exercises, or data on cell combinations that were treated as pseudo-diagonals. The post-processing illustrated in Fig. 6.1 attaches scale scores and standard error approximations for them to the data by using a brief routine that involves match-merging scores against the original datasets. Elaboration on relevant techniques of post-processing is provided on the CAMSIS project website.

## 6.11 Automated CAMSIS Scale Derivation Tools

Whist the stages of activity outlined above can be quite labour intensive, the CAMSIS project webpages also include some software tools that support the derivation of CAMSIS scales in a manner that is substantially automated. For this purpose, we have written Stata format programming 'macros' (on the CAMSIS website, under 'make_CAMSIS') which can be implemented by declaring a number of 'arguments' concerned with the data—for instance, to indicate the names and locations of data files and variables, thresholds to be used in recoding occupational categories, and information specifying which if any occupational combinations are to be excluded from the analysis as diagonals or pseudo-diagonals.

However, we would attach two 'health warnings' to prospective use of this automatic scale derivation programme. First, the automated programme in its current form is rather complicated to implement! The prospective user will still have some considerable work in hand to correctly specify each argument and ensure that data is placed in suitable formats and locations. Second, the automated programme makes some rather strong assumptions about the SID analysis solution and the underlying occupational data. It is important that a user understands broadly what the programme is doing, as there is a chance that it could generate rather inappropriate results which might only be recognised with some expertise in the topic area (e.g. it might merge together many more occupations than the user thinks is substantively reasonable; in extreme situations, it might pick out a dimension from the correspondence analysis solution that is not in fact the most compelling candidate as the main dimension of the SID structure). In summary, the automated programme for deriving CAMSIS scales is an interesting tool that we would encourage analysts to consider (including perhaps by using it as a baseline comparison, before undertaking a full derivation in a more manual way). However, the automated derivation macro is also a tool that should be used with caution.

## Notes

1. Copies of some of the supplementary files used in the analysis can be downloaded from the CAMSIS project website; see www.camsis.stir.ac.uk.
2. One treatment for clustered cases could be to downweight the related combinations, so that their total influence is reduced to a single unit (or randomly select one case from within the cluster). However, these treatments are probably unduly conservative. Model parameters could potentially be added to reflect the link between connections, such as random effects for the clustering, or fixed effects for characteristics of the clusters, but this would require a non-standard model specification.
3. Aside from employment status, disaggregation might also be made, for example, by education, age, or ethnicity. Most previous SID studies have not disaggregated by other characteristics, with the important exception of gender (see Sect. 6.8); we discuss disaggregation by education measures in Chap. 10.
4. The value is usually calculated as the weighted average of the scores for the different employment status categories.
5. Our illustrative code in Fig. 6.1 uses the original occupation codes—ISCO-08 three-digit units—as supplied on the data without further discussion. We might have explored alternative aggregations of these, such as two-digit or one-digit ISCO units.

6. Although it is not apparent from the graph alone, further inspection reveals that industrial sector and 'situs' are the most important subsidiary forces in the two-digit solution, but gender segregation and associations between advantaged professions that are perhaps related to shared educational requirements are relatively more important in the four-digit solution.
7. The problems expand if more occupational units are used. An occupational taxonomy with, say, 450 occupational units, and 4 possible employment status categories of interest, yields an eye-watering 1800 plausible occupational units and 3.2 million possible husband-wife permutations.
8. The CAMSIS website includes some relevant examples of code. The broad approach is to define one or more 'fall-back' occupational codes for existing units. Then, an algorithm is written to evaluate the number of cases representing an occupational unit, and, if it is below an agreed minimum, the unit is recoded to the nominated 'fall-back', and the algorithm repeated.
9. This terminology comes from the analysis to tabular data. When the ego and alter occupations are cross-classified, 'diagonal' combinations are those occupying the line of equality in the table (when ego and alter have the same occupation).
10. Analysts often leave all occupations in the solution, including diagonals and pseudo-diagonals. It is also common for them to exclude all occupational combinations that are labelled as diagonals or pseudo-diagonals. Exclusion can be achieved both by cutting the cases from the dataset and by fitting explicit model parameters for the relevant combinations (which leads to the equivalent results); the latter approach is commonly employed for 'diagonals' in the social mobility research tradition (e.g. Luijkx 1994). However, for a large cross-tabulation of data, the specification of model parameters can be cumbersome (it may require the construction of a large 'design matrix'). In addition, common implementations of correspondence analysis in statistical software do not feature tools to readily specify such parameters. A further possibility, implemented for historical data by Lambert et al. (2013), is to retain the pseudo-diagonal combinations in analysis, but substantially downweight them, according to some other criteria.
11. We do not show labels for the individual occupations involved for reasons of succinctness, but the categories correspond to the microclass categories reported in Griffiths and Lambert (2012).

12. The correlations in this example are the square root of the regression $r^2$ statistic, for a regression predicting the measure as a function of whether or not the occupation is over-represented; the underlying population excludes diagonal instances where the male and female job in the combination is the same.
13. For instance, if there are two different occupational units that can both be plausibly used as codes for the same jobs, an artefactual peak in social interactions between them is likely to occur that might reasonably be modelled as a pseudo-diagonal. Academic sociologists, for instance, may sometimes code themselves as 'Social science researchers' and sometimes as 'Higher education teachers', inducing an artefactual peak in social connections between these two categories.
14. In social mobility studies, structural schemes have been used to identify such forces and partial them out from the core analysis (especially Yamaguchi 1983; Breen 2004).
15. In some software settings, the simplest way to enforce this constraint is to make a duplicate of the data, reversing the gender order, then pool it with the original data, and undertake analysis on the pooled data. Indicatively, if the original analysis was on [M → F] (indicating male records linked to female records), the revised format would involve an analysis on [{M,F} → {F,M}] * 0.5 (indicating male-female and female-male links jointly analysed and downweighted appropriately).
16. Another relevant technique is 'latent class analysis'. This technique would in broad terms seek to identify and define aggregated categories or clusters within the table, after which those could then be potentially ordered (e.g. Goodman 2002; Croon 2002). In this framework, we would seek to group occupations into categories defined by patterns in the structure of social connections between occupations, rather than seeking to identify underlying quantitative dimensions.
17. In many cases, a given occupation may have two scale score estimates, for instance the score on the male scale and the score on the female scale. In such examples, it may well be the case that the occupation is represented by different numbers of units on the two scales (e.g. by many men, but few women). Accordingly, the same occupation might have different standard errors (as well as scale score estimates) in the different contexts.
18. Given that the underlying populations of social connections from which the data for a SID analysis is constructed often have unusual features (for instance, being restricted to both-working couples; excluding a range of

'pseudo-diagonal' combinations), it is plausible that conventional approaches to calculating margins of error are inefficient, and 'robust' estimates should be used that widen the standard errors.

## Bibliography

Bakker, B. F. M. (1993). A New Measure of Social Status for Men and Women: The Social Distance Scale. *Netherlands Journal of Social Sciences, 29*, 113–129.

Bennett, T., Savage, M., Silva, E. B., Warde, A., Gayo-Cal, M., & Wright, D. (2009). *Culture, Class, Distinction*. London: Routledge.

Bott, E. (1957). *Family and Social Network*. London: Tavistock Publications.

Breen, R. (Ed.). (2004). *Social Mobility in Europe*. Oxford: Oxford University Press.

Chan, T. W. (Ed.). (2010). *Social Status and Cultural Consumption*. Cambridge: Cambridge University Press.

Chan, T. W., & Goldthorpe, J. H. (2004). Is There a Status Order in Contemporary British Society? *European Sociological Review, 20*(5), 383–401.

Charles, M., & Grusky, D. B. (2004). *Occupational Ghettos: The Worldwide Segregation of Women and Men*. Stanford, CA: Stanford University Press.

Clogg, C. C. (1982). Using Association Models in Sociological Research: Some Examples. *American Journal of Sociology, 88*(1), 114–134.

Crompton, R., & Sanderson, K. (1990). *Gendered Jobs and Social Change*. London: Unwin Hyman.

Croon, M. A. (2002). Ordering the Classes. In J. A. Hagenaars & A. McCutcheon (Eds.), *Applied Latent Class Analysis* (pp. 137–162). Cambridge: Cambridge University Press.

de Luca, D., Meraviglia, C., & Ganzeboom, H. B. G. (2010, April 14). *ICAM: The International CAMSIS Scale*. Hafia: First Presented at ISA RC28, Hafia. Software Retrieved from http://www.harryganzeboom.nl.

Elias, P. (2000). Status in Employment: A World Survey of Practices and Problems. *Bulletin of Labour Statistics, 1*, 11–19.

Foner, N. (2000). *From Ellis Island to JFK: New York's Two Great Waves of Immigration*. New Haven: Yale University Press.

Gilbert, G. N. (1993). *Analyzing Tabular Data: Loglinear and Logistic Models for Social Researchers*. London: UCL Press.

Goodman, L. A. (1979). Simple Models for the Analysis of Association in Cross-Classifications Having Ordered Categories. *Journal of the American Statistical Association, 74*, 537–552.

Goodman, L. A. (2002). Latent Class Analysis: The Empirical Study of Latent Types, Latent Variables, and Latent Structures. In J. A. Hagenaars & A. McCutcheon (Eds.), *Applied Latent Class Analysis* (pp. 3–55). Cambridge: Cambridge University Press.

Greenacre, M., & Blasius, J. (1994). *Correspondence Analysis in the Social Sciences: Recent Developments and Applications*. London: Academic Press.

Griffiths, D., & Lambert, P. S. (2012). Dimensions and Boundaries: Comparative Analysis of Occupational Structures Using Social Network and Social Interaction Distance Analysis. *Sociological Research Online, 17*(2), 5.

Guinea-Martin, D., Blackwell, L., & Elliott, J. (2010). Occupational Segregation and Concentration: An Analysis by Sex, Employment Status and Ethnic Group in England and Wales. In J. Stillwell, P. Norman, C. Thomas, & P. Surridge (Eds.), *Spatial and Social Disparities: Understanding Population Trends and Processes* (Vol. 2, pp. 99–111). Dordrecht: Springer.

Guttman, L. (1944). A Basis for Scaling Quantitative Data. *American Sociological Reivew, 9*(2), 139–150.

Hakim, C. (1998). *Social Change and Innovation in the Labour Market: Evidence from the Census SARs on Occupational Segregation and Labour Mobility, Part-Time Work and Student Jobs, Homework and Self-Employment*. Oxford: Oxford University Press.

Jarman, J., Blackburn, R. M., & Racko, G. (2012). The Dimensions of Occupational Gender Segregation in Industrial Countries. *Sociology, 46*(6), 1003–1019.

Kalmijn, M. (1998). Intermarriage and Homogamy: Causes, Patterns, Trends. *Annual Review of Sociology, 24*, 395–421.

Lambert, P. S., Zijdeman, R. L., Maas, I., van Leeuwen, M. H. D., & Prandy, K. (2013). The Construction of HISCAM: A Stratification Scale Based on Social Interactions for Historical Research. *Historical Methods, 46*(2), 77–89.

Laumann, E. O. (1966). *Prestige and Association in an Urban Community*. Indianapolis, IN: Bobbs-Merrill.

Laumann, E. O., & Guttman, L. (1966). The Relative Associational Contiguity of Occupations in an Urban Setting. *American Sociological Review, 31*, 169–178.

Levine, J. H., & Spadaro, J. (1988). Occupational Mobility: A Structural Model. In B. Wellman & S. D. Berkowitz (Eds.), *Social Structures: A Network Approach* (pp. 452–475). Cambridge: Cambridge University Press.

Longhi, S., & Nandi, A. (2015). *A Practical Guide to Using Panel Data*. London: Sage.

Luijkx, R. (1994). *Comparative Loglinear Analyses of Social Mobility and Heterogamy*. Tilburg: Tilburg University Press.

Minnesota Population Center. (2015). *Integrated Public Use Microdata Series, International: Version 6.4 [Machine-Readable Database]*. Minneapolis, MN: University of Minnesota.

More, D. M., & Suchner, R. W. (1976). Occupational Situs, Prestige and Stereotypes. *Work and Occupations, 3*(2), 169–186.

Morris, R. T., & Murphy, R. J. (1959). The Situs Dimension in Occupational Structure. *American Sociological Review, 24*(2), 231–239.

Playford, C., Gayle, V., Connelly, R., & Gray, A. J. G. (2016). Administrative Social Science Data: The Challenge of Reproducible Research. *Big Data and Society, 3*(2), 1–13.

Prandy, K. (1979). Ethnic Discrimination in Employment and Housing. *Ethnic and Racial Studies, 2*(1), 66–79.

Prandy, K. (1986). Similarities of Life-Style and the Occupations of Women. In R. Crompton & M. Mann (Eds.), *Gender and Stratification* (pp. 137–153). Cambridge: Polity Press.

Prandy, K., & Bottero, W. (1998). The Use of Marriage Data to Measure the Social Order in Nineteenth-Century Britain. *Sociological Research Online, 3*(1), U43–U54.

Prandy, K., & Lambert, P. S. (2003). Marriage, Social Distance and the Social Space: An Alternative Derivation and Validation of the Cambridge Scale. *Sociology, 37*(3), 397–411.

Rose, D., & Harrison, E. (Eds.). (2010). *Social Class in Europe: An Introduction to the European Socio-economic Classification*. London: Routledge.

Rytina, S., & Bianchi, S. M. (1984). Occupational Reclassification and Changes in Distribution by Gender. *Monthly Labor Review, 107*(3), 11–17.

Shaw, A. (2000). *Kinship and Continuity: Pakistani Families in Britain*. Amsterdam: Harwood Academic.

Sokoloff, N. J. (1992). *Black Women and White Women in the Professions: Occupational Segregation by Race and Gender 1960–1980*. New York: Routledge.

Stewart, A., Prandy, K., & Blackburn, R. M. (1980). *Social Stratification and Occupations*. London: Macmillan.

Treiman, D. J. (2009). *Quantitative Data Analysis: Doing Social Research to Test Ideas*. New York: Jossey Bass.

Turner, H., & Firth, D. (2007). *Generalized Nonlinear Models in R: An Overview of the GNM Package*. Southampton: ESRC National Centre for Research Methods, Paper 6/07 of the NCRM Working Paper Series.

Waldinger, R., & Lichter, M. I. (2003). *How the Other Half Works: Immigration and the Social Organization of Labor*. Berkeley: University of California Press.
Weeden, K. A., & Grusky, D. B. (2012). The Three Worlds of Inequality. *American Journal of Sociology, 117*(6), 1723–1785.
Wong, R. S. K. (2010). *Association Models*. Los Angeles: Sage.
Yamaguchi, K. (1983). The Structure of Intergenerational Occupational-Mobility: Generality and Specificity in Resources, Channels, and Barriers. *American Journal of Sociology, 88*(4), 718–745.

# 7

# Networked Occupations

## 7.1 Introduction

In this chapter, we explore alternative approaches that can be used to identify and summarise specific combinations of occupations between which social connections are unusually common. Whereas the social interaction distance approach (Chaps. 4, 5, and 6) highlights dimensional structures that reflect social interactions between occupations, we can sometimes draw different conclusions by looking at the same social interaction patterns through different analytical approaches.

## 7.2 Defining 'Networked Occupations'

### 7.2.1 Definitional Criteria

We start once again with data on pairs of occupations, and the number of social interactions between them (it is convenient to work with data in the same format as described in Chap. 6, Table 6.2). We define as 'networked occupations' any specific pairs of (different) occupations that



P. Lambert, D. Griffiths, *Social Inequalities and Occupational Stratification*,
https://doi.org/10.1057/978-1-137-02253-0_7

occur disproportionately often. It is useful to label these as 'networked' because we are exploring patterns of social links between occupations—that is, when occupations are disproportionately often linked by social connections. Of course, what volume of links will constitute a disproportionately common occurrence is open to negotiation. In principle, because many occupations will have no social connections between them at all, it could reasonably be argued that pairs are 'networked' if there are any observed social connections between them at all. However, in the forthcoming analysis, we use statistical criteria that take account of both the absolute and relative prevalence of the combination, and these usually mean, in practice, that multiple realisations of the social connection must occur before it is defined as 'networked'.

To provide an example, imagine a society in which 5% of all married women are nurses. If marriage ties are uninfluenced by occupation, then 5% of the married men in every occupation might be expected to have a spouse who is a nurse. If we found, however, that 20% of male doctors were married to a nurse, this would suggest that this combination was disproportionately common. Indeed, if 1% of all married working men were doctors, we might expect that 0.05% of all marriages (1 in every 2000, i.e. 5% of 1%) were between a female nurse and a male doctor; if we actually found this combination occurring once in every 500 couples, we could say it arises four times more often than we would expect if marriage ties were distributed by chance.

To define a statistic that will indicate whether pairs of occupations might be considered to be 'networked occupations', we produce a 'representation ratio' for occupational combinations by comparing the number of times we observe a social connection between two occupations, with the number we would expect to see, given the size of the two groups, if social connections were distributed randomly. If the above-mentioned example was applied to a sample of two million both-working heterosexual couples, we would anticipate seeing 1000 instances of a male doctor married to a female nurse. By comparing the number actually observed to that expected figure, the representation ratio would tell us whether more ties, or fewer than expected, occurred. Using IPUMS-I data on the 2000 US Census, we observe that amongst both-working couples there are 15,545 male physicians and surgeons (0.7% of all males) and 80,437 female nurses (3.7% of all females). Given there were 2,191,104 couples in which both partners had a current job, this means we would expect a male doctor to be married in a female nurse in 0.026% of cases (3.7% of 0.7%), or 571 records. This combination is actually observed 2151 times in the data, implying a representation ratio of 3.8: it occurs 3.8 times more often than we might expect if relationships we formed by chance. Typically, we would define 'networked occupations' as those combinations (of non-equivalent occupations) that have a representation ratio that exceeds an agreed threshold. Different values of the representation ratio can be used for this threshold, but we have most often used the value 2. That is, we usually define networked occupations as combinations of different occupations between which social interactions occur two or more times as often as would be expected if social connections arose by chance.[1]

For relatively large occupational groups and with large datasets, it may be appropriate to use the representation ratio criterion without any further qualifications. However, with smaller groups or datasets, a chance combination between two sparsely populated occupations could be unduly influential. For example, within the IPUMS-I US Census 2000 sample, there are just 59 male gaming cage workers and six female ships engineers: statistically, if marriage ties were formed by chance, we might anticipate observing one instance of this combination within each 13.6 billion marriages. This in turn means that just a single occurrence of this combination in the IPUMS-I dataset (of two million couples) would be enough to generate a representation ratio of more than 6000.

Accordingly, it usually makes sense to set further conditions for defining 'networked occupations' in addition to the value of the 'representation ratio'. One option is to require that a certain number of instances of the combination must be observed. We have used this approach in some applications (e.g. Griffiths and Lambert 2012), and it is also used in some other studies with a similar approach (e.g. Toubol and Larsen 2017). For instance, we have sometimes defined a combination as a pair of networked occupations if the representation ratio exceeds 2 and, in addition, if there are at least ten occurrences of the combination in the dataset, and if the combination occurs at least once in every 200,000 occupational pairings. This approach is conveniently simple, but the criteria themselves are somewhat arbitrary and might not be optimal for all sample sizes—for instance, in a sample of 200,000 couples or less, any observed couples would immediately meet the second criteria; in a sample of 10,000 couples, it is plausible that many combinations might occur more than once, but less than 10 times.

Arguably a more flexible approach is to construct a confidence interval for the representation ratio, and then only define combinations as networked occupations if the lower boundary of the confidence interval is above the nominated threshold. A confidence interval constructed in this way is likely to be much larger if the occupational combination involves relatively few cases, which means that the representation ratio's lower bound would be unlikely to exceed the nominated threshold, even though the value of the ratio did so. In most circumstances, the use of confidence intervals for representation ratios seems to us to be a more appropriate way to control for variations in the number of cases in occupations (see also Sect. 7.2.3).

Table 7.1 shows some of the most over-represented 'microclass' combinations in the USA in 2000 based upon IPUMS-I data. The microclasses are aggregations of occupations that retain a relatively fine level of occupational detail (e.g. Weeden and Grusky 2012). The data is based upon a large sample. Every combination that is listed in this table meets the criteria that the lower bound of the 95% confidence interval for the representation ratio exceeds the value 2.

The top five rows of Table 7.1 show which combinations have the highest representation ratios for the entire dataset (to be precise, we are showing those that are estimated to have the highest lower bound to their 95% confidence interval for the representation ratio). Perhaps unsurprisingly, these are all examples where the husband and wife are in the same job ('diagonals' in the language of Chap. 6). The most over-represented connection is between ships officers (its representation ratio point estimate is 298). There are only eight such partnerships, but there are only 44 female ships officers with a working partner in the entire data extract—thus 18% of female ships officers are occupationally homogamous, which constitutes many more partnerships that would be expected if relationships were distributed by chance. Ordinarily, however, we would not label any of these diagonal cases as 'networked occupations', because they do not tell us about connections between different occupations.

**Table 7.1** Most over-represented 'microclass' combinations for married/cohabiting couples in the USA in 2000

| Male job | Female job | Expected | Observed | Representation ratio *(95% CI)* |
|---|---|---|---|---|
| *Most over-represented* | | | | |
| Building managers and proprietors | Building managers and proprietors | 3.02 | 577 | 175–206 |
| Aircraft pilots and navigators | Aircraft pilots and navigators | 0.37 | 68 | 140–227 |
| Ships officers | Ships officers | 0.027 | 8 | 91–504 |
| Forestry workers | Forestry workers | 0.57 | 57 | 75–127 |
| Armed forces | Armed forces | 2.06 | 178 | 74–108 |
| *Most over-represented (excluding same job)* | | | | |
| Farmers | Farm labourers | 11.7 | 238 | 17.8–22.9 |
| Bakers | Food processors | 1.9 | 26 | 8.4–19.0 |
| Tailors | Textile workers | 13.7 | 126 | 7.6–14.9 |
| Fishermen | Farm labourers | 158.0 | 966 | 5.7–6.5 |
| Launderers | Tailors | 5.4 | 37 | 4.6–9.1 |
| *Most over-represented (without shared workplace)* | | | | |
| Jurists | Statistical and social scientists | 67 | 295 | 3.9–4.9 |
| Creative artists | Authors and journalists | 94 | 386 | 3.7–4.5 |
| Health professionals | Statistical and social scientists | 89 | 357 | 3.6–4.4 |
| Jurists | Authors and journalists | 116 | 426 | 3.3–4.0 |
| Teachers | Librarians | 192 | 667 | 3.2–3.7 |
| Protective service workers | Newsboys and deliverymen | 143 | 497 | 3.2–3.8 |
| Sawyers | Textile workers | 41 | 150 | 3.0–4.2 |
| Statistical and social scientists | Authors and journalists | 31 | 106 | 2.8–4.1 |
| Statistical and social scientists | Social workers | 88 | 271 | 2.7–3.4 |
| Painters | Housekeepers | 330 | 908 | 2.6–2.9 |
| Architects | Authors and journalists | 26 | 82 | 2.5–3.8 |

Notes: IPUMS-I data, $N$ = 2,191,104 married/cohabiting couples

The middle section of Table 7.1 shows the five most over-represented pairings from those that involve two different occupations. Upon inspection, we see that all the instances of these combinations involve occupations whose incumbents are people who are likely to share a workplace environment (or be part of the same small enterprise). The driving force of these combinations may well be 'situs', the overlapping working environment, which can be expected to foster new contacts between people from the different occupations (and/or to foster recruitment to the occupations that draws upon existing social connections). In the language of Chap. 6, all of the combinations in the middle panel might be labelled as 'pseudo-diagonals'. They certainly count as 'networked occupations'; however, there is some ambiguity over whether we would seek to focus upon them analytically, if we consider that the reason for their occurrence is transparent or trivial (see further in Sect. 7.2.2).

The lowest panel of Table 7.1 summarises combinations that occur disproportionately often but for which there is no obvious 'situs'. The strongest connection appears to be between jurists and 'statistical or social scientists'—there are 4.7 times more couples in this combination than would be expected by chance. Although there could conceivably be some forces of situs that push people in these jobs together (e.g. perhaps both solicitors and researchers spend long hours in the same libraries), we would be more inclined to believe that similarity in social stratification circumstances and educational backgrounds are principal drivers of the over-occurrence of this combination. In such ways, by identifying combinations of networked occupations, we can gain insights to social mechanisms and social structure at a relatively fine-grained level. Indeed, because many of the most over-represented occupational combinations arise for apparently obvious or trivial reasons (such as of shared occupational situs), it may be less sociologically revealing to focus upon them in an analysis, and more compelling to redirect our attention primarily to those networked occupations that don't have obvious explanations—such as those in the third panel of Table 7.1 (but see also Sect. 7.2.2).

The practical identification of 'networked occupations' can be an iterative process. If we identify some combinations with an obvious reason for over-representation, we could exclude them and then recalculate the representation ratios for each remaining combination. We didn't do this in

Table 7.1, where the figures for panels 2 and 3 are based on using all the cases in the dataset, but there are scenarios when it is sensible to work on subsets of the data in this way. More generally, we should also recognise that the presence of a networked occupation limits the opportunity for the same occupations to be involved in other over-represented combinations. In Table 7.1 (panel 3), for example, we can see that female statistical and social scientists have 652 ties to two occupations (jurists, and health professionals) when we would only expect them to have 137. This means there are 515 fewer to be dispensed to other groups. As there are 6625 females in that role within the dataset, we could expect an average representation ratio of 0.92 (not 1) for all other microclasses. That is, when connections are made between some occupations, the potential for other over-representations to be observed is reduced.

An alternative approach is to use a 'popularity' method to define networked occupations. In this approach, we would identify the other (non-diagonal) occupational group that has the most frequent social connections to a given occupation (irrespective of the relative volume of occurrences). This would provide a systematic summary of networked occupations across the occupational distribution. Figure 7.1 shows the distribution of the most common non-diagonal husband-wife combinations for each of the 475 male occupations in the 2000 US census, by CAMSIS score. A relationship between the most popular ties and social stratification is evident—there is a weighted correlation between the CAMSIS scores of the two occupations involved in popular ties of 0.80 (approximately twice the value of the correlation between all husband-wife occupational combinations). Nevertheless, it is also evident that some of the most popular combinations are not so close in their CAMSIS scores. On inspection (the details are not shown within Fig. 7.1), we found that the exceptions could usually be linked to 'pseudo-diagonal' relationships reflecting workplace situs and/or urban-rural geography. In addition, a few of the exceptions seemed to reflect other orthogonal dimensionality in social interactions, such as links that occur disproportionately often between certain jobs that have distinctive ethnic group profiles.

Clearly, if we focus only upon the most popular combination for a given occupation, we could exclude from attention other combinations

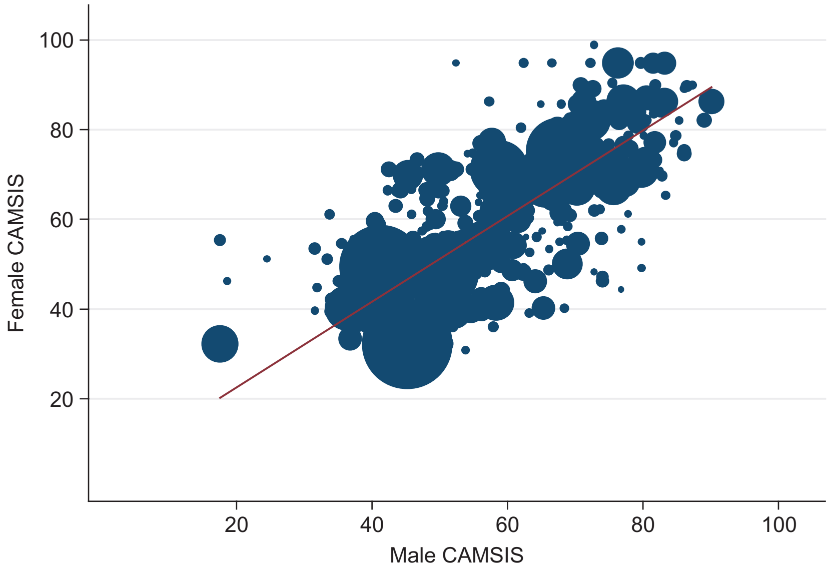



**Fig. 7.1** Most over-represented connections for male occupations. Source: USA 2000 census: IPUMS-I. R = 0.80

which are considerably over-represented and occur often, but are not the most over-represented for the relevant job. In Table 7.1, for example, there were several occupations that were associated with more than one combination of 'networked occupations', but many of these cases would not be highlighted by the 'popularity' criterion (because they would not be the most popular of the ties involving the job). A further drawback is that we might highlight some combinations that, in relative terms, are only rather modestly over-represented. In the US data, for example, for three occupations (cargo and freight agents; couriers and messengers; retail salespersons), the highest over-representation ratio is less than 2. A further practical difficulty of the 'popularity' approach arises when working with 'directed' social interaction records, such as connections between male and female partners, because the list of most popular combinations will be different depending on whether we focus on the most popular connections to egos (e.g. male occupations) or to alters (e.g. female occupations). In these situations, it is important to make a clear statement on which type of directed ties are summarised. For all of these reasons, we have generally favoured using the representation ratio, rather than a popularity principle, as a means of identifying 'networked occupations'.

### 7.2.2 Networked Occupations and 'Pseudo-Diagonals'

As evident from Table 7.1, many 'networked occupations'—however they are identified—could also be described as 'pseudo-diagonals' in the social connections between occupations. We used this term in Chap. 6 to refer to combinations of specific occupations that occur disproportionately often for reasons that do not seem to be due to the general influence of social stratification, but instead represent some other evident social mechanism. In the SID tradition, it was often simplest to remove these combinations from further analysis—this is typically done to better characterise the wider structure lying beneath these mechanisms.

In some scenarios, it makes good sense to focus upon only the less clearly explained examples of networked occupations. In Table 7.1, for example, this means focussing attention upon the third panel of the table. This could direct our attention towards more general and less overtly defined mechanisms that influence the volume of social interactions. A practical limitation, however, can be that the volume of unusually common social connections between occupations that are neither pseudo-diagonal nor diagonal can often be relatively modest. In the USA, for example, 11% of male farmers are married to female farm labourers (who comprise 0.6% of married female workers). This large pseudo-diagonal pattern accounts for a hefty proportion of non-diagonal social connections involving both male farmers, and female farm labourers; by corollary, the opportunities for identifying other networked occupations involving either of these two groups are diminished.

If we are studying specific patterns of over-representation between occupations (i.e. 'networked occupations'), it is less clear that we would always exclude 'pseudo-diagonal' circumstances. An analysis of 'networked occupations' might be designed to reveal the broad range of disproportionately common social connections, meaning that it might be sensible to include all combinations in an analysis, including those that might seem to have very obvious explanations. Also, networked occupations that are pseudo-diagonals might have important roles in a wider set of relations between occupations—perhaps acting as 'bridging ties' that connect otherwise disparate positions.

In any case, whilst in the SID tradition we might exclude pseudo-diagonals as a means of concentrating upon the 'core' dimensions of social interaction patterns, this doesn't mean that pseudo-diagonal combinations are completely independent from other social forces. For example, Table 7.2 shows 'representation ratios' for the social connections between male physicians/surgeons, and their female partners in other occupations from the US health/dentistry sector. Within the sector, we might conventionally treat all of the social connections that are listed as influenced by 'situs' (workplace contiguity)—in a SID analysis, we might code all of these combinations as 'pseudo-diagonals' and discount them from the analysis. Nevertheless, Table 7.2 suggests that the general force of social stratification also influences the volume of these social interactions—for example, the extent to which ties occur is much higher for those occupations that are more advantaged in the stratification structure as based on their CAMSIS score. In the SID case, excluding these combinations from analysis should not have negative consequences for our ability to depict the wider underlying dimensional structures linked to social interactions. However, in terms of an analysis

**Table 7.2** Marriage/cohabitations between male physicians and surgeons, and selected other health sector occupations

| | Lower bound of representation ratio | Female CAMSIS score |
|---|---|---|
| Dentists | 15.0 | 87 |
| Optometrists | 9.8 | 83 |
| Registered nurses | 4.2 | 71 |
| Dietitians and nutritionists | 3.8 | 77 |
| Radiation therapists | 1.7 | 58 |
| Dental hygienists | 0.8 | 67 |
| Emergency medical technicians and paramedics | 0.4 | 50 |
| Dental assistants | 0.3 | 44 |

Source: US Census 2000, accessed from IPUMS-I of ‘networked occupations’ which seeks to identify and interpret over-represented combinations, it is much less convincing to think that the over-represented combinations from Table 7.2 should be excluded from further attention.

### 7.2.3 Occupations with Few Incumbents

Occupational categories that are represented by few incumbents are common in most datasets on social connections involving occupations. In a SID analysis, it is common practice to combine smaller occupational groups in order to produce statistically robust results (see Sect. 6.7), but when focussing upon ‘networked occupations’ this is not as convincing as a strategy. For example, in an analysis using 2002 Romanian census data, we identified only four male typists, and five female ship’s deck crew (Griffiths and Lambert 2012). In a SID analysis, we recoded these categories to merge them with other similar occupations from respectively the male and female distributions, but in identifying networked occupations, we used the original occupational codes. Retaining the original codes did, however, effectively prevent either group from having any chance of being listed within a combination of networked occupations (because the confidence interval or minimum number of cases criteria would never be met). Indeed, in that analysis we suggested that nearly a quarter of occupations in census datasets across countries may contain too few women from both-working couples for the women’s occupations to realistically produce any networked occupations at all (Griffiths and Lambert 2012).

It is sensible to retain even sparsely represented occupations in their original units because an analysis of networked occupations is usually designed to identify specific combinations that occur unusually often. As such, those combinations that involve smaller occupational groups are intrinsically of limited interest. In practice, occupations that are not represented by many incumbents would usually be precluded from being flagged as networked occupations, because they would not usually meet the standard criteria associated with minimum sample size, and/or they would yield wide margins of error that overlap the defined threshold.[2]

This may be preferable, however, to merging the sparse occupations with others, which would risk conflating different social mechanisms linked to the different occupations.

As mentioned above, a confidence interval for the representation ratio often provides useful data about the social connection. If we treat the data on social interactions between occupations as if it were from a random sample,[3] a 95% confidence interval for the representation ratio provides a statistic that is informed by the number of cases representing the combination, and the sample as a whole, and reflects a range of values in which it is very plausible that the true population representation ratio will lie. Thus, we can treat two occupations as 'networked' if the confidence interval for their representation ratio is such that its population value is unlikely to be smaller than the target threshold—that is, the lower bound of its confidence interval exceeds the nominated threshold, such as being greater than 2. A lower boundary for the confidence interval can be estimated by finding the standard error for the observed proportion of the combination in the population, multiplying it by the appropriate normal distributional statistic (e.g. 1.96 for a two-sided 95% threshold), and then subtracting that from the observed proportion. The standard error for a proportion '$p$' can be calculated as the square root of $(p * (1 - p))/n$, when n represents the total sample size.[4] For the example discussed earlier, the combination of male doctors married to female nurses in the IPUMS-I sample for the USA in 2000 occurred 2151 times in the sample of 2,191,104 marriages. This leads to an observed proportion of 0.00098, and this number would be contrasted with the proportion expected if there were no relationship between spouses' jobs (0.00026, i.e. the proportion of all males who are doctors multiplied by the proportion of all females who are nurses). For the observed proportion we can calculate a standard error of 0.00002, which in turn means the lower bound of the representation ratio would equal $(0.00098 - (1.96 * 0.00002))/0.00026 = 3.62$.[5] This value compares to the point estimate for the representation ratio, 3.77. To adjudicate on whether the combination were treated as 'networked occupations', therefore, we could assess whether the lower bound value of 3.62 (rather than 3.77) was in excess of the agreed threshold. Using the lower bound of the confidence interval in this way has the advantage that it avoids using an arbitrary minimum number criteria for selecting combinations, which should in turn provide a more balanced response to limitations associated with the overall sample size.

A confidence interval calculation is especially helpful when there are few cases representing a given occupational combination, yet nevertheless the combination has a high representation ratio (typically because the occupations themselves are uncommon). Table 7.3 shows five combinations which occurred only once within the US 2000 census data, but which were statistically unlikely to have occurred, given the respective number of incumbents of each occupation. For instance, the single instance of a male typist marrying a female earth driller means that that combination has a representation ratio of more than 1000! However, as Table 7.3 shows, once we attach 95% confidence intervals to these ratios, we see in all cases a very wide margin of error, which overlaps a commonly used threshold value, and implies that we would not define these combinations as examples of networked occupations (a decision that makes sense intuitively). At the same time, for more fully represented combinations of occupations, confidence intervals for the representation ratio will be much smaller, and unusually common occurrences are still

**Table 7.3** The most unusual occupational combinations amongst heterosexual couples as observed in the USA in 2000

| Husband | Wife | Expected no. | Observed number | Representation ratio (RR) | RR 95% confidence interval |
|---|---|---|---|---|---|
| Word processors and typists | Earth drillers | 0.0007 | 1 | 1367 | 0.0003–2734 |
| Hotel desk clerks | Earth drillers | 0.0019 | 1 | 515 | 0.0001–1031 |
| Materials engineers | Derrick and rotary drill operators | 0.0021 | 1 | 469 | 0.0001–938 |
| Other extraction workers | Riggers | 0.0031 | 1 | 324 | 0.0001–647 |
| Message therapists | Motion picture projectionists | 0.0031 | 1 | 321 | 0.0001–643 |

Source: As Table 7.2

**Table 7.4** The most common combinations of occupations amongst heterosexual couples in the USA in 2000

| Husband | Wife | Expected no. | Observed no. | Representation ratio (RR) | RR 95% confidence interval |
|---|---|---|---|---|---|
| Truck drivers | Secretaries | 7118 | 6913 | 0.971 | 0.959–0.982 |
| Truck drivers | Primary school teachers | 4662 | 2253 | 0.483 | 0.473–0.493 |
| Truck drivers | Nurses | 3926 | 2636 | 0.671 | 0.658–0.684 |
| Retail supervisors | Secretaries | 3392 | 3790 | 1.11 | 1.10–1.14 |
| Truck drivers | Accounts clerks | 3151 | 3303 | 1.05 | 1.03–1.07 |

Source: As Table 7.2

likely to be confirmed as beyond the threshold of a networked occupation—Table 7.4, for instance, illustrates the opposite scenario, listing some of the most common male-female combinations in the US 2000 data and showing the very small estimated margin of error around their ratio.

## 7.3 Patterns of Networked Occupations in the Contemporary USA

The next section provides some elaboration and interpretation on patterns in networked occupations for the USA in 2000. Amongst other possibilities, the large amounts of data mean that it is plausible and interesting to examine how patterns vary between different states.

For the US census data, there were 475 occupational groupings, meaning 225,150 possible male-female permutations, excluding diagonals. In fact, 48% (107,205) of the possible permutations occur at least once in the data, but over a quarter of these have only one instance, and less than a third involve five or more couples. This sparsity arises in part because of the social structuring of social connections, the phenomena that we are focussing attention on; it also arises, however, simply because many of the permutations that feature no connections, or very few connections, involve occupations which themselves have relatively few incumbents.

### 7.3.1 Common Networked Occupations

In the US 2000 census data, as is a standard finding in any society with high levels of formal female employment, occupational homogamy is quite strong (e.g. Blossfeld and Timm 2003; McPherson et al. 2001). The nature of homogamy can be explored in a number of ways. As was illustrated in Fig. 7.1, one option is to code the occupations by their CAMSIS scale scores and report the correlations between the husbands' and wives' scores (very similar results would be revealed by using occupation-based measures other than CAMSIS). Across the whole population, we see a Pearson's correlation of 0.37 between the husbands' and wives' CAMSIS scores, although this relationship is much stronger (0.80) if we focus only upon networked occupations. Nevertheless, whilst the occupations of husbands and of wives are likely to be socially similar, they are by no means identical (in part a consequence of occupational gender segregation): across the US 2000 sample, for instance, only 4.2% of couples are in exactly the same occupation, and 8.3% of couples are in occupations that fall into the same 'microclass'.

Many of the most commonly observed 'networked occupations' in the USA share workplace environments, suggesting the influence of 'situs'. Aside from the examples evident in Tables 7.1 and 7.2, it is disproportionately common to find waitresses married to cooks, chefs married to food service managers, and female nurses married to male doctors, pharmacists and health managers. It is also common to find farmers married to agricultural workers, and intermarriage within both the educational and legal sectors. Clearly, the disproportionate frequency of many combinations can be explained, at least in part, by sectoral proximity.

However, we also see evidence from the same tables of a general influence of social stratification in defining the most commonly connected occupations. Not only do networked occupations commonly occupy a similar position in the social stratification structure (e.g. Fig. 7.1, Tables 7.1 and 7.2), but the results also suggest that some of the tightest social bonds within occupations in the USA are found in some of the most advantaged occupations, namely, those that involve professionalised jobs that often feature extended university education (e.g. Table 7.1, panel 3). Table 7.5 provides an additional demonstration. By focussing upon those

**Table 7.5** Networked occupations that occur in at least 75% of sampled US states (USA Census 2000)

| #States | Male occupation | Female occupation |
|---|---|---|
| 23 | Physicians and surgeons (82.5) | Lawyers (81.5) |
| 22 | Clergy (61.2) | Musicians (67.5) |
| 20 | Postsecondary teachers (79.8) | Physicians and surgeons (82.4) |
| | Education administrators (71.2) | Counsellors (65.0) |
| 19 | Clergy (61.2) | Primary school teachers (66.2) |
| | Postsecondary teachers (79.8) | Psychologists (86.5) |
| | Lawyers (81.5) | Postsecondary teachers (79.8) |
| 18 | Lawyers (81.5) | Physicians and surgeons (82.5) |
| | Physicians and surgeons (82.5) | Postsecondary teachers (79.8) |
| 17 | Postsecondary teachers (79.8) | Physical scientists (85.9) |
| | Police patrol officers (53.1) | Dispatchers (43.6) |

Source: Table 7.2. Table 7.2 shows occupational titles and corresponding CAMSIS scale scores. The combinations shown are restricted to those that do not feature an obvious workplace-based connection and have lower bound of representation ratio confidence interval above 2

occupations that consistently emerge as networked occupations (within the 24 US states used in Sect. 7.3.2, and excluding examples that might be classified as 'pseudo-diagonals'), it picks up patterns of social connections between occupations that are likely to be particularly stable and robust. In Table 7.5, links between 'professionals' dominate—for instance, associations involving doctors, lawyers, postsecondary teachers, and physicians being married to each other.

Indeed, for the US 2000 data, only one of the top ten networked occupations involved non-manual occupations (Table 7.1, panel 3). This implies that the tightly bonded occupational combinations which occur are frequently characterised by the possession of university education. There is less evidence of systematic bonding amongst occupations elsewhere in the stratification order; the analysis of networked occupations in the USA seems to point to educational participation as a key influence upon social connections between occupations.

### 7.3.2 State-Level Variations

Figure 7.2 shows further information on the spousal association in CAMSIS scores, broken down by US states for the 24 states which

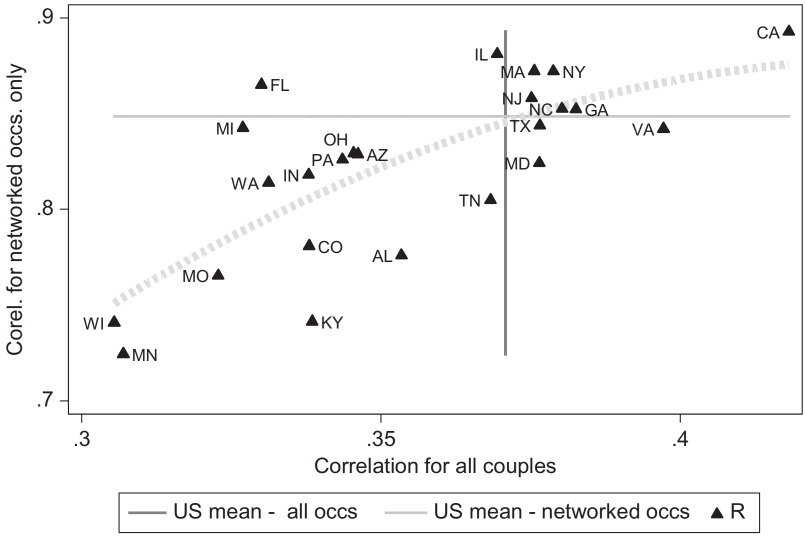



**Fig. 7.2** State-level correlations between husbands and wives in CAMSIS score for all occupations and for networked occupations (USA 2000). Source: IPUMS-I data for US 2000, 24 largest states. $N$ = 1.7 m couples (all couples, $r$ = 0.37); 85k couples (in networked occupations only, $r$ = 0.85)

contain at least 30,000 both-working couples. The horizontal axis shows the association between all couples in the state—it shows moderate levels of correlation and some variance between states. The association for 'networked occupations', shown on the vertical axis, is the correlation in scores for only the group of husband-wife occupational combinations which have social connections between them at least twice as often as would be expected is social connections were made randomly. This correlation (0.85) is generally much stronger than the correlation amongst all couples. Again, there is some variation at state level in these correlations. In general, there is also a fairly strong relationship between the two correlation values for each state.

The patterns of difference between states that are evident from Fig. 7.2 could be interpreted in several different ways. Generally speaking, the more populous and more urbanised states seem to have higher correlation patterns, but there are exceptions. The correlations for the networked occupations do seem however to be more closely associated with a dimension of size and/or urbanisation (compared to differences in the overall correlations). The implication of that pattern is that when occupational combinations are particularly common in larger and more urban states, they tend to be quite similar in terms of stratification position; in smaller and less urbanised states, it is apparently more common to find networked occupations that are not as strongly defined around a social stratification dimension.

Figure 7.3 then looks at the relationship between the husband-wife CAMSIS score correlation and three measures of state-level characteristics: income inequality (based on state-level Gini statistics, derived from American Community Survey data; see Ruggles et al. 2009), the educational profile of the state (based upon the percentage of residents with university degrees, calculated from the 2000 Census data), and a 'social capital index' (Putnam's state-level index; see Putnam 2000).

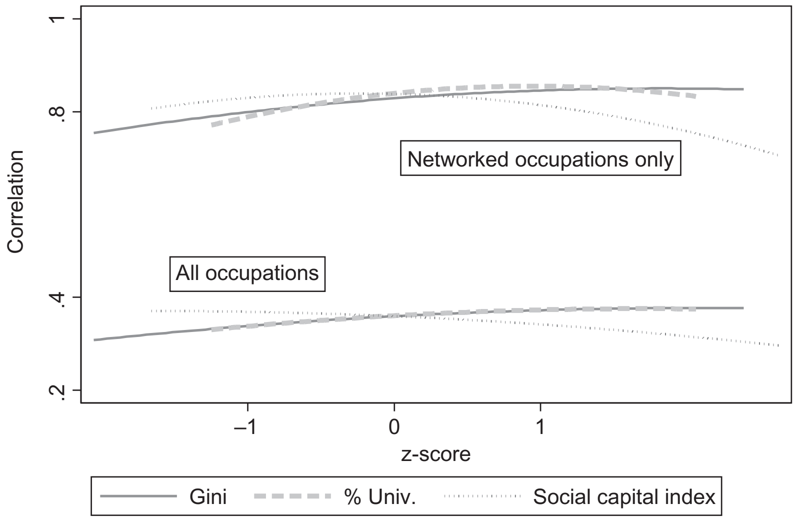



**Fig. 7.3** The relationship between state-level husband-wife CAMSIS correlations, and other state-level indicators (USA 2000). Source: IPUMS-I data for US 2000, 24 largest states. *N* = 1.7 m couples (85k couples in state-specific networked occupations)

Figure 7.3 suggests that variations in the husband-wife correlation are somewhat positively correlated with variations in levels of income inequality and in education levels, and may have a slight negative correlation with the indicator of social capital. At face value, the associations between husband-wife correlations and income inequality might suggest that both serve to index the extent and scale of wider social inequalities, but they might also suggest simply that, as educational levels rise, individuals increasingly develop social contacts around their educational background. The social capital index, on the other hand, suggests the relevance of a different process again. Putnam (2000) argues that social capital tends to be higher for people with more diverse networks, so the relationship with this index suggests that those with more variety in their occupational connections do indeed enjoy higher—potentially more beneficial—social capital (the index is designed to capture differences such as engagement in voluntary organisations and public affairs, volunteerism, sociability and social trust). Accordingly, Fig. 7.3 also reveals the intrinsic difficulties of drawing conclusions from state-level patterns, because there are often several different features of states that tend to be correlated to each other, and there is no simple way to disentangle the influences. Wilkinson and Pickett (2009) stressed the potential relevance of income inequality patterns in accounting for other social differences, but in this analysis we see that whilst a relationship with income inequality is plausible, it is just one of a few indicators that might correlate social connections patterns.

In summary, states with lower levels of homogamy appear to be smaller and less urbanised and have lower proportions of graduates, lower levels of income equality, and higher levels of social capital. This is an interesting, and not necessarily expected, finding. Arguments from Wilkinson and Pickett (2009) and Putnam (2000) assert that increased education should lead to increased social capital and less income inequality, leading to further social benefits. This analysis suggests the opposite position could be true, namely, that educational attainment can increase social distance, lessening the volume of interactions across social divides, which in turn lowers social capital and increases income inequality. This claim is so far based only on a single national analysis, but it shows how data on occupational connections might inform important debates in the social sciences.

## 7.4 Summary

The metaphor of the 'social resin' seems to provide quite a good account of the patterns associated with those occupations that are most commonly connected to each other. The social connections of occupations align in a structure that is substantially oriented around an axis of social stratification, yet the connections are also influenced by other processes and points of connection, such as by 'pseudo-diagonals' in occupational connections. The analysis of networked occupations, therefore, helps demonstrate the dual character of the 'social resin'—it is at once malleable, in that it is not entirely constraining, but it is also deeply engrained in the social stratification structure itself. Indeed, when looking at specific pairs of occupations, just as when using SID techniques to identify dimensions of difference between occupations, the underlying influence of stratification structure seems to be discernible over and above apparent 'noise' arising from other influences upon social connections. Whilst SID techniques focus on dimensional descriptions, the analysis of networked occupations offers alternative insights into the volume and scale of specific occupational social connections.

Our choice of the term 'networked occupations' was not incidental. We could, alternatively, have described them as being 'bonded' or 'linked' or with another comparable term. As many of the tables above show, there are some occupations which are 'networked' with several others (particularly amongst professionalised occupations). A natural next step is to explore the matrix of 'networked occupations' as a social network, a theme that we turn into in the next chapter. However, we believe that identifying and exploring those specific social connections between occupations that occur unusually often is a useful device that has a role to play in understanding social interactions directly. Whilst the tools of social network analysis offer some helpful options (see Chap. 8), they should not be thought of as essential in identifying and exploring 'networked occupations'.

## Notes

1. Toubol and Larsen (2017) undertake a similar exercise involving occupations that are connected by career mobility rather than social interactions. They focus on all occurrences with a ratio greater than 1, since they wish to target those occupational combinations between which career mobility is not uncommon.
2. They might, however, be flagged if we used the 'popularity' method for identifying networked occupations.
3. In some scenarios, data on social interactions might be taken from complete population datasets, in which case there is a plausible argument that standard error statistics are not required. However, even in this situation, we would argue that uncertainty statistics based upon sampling theories can help us assess the robustness of given results.
4. Other formulations for the standard error of a proportion might be considered, for instance using adjustments that more appropriately reflect the skew associated with low proportions within a dataset.
5. Our formulation allows for an uncertainty estimate around the observed proportion (the number of observed partnerships), but it does not allow for a corresponding uncertainty estimate around the expected proportion, which could in principle also be calculated.

## Bibliography

Blossfeld, H. P., & Timm, A. (Eds.). (2003). *Who Marries Whom: Educational Systems as Marriage Markets in Modern Societies*. London: Kluwer Academic.

Griffiths, D., & Lambert, P. S. (2012). Dimensions and Boundaries: Comparative Analysis of Occupational Structures Using Social Network and Social Interaction Distance Analysis. *Sociological Research Online, 17*(2), 5.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology, 27*, 415–444.

Putnam, R. D. (2000). *Bowling Alone: The Collapse and Revival of American Community*. London: Simon and Schuster.

Ruggles, S., Sobek, M., Alexander, T., Fitch, C. A., Goeken, R., Hall, P. K., King, M., & Ronnander, C. (2009). *Integrated Public Use Microdata Series: Version 4.0 [Machine-Readable Database]*. Minneapolis, MN: Minnesota Population Center [producer and distributor]. Retrieved from http://usa.ipums.org/usa/.

Toubol, J., & Larsen, A. G. (2017). Mapping the Social Class Structure: From Occupational Mobility to Social Class Categories Using Network Analysis. *Sociology, 51*(6), 1257–1276.

Weeden, K. A., & Grusky, D. B. (2012). The Three Worlds of Inequality. *American Journal of Sociology, 117*(6), 1723–1785.

Wilkinson, R. G., & Pickett, K. E. (2009). *The Spirit Level: Why More Equal Societies Almost Always do Better*. London: Allen Lane.

# 8

# Social Network Analysis of Occupational Connections

## 8.1 Introduction

Social network analysis (SNA) can be thought of as any approach to analysis that focusses upon the role of social connections between units. Marin and Wellman (2011, p. 11) argue that 'social network analysis takes as its starting point the premise that social life is created primarily and most importantly by relations and the patterns formed by these relations'. The forms of analysis of social connections between occupations that we have discussed previously could all be conceived of as examples of social network analysis, but in this chapter we look at analysing the same data using a range of analytical techniques that might be presented as the 'classical' tools of SNA.

'Classical' SNA approaches typically begin with data in a matrix which informs us whether and how different units within the network are connected to each other. Network members are often called 'actors' or 'nodes', and the connecting attributes are known as 'ties' or, depending on circumstances, 'edges' or 'arcs'. Data captured in the matrix is then described through a number of devices, one of the most common being the 'sociogram', a graphical presentation that visualises the ties between different nodes: the 'nodes' are displayed as symbols (often circles), and the 'ties'



P. Lambert, D. Griffiths, *Social Inequalities and Occupational Stratification*,
https://doi.org/10.1057/978-1-137-02253-0_8

are displayed as lines connecting them. Various texts provide an overview of methods of SNA (e.g. Borgatti et al. 2013; Kadushin 2012; de Nooy et al. 2011; Knoke and Yang 2008), or summarise the ways in which SNA has been used in social science research (e.g. Crossley 2015; Giuffre 2013; Knoke 2012; Lin 1999; Wellman and Berkowitz 1988).

In this chapter, we explore how descriptive methods of SNA can help us in analysing the social connections between the incumbents of occupations. The basic idea involves defining occupations as 'nodes', and defining the presence or absence of a tie between the nodes by using criteria about the relative prevalence of social connections between the incumbents of occupations (for instance, whether or not a pair of occupations are 'networked occupations').

## 8.2 Analysing Occupations as Nodes

### 8.2.1 Nodes and Ties

Whenever we have information on the occupations of two connected individuals, we could formulate the data in the style of a social network record. In this scenario, we would be treating the occupations as the nodes within the analysis, and the data on the social connections between occupations would contribute to how we recorded ties between each occupation. Just as was discussed in the example of constructing CAMSIS scales, this data could be preserved in a matrix format showing the relationship between ego and alter occupations (e.g. Table 6.1). Alternatively the same information could be preserved in a 'table' format data structure that lists every combination of ego-alter occupations and the number of connections (if any) between them (e.g. Table 6.2). In the language of SNA, the table format representation of the network matrix is known as an 'edge list'.

The origins of the data on social connections are the same as used in a SID analysis. For instance, we might extract ego-alter data on husband-wife combinations, the occupations of friends, or use administrative documents such as the jobs declared by witnesses on marriage certificates.

It is important however to recognise that the same occupations usually feature in the list of ego and alter records. Because of this, in the network analysis tradition it is consequential whether we define ties as 'directed' or 'undirected'. Undirected ties refer to social connections that are analysed as if they were reciprocal in terms of egos and alters, whereas directed ties do not assume reciprocity. Most data on the social connections between occupations comprises directed ties, because the number of links between an ego occupation (A) and a separate alter occupation (B) is not the same when the occupations are reversed and we consider the relation from ego occupation (B) to alter occupation (A) (for instance, the relative occurrence of male doctors married to female nurses is not usually the same as the relative occurrence of female doctors married to male nurses). In this situation, it is normally more appropriate to differentiate and summarise patterns of ties in a manner that recognises their direction.[1]

### 8.2.2 Graphical Depictions of Networks: The Sociogram

Figure 8.1 summarises data on the occupations of 27 'both-working' couples from a small locality in Scotland in the nineteenth century. We could have retrieved similar data from various sources, but this extract is derived from the North Atlantic Population Project database (Minnesota Population Center 2015; we selected, not impartially, a rural area of Scotland in which both of the authors have lived). The numerical data on the occupations of couples and their relative occurrence is summarised in the left part of the figure.

The right-hand side of Fig. 8.1 shows a graphical depiction of the numerical data, namely, a 'sociogram'. This is a popular descriptive tool of SNA. In this example, the sociogram plots a point for each occupation that features within the database and draws a line connecting any of the occupations that have a social connection between them ('ties'). For example, occupation 56 (domestic servants) has the most ties to other occupations—there are five ties shown from this occupation, because the data features records that link domestic servants to five other occupations within the set of couples. For simplicity, the sociogram shown in Fig. 8.1

| ***Husband's job*** | ***Wife's job*** | ***# ties*** |
|---|---|---|
| 2. Civil Service (officers and clerks) | 2. Civil Service (officers and clerks) | 1 |
| 5. Police | 283. Milliner, Dressmaker, Staymaker | 1 |
| 100. Farmer, Grazier | 56. Domestic Indoor Servant | 3 |
| 100. Farmer, Grazier | 100. Farmer, Grazier | 2 |
| 103. Agricultural Labourer, Farm Servant, Cottager | 103. Agricultural Labourer, Farm Servant, Cottager | 1 |
| 103. `` | 56. Domestic Indoor Servant | 2 |
| 104. Shepherd | 56. Domestic Indoor Servant | 1 |
| 110. Woodman | 110. Woodman | 1 |
| 118. Gamekeeper | 118. Gamekeeper | 1 |
| 134. Millwright | 56. Domestic Indoor Servant | 1 |
| 168. Carpenter, Joiner | 66. Merchant | 1 |
| 170. Mason | 283. Milliner, Dressmaker, Staymaker | 1 |
| 223. Milkseller, Dairyman | 223. Milkseller, Dairyman | 1 |
| 282. Tailor | 283. Milliner, Dressmaker, Staymaker | 2 |
| 290. Shoe, Boot - Maker, Dealer | 253. Cotton, Cotton Goods Manufacture | 1 |
| 362. Platelayer | 56. Domestic Indoor Servant | 1 |
| 377. Blacksmith | 377. Blacksmith | 1 |
| 404. General Labourer | 56. Domestic Indoor Servant | 3 |
| 404. General Labourer | 283. Milliner, Dressmaker, Staymaker | 1 |
| 404. General Labourer | 404. General Labourer | 1 |
| Total | (from 20 different ties) | 27 |

*Sociogram summarising 27 occupational connections:*

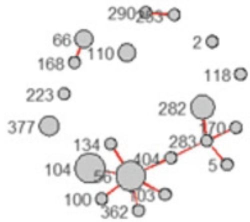



**Fig. 8.1** Occupational networks amongst both-working couples in West Perthshire in 1881. Source: Data for 27 couples, covering 20 different ties, accessed from NAPP, Scotland 1881, for selected parishes in rural West Perthshire

shows undirected ties, in this case meaning that we don't know if the ties drawn refer to connections from husbands to wives or from wives to husbands; alternatively the sociogram could be redrawn to distinguish the two directions, for instance, using arrow symbols or colour coding to illustrate the direction of the link. The general principle in this descriptive approach to SNA is that we could learn something about social structure by examining the profile of ties in the sociogram and considering the circumstances of those occupations which do and do not have ties.

As shown in the example in Fig. 8.1, within a small sample of the social relations between occupations, there could be many occupations that are only represented in the data by a single individual, and other interesting occupations may not be represented at all in the sampled population. It is also quite plausible that the patterns shown in Fig. 8.1 are

| *Husband's job* | *Wife's job* | |
|---|---|---|
| 2. Civil Service (officers and clerks) | 2. Civil Service (officers and clerks) | 9 |
| 2. `` | 56. Domestic Indoor Servant | 11 |
| 2. `` | 60. Office Keeper (not Government) | 1 |
| 2. `` | 99. Telegraph, Telephone Service | 1 |
| 2. `` | 232. Confectioner, Pastrycook | 2 |
| 2. `` | 236. Grocer. Tea, Coffee, Chocolate Maker, Dealer | 2 |
| 2. `` | 280. Hatter, Hat Manufacture | 1 |
| 2. `` | 282. Tailor | 1 |
| 2. `` | 283. Milliner, Dressmaker, Staymaker | 6 |
| .. | .. | .. |
| 411. Chimney Sweep, Soot Merchant | 56. Domestic Indoor Servant | 5 |
| 411. `` | 64. Hospital and Institution Service | 1 |
| 411. `` | 103. Agricultural Labourer, Farm Servant, Cottager | 2 |
| 411. `` | 215. Lodging, Boarding House Keeper | 2 |
| 411. `` | 240. Woollen Cloth Manufacture | 1 |
| 411. `` | 253. Cotton, Cotton Goods Manufacture | 4 |
| 411. `` | 262. Hemp, Jute, Cocoa Fibre Manufacture | 2 |
| 411. `` | 269. Weaver (undefined) | 3 |
| ... | ... | ... |
| Total | (from 4837 different ties) | 27852 |

Sociogram summarising 27k occupational connections:

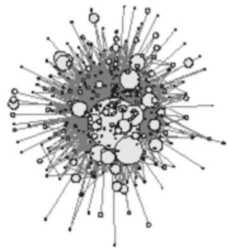


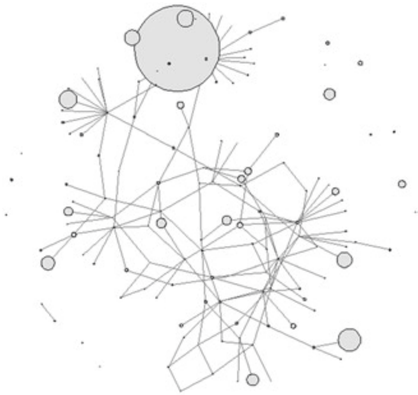
**Fig. 8.2** Occupational networks amongst both-working couples in Scotland in 1881. Source: Data for 27k couples, covering 4837 different ties, accessed from NAPP, Scotland 1881

'overanalysing' the data, since they describe just a few marriages and perhaps are shaped by happenchance circumstances. For such reasons it is rational to use much larger-scale data on the social connections between occupations if it is available. However, the simplest graphical summarises of network data do not work as effectively when the number of nodes and ties is much larger. The right-hand side of Fig. 8.2, for instance, depicts a sociogram for a sample of over 27,000 pairs: in this instance, there are so many occupations, and so many ties between occupations, that it is hard to interpret anything useful from the image. The exert from the data on all occupations (on the left of the figure) does suggest that there may be some interesting social patterns of relative difference in the volume of social connections between different jobs, but the scale of the data is such that the initial sociogram is not suited to discerning them.

A useful resolution, as anticipated in Chap. 7, is to use large samples of data on occupations, but define ties between occupations by criteria related to the relative occurrence of the tie. This means that a much smaller number of ties will be identified, which in turn makes it more feasible to interpret the distribution of those ties in an effective manner. In short, it makes sense to explore ties between only the 'networked occupations', and use the tools of SNA to uncover patterns in these relationships.

Taking the idea of 'networked occupations' (Chap. 7), we define a tie as existing between two occupations if social interactions occur between the occupations a certain number of times more often than would be expected if social connections between occupations were distributed by chance. As described more fully in Chap. 7, we would typically decide on a given threshold for the 'representation ratio' which characterises this over-representation (e.g. if the representation ratio is bigger than two, which means that link occurs more than twice as often as would be expected). We would typically also apply further criteria to exclude those ties that are only represented by very small numbers of cases, for instance, by setting a minimum requirement for the number of ties or by calculating a confidence interval for the representation ratio and requiring its lower bound to exceed the agreed threshold.

Figure 8.3 illustrates typical results that emerge after applying this approach. Using the historic Scottish data that was also used in Fig. 8.2, we now draw ties only for those husband-wife occupational combinations which occur more than twice as often as would be expected if marriage patterns were randomly distributed (and if there are at least five instantiations of the tie within the dataset). It is perhaps clear to see that by focussing only on 'over-connected' occupations, we are suddenly able to identify distinctive patterns or channels of socially connected occupations through the sociogram's visualisation.



| Male job | Female job | # ties | Male job | Female job | # ties |
|---|---|---|---|---|---|
| … | … | | … | … | |
| 344. Coal Merchant | 399. General Shopkeeper, Dealer | 6 | 6. Municipal, Parish, Union, District, Officer | 6. Municipal, Parish, Union, District, .. | 12 |
| 348. Stone Quarrier | 236. Grocer. Tea, Coffee, … | 8 | 6. Municipal, Parish, Union, District, Officer | 34. School Service, & others connected | 6 |
| 348. Stone Quarrier | 270. Dyer, Printer, Scourer, Bleacher, … | 16 | 32. Schoolmaster | 32. Schoolmaster | 100 |
| 348. Stone Quarrier | 285. Shirt Maker, Seamstress | 9 | 47. Musician, Music Master | 47. Musician, Music Master | 35 |
| 348. Stone Quarrier | 348. Stone Quarrier | 6 | 50. Actor | 47. Musician, Music Master | 6 |
| 360. Road Labourer | 82. Toll Collector, Turnpike Gate Keeper | 10 | 50. Actor | 50. Actor | 59 |
| 360. Road Labourer | 404. General Labourer | 7 | … | … | |
| | | | Total | (ties involving 883 couples) | 246 |

**Fig. 8.3** Marriage links between occupations in Scotland, 1881, using 'Threshold method'. Source: As Fig. 8.2. Data comprises the subset of 'networked occupations' from sample of 27k husband-wife combinations

More work is still required to discern meaningful patterns within the social network structure that is depicted in Fig. 8.3. At this stage, we have not yet attached labels to the nodes from Fig. 8.3 (cf. Fig. 8.1). Without labels it is hard to draw conclusions about the social connection patterns; but on the other hand, adding labels to the occupations would considerably clutter the graphical depiction. In Fig. 8.3, the nodes have been sized by the number of people in those jobs, although it transpires that there is not an obvious link between the network connections and the size of occupations. We could also have sized and/or shaded the nodes according to some other attribute of the occupations—such as their CAMSIS score, average educational level, gender profile or rural-urban statuses. In later examples, we discuss sociograms that feature each these inputs as an aid to interpretation of the network structure.

One important interpretation that can be garnered from sociograms concerns the number and character of different 'components' that are revealed in the network structure. Components are subsections of the network which are joined together through ties. Most of the occupations in Fig. 8.3 form part of the largest component, although we can see that a number of other occupations apparently have no ties (and are isolated), and there is one other small component where two nodes link to each other and are isolated from all others. Typically the distribution of components evident from a sociogram might be explored and compared against patterns that we might have expected according to a theory or hypothesis about the social structure. For the case of occupations, for example, we might expect to observe separate large components for manual and non-manual occupations (anticipating that ties within the sectors are common, but ties between them are uncommon). Thinking in terms of social stratification, we might also expect to see a series of different components that might represent different social classes. We might also have expected to see some isolated components that might represent occupations located in extreme positions in the social stratification structure. As it happens, few of these hypotheses seem plausible for the historical data summarised in Fig. 8.3: the existence of one large component that connects together the large majority of occupations could be presented as an argument against the existence of discrete boundaries in the social structure, and instead may be more consistent with a model of gradational differentiations within the component (see Griffiths and Lambert 2012). However, the character of network components can be highly contingent upon the thresholds set when defining the presence or absence of ties between nodes, so more attention should be paid to the definition of networked occupations (and other social structural information about the occupations) before drawing stronger conclusions on the basis of Fig. 8.3.

Sociograms are also often used to ask whether obvious structural patterns can be seen amongst the links within components. For example, ties within the large component in Fig. 8.3 are highly structured, as most occupations within it connect only to a few others within the component. Typically, with data on the social connections between occupations, one of the main influences upon structure within components is proximity in the social stratification structure: although it is not evident from Fig. 8.3, most of the links within the main component are made between occupations that are fairly close to each other in the stratification structure. In general, therefore, the social connections between occupations can be explored through sociograms that depict networked occupations, with dual interest in the breakdown of different network components, and patterns in the characteristics of occupations that do and do not have ties.

## 8.3 Comparative Analysis of Social Inequality Using Sociograms of Networked Occupations

Descriptions of networked occupations offer some appealing contributions to the comparative analysis of inequality structures between countries or over time. We have published some relevant findings elsewhere (e.g. Griffiths and Lambert 2012, 2013), but it is useful to reiterate two of the most important issues, as they help to illustrate the contribution made by a descriptive network-based analysis.

### 8.3.1 Patterns of Network Components for Social Interactions Do Not Support Categorical Models of Social Class

A useful way to evaluate the structures of networked occupations is to compare the observed structure against the hypothetical structure that we might expect under a relevant theoretical model. Figure 8.4 (also used in Griffiths and Lambert 2012) is one example. On the left is a representation of the hypothetical network structure that we might expect according to the literature on ‘microclasses’ (e.g. Weeden and Grusky 2012).

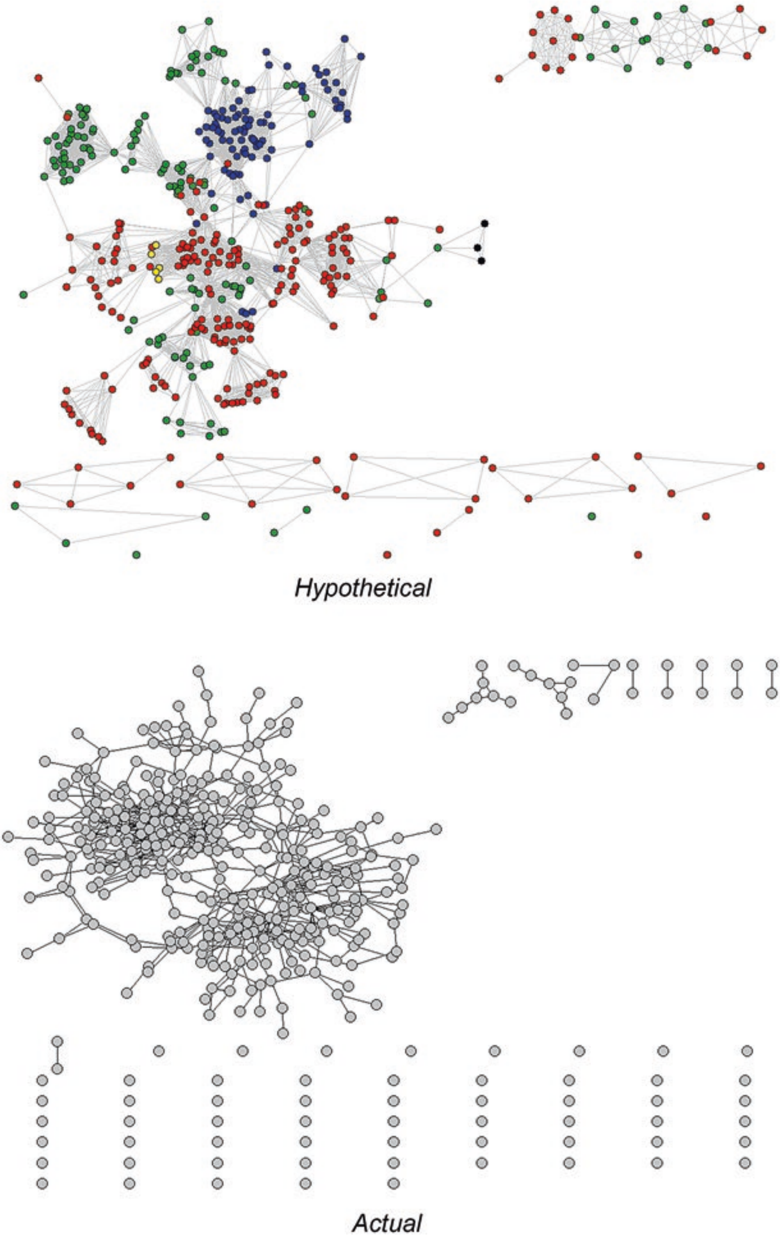



**Fig. 8.4** Actual and hypothetical network structures for social connections between occupations in the US 2000 census data. Source: Actual data based on IPUMS-I extract of husband-wife occupational ties

This literature advocates a model of many distinctive and relatively small social classes, which involve very similar occupations with high levels of social contact. The microclasses themselves also fall naturally into larger aggregates of relatively similar circumstances ('macroclasses). Accordingly, the microclass model would anticipate clusters of occupations (within microclasses) that have many links between each other, whilst there might only be occasional links connecting occupations from different microclasses (most of these would be within 'macroclasses', but some may not be, if, for instance, they reflect 'pseudo-diagonal' connections fostered when occupations from different microclasses are located in shared workplace environments). The hypothetical sociogram in Fig. 8.4 is intended to indicate the nodes and ties that might emerge under this model—it shows dense network connections within microclasses and structured patterns in the occasional ties outwith microclasses.

Nevertheless, the actual structure of networked occupations in the contemporary USA (on the right of Fig. 8.4) has little resemblance to the hypothetical depiction. On the contrary, it suggests a more gradational spread of occupational connections and does not seem to indicate strong boundaries between different groups of occupations such as microclasses (or larger aggregations of them). Though not shown, the actual network structure of Fig. 8.4 is fairly typical of the network structures that we have seen in applications across a range of countries and time periods (e.g. Griffiths and Lambert 2012). To us, this structure is an argument against the existence of crisp boundaries between social classes, at least in terms of social interaction patterns.

### 8.3.2 Network Components Are Structured by Social Stratification But Interwoven by Other Social Mechanisms

A second consistent feature of network structures that are defined by social connections between occupations is that the broad contours of the structure, both within and between network components, suggest an interplay between an important structure of social stratification, plus further influences associated with other more specific mechanisms in social

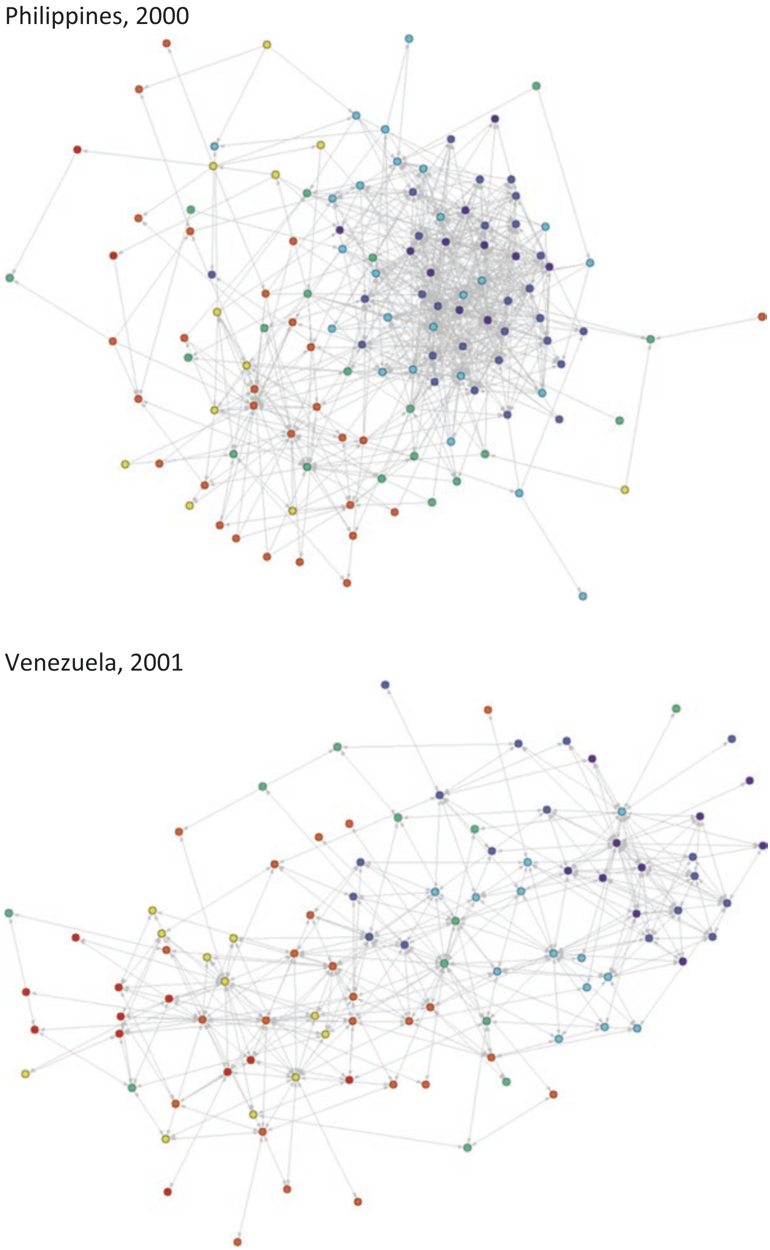



**Fig. 8.5** Social networks for social connections between occupations in the Philippines and in Venezuela. Notes: Based on data on both-working couples obtained from IPUMS-I

connections. Figure 8.5, which is also based upon results discussed in Griffiths and Lambert (2012), gives two typical examples. Taking account of the shading patterns, it is evident from the figure that in both the Philippines and Venezuela, a general structure of social stratification pervades the realised social connections between occupations: generally speaking, most social connections that are unusually common are between occupations of a similar position in the stratification structure. The influence of the stratification structure on network interactions also seems to be largely gradational rather than categorical (i.e. there are no obvious boundaries between groups or major gaps in the network patterns).

At the same time, there are also some connections which make a bridge between more different positions in the stratification structure. Upon inspection (the results are not evident from the figure), most of these bridging connections seem to be accounted for by specific social mechanisms that are not linked to social stratification—for instance, workplace situs, or shared educational institutions. A valuable contribution of the descriptive approach to SNA, indeed, is the potential insight in recognising and exploring these 'bridging' occupations, which might perhaps open channels of social connections that would not otherwise be available.

## 8.4 Case Study: Network Structures at the State Level in the Contemporary USA

A descriptive analysis of social network structures in the social connections between occupations can also be revealing when making relatively small-scale comparisons. In this section, we illustrate a network analysis of occupations for two states from the 2000 US census, Wisconsin and Texas. These had amongst the largest (Texas) and smallest (Wisconsin) correlations in CAMSIS scores within couples (Fig. 7.2), and are generally different from each other in many significant ways. For instance, Wisconsin is a medium-sized northern state with a mixed economy, and

Texas a large southern state with large agricultural and mining sectors. By comparison to national averages, both Wisconsin and Texas have relatively low educational profiles, whilst both states are cited regularly in Wilkinson and Pickett's (2009) analysis as examples of respectively low and high levels of social inequality. Equally, whereas Wisconsin has a mainly white population with European ancestry (78% in 2016),[2] Texas has a large Hispanic population (37%) and a much smaller White ethnic group (44%). In every presidential election from 1988 to 2012, Wisconsin voted Democrat and Texas Republican. Such differences in the sizes, ethnicity, economies, geography, and social outlooks in the two states could all influence the social interaction patterns as measured by occupation.

Figures 8.6 and 8.7 show network diagrams for connected occupations, for Texas and Wisconsin, respectively. Lines are drawn between occupations if connections occur at least twice as frequently as would be

**Fig. 8.6** Texas network. Source: Analysis of data for both-working couples from IPUMS-I data for 2000

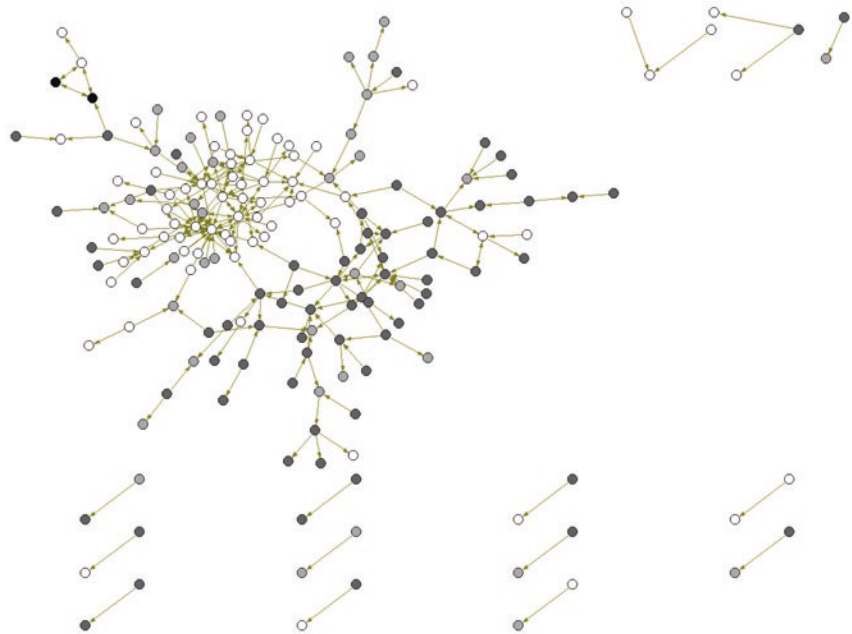


**Fig. 8.7** Wisconsin network. Source: As Fig. 8.6

expected if they were made randomly (and if the lower bound of the confidence intervals for the representation ratio exceeds 2—see Chap. 7). The occupations are shaded by five categories of a hierarchical social class scheme,[3] whereby the darker shades represent the more disadvantaged groups. Occupational nodes are only shown if they have one or more disproportionately common connections to other occupations—in both states, far fewer than the 475 available occupations are drawn. This indicates that many occupations did not have ties that were both unusually common, and met the confidence interval criteria (in practice, the latter was the more decisive influence, because many occupations had relatively few incumbents within each state).

Both figures suggest that occupational advantage has a strong relationship to social connections of marriage. In both states, most connections are between occupations in the same or adjacent class category, and there are few direct linkages between occupations in the most, and least, advantaged classes. However, most occupations are part of a single major component, which means, for instance, that indirect links between the most and least advantaged positions do exist (although they are 'brokered' by intermediate occupational positions). The main substantive conclusion from both network diagrams is similar: the gradient of social stratification is a major influence on social connections between occupations in Texas and Wisconsin alike.

Within Texas there are only two isolated components aside from the main component, and most occupations are part of the main component structure. In Wisconsin, however, there are 13 isolated components, within which the relevant occupations are strongly connected to each other, but no others. Assessment of the specific ties involved in the isolated components does not reveal obvious social mechanisms (results not shown). The higher number of components in Wisconsin may well arise as an artefact, since the sample used for Wisconsin was smaller—that is, these occupations might not really be isolated, but those connections that they do have with other occupations are not voluminous enough to meet the minimum threshold for inclusion in the sociogram.

The composition of the main network component in each state does seem to differ somewhat. Within Texas there are many darker (manual) occupations which commonly interlink, whilst the ties amongst the lighter (non-manual) occupations appear to be more sparsely distributed. By contrast, within Wisconsin it appears the lighter non-manual occupations have more instances of bonding to each other, with fewer connections amongst the darker manual categories. This might suggest that there are some class-based patterns to occupational homogamy, involving the more disadvantaged occupations in Texas, but the more advantaged in Wisconsin.

Similarly, within Texas there are many occupations which sit between the advantaged and disadvantaged segments of the main component. Because of this, within Texas there is arguably a greater overall distance between the manual and non-manual segments of the largest component (because most connections between the extremes are brokered by intermediate positions). This pattern is less evident in Wisconsin, suggesting overall a less dramatic gradient of social distance in Wisconsin than in Texas. One plausible explanation could be that the more substantial

**Table 8.1** Exemplar differences in disproportionately common linkages in Texas and Wisconsin

| | Texas | Wisconsin | Both |
|---|---|---|---|
| Lawyers (male) | Chief executives, marketing/sales managers, management analysts, counsellors, paralegals, misc. legal workers, physicians, and surgeons | Other managers, tax preparers, primary and secondary school teachers, librarians, real estate brokers | Postsecondary teachers |
| Metalworkers (male) | Building cleaners, sewing machine operators, production line workers, inspectors, and testers | Electrical assemblers, printing machine operators, packagers and fitters, freight movers | Miscellaneous assembly line workers |
| Sales managers (female) | Financial managers, civil engineers, lawyers, sales representatives | Computer and information managers, engineering managers, accountants, sales supervisors, and carpet/floor fitters | Chief executives |
| Accountants (female) | Financial managers, management analysts, personal financial specialists, network systems analysts | Computer and information managers, computer scientists, agricultural scientists | |

bonding within disadvantaged occupations in Texas serves to pull relatively more individuals away from bridging contacts with more advantaged groups, whereas by comparison the bonding within more advantaged positions in Wisconsin does not have as pronounced an exclusionary effect.

Table 8.1 lists some of the specific over-represented linkages for four selected occupations in both states: lawyers and metalworkers for men and sales managers and accountants for women. Whilst there are many connections for each occupational group, there are only three ties which are shared between both states. Workplace connections characterise a number of the ties, but there also appears to be a more general pattern of greater diversity amongst the linkages in Wisconsin. Male lawyers in Texas have disproportionate connections within their sector or with other, particularly advantaged, occupations in their sector, whereas in Wisconsin, lawyers seem to have more diverse ties, such as to school teachers, librarians, and real estate brokers. Female sales managers in both states are mostly linked to other professionals, but in Texas their only disproportionate ties to manual workers are within the workplace (sales representatives), whilst in Wisconsin they might not be (e.g. carpet and floor fitters). Meanwhile, for the male metalworkers, ties in Wisconsin appear to centre upon shared workplace, but those in Texas feature outward ties to other occupations associated with a similar position in the stratification structure (building cleaners and sewing machine operators). This interpretation of the data is subjective and alternatives are plausible, but it again suggests that social stratification position appears to drive social interaction patterns more strongly in Texas than in Wisconsin.

## 8.5 Methodological Issues in Analysing Network Structures for the Social Connections Between Occupations

### 8.5.1 Criteria for Including Network Ties

In network studies, different approaches to the measurement of a connection may well produce different network structures (e.g. Marsden 2011). In our examples, different values for the threshold for defining a tie (based on the representation ratio), and different supplementary inclusion criteria for the tie (such as requiring a nominated minimum number of cases), could both have consequences for results.

Figure 8.8 indicates some of the variations that might arise when using different threshold values and other related criteria. Using the same underlying data for all of the outputs, the first two panels show the

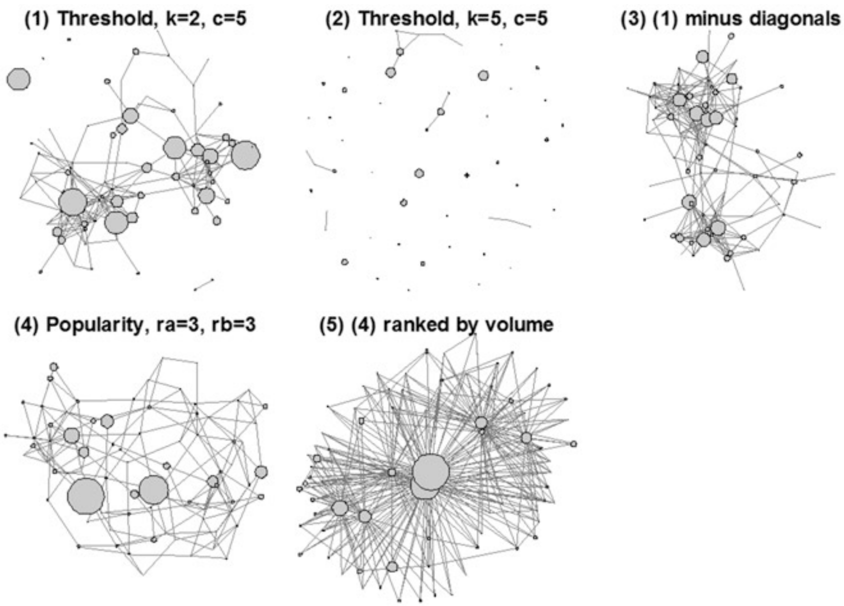



**Fig. 8.8** Network patterns for US data using different criteria to select 'networked occupations'. Source: Data on 2.1 million occupational connections by marriage/cohabitation. Data for USA 2000, accessed from IPUMS-I

sociograms that emerge if we require a minimum representation ratio ('k') of, respectively, 2 and 5, but keep the same minimum inclusion criteria ('c'), namely, that the combination is represented by at least five pairs. As might be anticipated, by using a higher representation ratio, we considerably reduce the number of network ties that are revealed—meaning that we concentrate on fewer, but more intense, network connections. The third panel then shows another variant of these structures, namely, the same data as in panel 1 but after removing all diagonal cases from calculations. This has the effect of focussing attention on the core component (although the component is marginally different, since it is based on calculations on a subset of the data). Next, the fourth panel in Fig. 8.8 shows a rather different network structure, that obtained by using the 'popularity' method for selecting ties (see also Chap. 7). This method includes the most disproportionately common three ties for both the male occupation ('ra=3') and the female ('rb=3'), but only if they occur at least five times in the data. Ties are included regardless of the scale of over-representation, and the network structure that is revealed is spread more evenly across the occupations. Lastly, the fifth panel uses a second version of the popularity criteria, now showing the three most popular ties as ranked by volume of cases, rather than by over-representation. This vastly increases the number of ties shown, and in this case leads to a sociogram structure that is too dense to be easily interpreted. In the light of such evidence, it is important when depicting networks that details on the thresholds and criteria that are used are clearly stated. Nevertheless, different threshold values can be helpful in different circumstances—they represent different conscious decisions about whether to 'zoom in' or 'zoom out' from the relevant structure, and either perspective might be helpful.

### 8.5.2 The Influence of Occupational Detail

The total number of nodes that are allowed for in a network study can also have important consequences for how networks are depicted. For instance, ties between units are usually more likely to occur when fewer nodes are specified (e.g. Hanneman and Riddle 2011), and many argue that network tools and summary statistics about networks cannot readily be compared when the number of nodes is different (e.g. Scott 2017). In the study of 'networked occupations', we often have some choice over the number of occupational units that we could study, depending upon whether the data is operationalised at a more or less precise level of occupational detail (see Chap. 3).

In some scenarios, there might be a clear theoretical reason for focussing upon a particular level of aggregation. For example, network studies on 'interlocking directorships' examine a phenomena that requires detailed data on specific firms and appointments; however, Lin et al. (1981), evaluating how networks were linked to 'status attainment', required only broad-brush characterisations of occupations in order to represent the 'prestige' and 'social resources' of a contact. We argued earlier that theories about social stratification mechanisms generally favour working with occupational data at as fine a level of disaggregation as possible, yet we also noted the perspective that the bulk of inequalities that are linked to occupations are substantially visible at a more coarse-grained level (e.g. Ganzeboom 2005). Considerable practical challenges often arise if working with disaggregated occupational data, such as low volumes of cases in some categories, and challengingly complex patterns to the connections between many units. Sometimes the level of occupational detail is in any case constrained by administrative considerations. However, aside from practical and bureaucratic pressures, in our research into network structures we have generally concluded that more disaggregation is desirable when possible, on the grounds that certain patterns of social connections involve very specific combinations of occupations that might be masked at a more aggregate level (e.g. Griffiths and Lambert 2012). Although a summary of the relative volumes of connections between a small number of highly aggregated occupational groups is likely to lead to a reasonably neat analytical representation, it seems

**Table 8.2** Extent to which the identification of networked occupations is replicated at different levels of occupational aggregation (using three-digit and two-digit ISCO)

| | Two-digit ties replicated at three-digit level (%) | Three-digit level replicated at two-digit level (%) | Three-digit level replicated at two-digit level (excluding same two-digit combination) (%) |
|---|---|---|---|
| Cuba (2002) | 99.1 | 22.8 | 30.6 |
| France (1999) | 97.8 | 57.9 | 71.3 |
| Philippines (2000) | 89.4 | 44.3 | 57.4 |
| Portugal (2001) | 97.6 | 58.1 | 67.1 |
| South Africa (2001) | 96.7 | 41.6 | 52.9 |
| Thailand (2000) | 99.6 | 48.3 | 57.7 |
| Venezuela (2000) | 97.0 | 66.7 | 73.1 |

Notes: Data from IPUMS-I. Analysis of networked occupations for heterosexual couples. Networked occupations are defined as those with a representation ratio of at least 2 and that occur at least once in every 15,000 pairs.

improbable to us that the full complexity of social connections are reasonably represented through aggregate units.

It is often unclear in advance of an analysis what the most instructive level of aggregation will be. A sensible strategy—in an ideal world—is to undertake sensitivity analyses across different levels of occupational detail, in order to adjudicate on the most effective approach. As an illustrative example, Table 8.2 shows, for a range of societies, how patterns of 'networked occupations' can vary substantially if two-digit or three-digit occupational units are analysed. The relevant studies for the seven countries illustrated in Table 8.2 all have data available from IPUMS-I that is coded into three-digit ISCO-88 occupational unit groups, involving up to 116 different categories, and this is readily recoded to two-digit level, with 28 different categories. The columns of Table 8.2 summarise to what extent networked occupations that are identified at one level of occupational detail are also detected at the other. In the first column, we see that nearly all of those combinations that are detected at two-digit level are also represented (by at least some component occupations) in an analysis at the three-digit level. For example, if at the two-digit level there was a disproportionately common tie between 'customer service clerks' (42) and 'models, salespersons and demonstrators' (52), then at the three-digit level, it is very likely that there will be several ties involving occupations from those groups (e.g. from 'cashiers, tellers and related clerks' (421) to 'shop salespersons and demonstrators' (522)). The second and third columns of Table 8.2, however, show that there are many pairs of occupations that would be identified as linked when analysed at the three-digit level, but that would not be identified as linked when analysed at the two-digit level. These patterns also vary between countries—in Cuba only 23% of pairs of occupations would be identified at both levels of analysis, but this rises to 67% in Venezuela. A complication concerns how 'diagonals' (husbands and wives in the same occupation) are treated. 'Diagonals' themselves are contingent upon the level of occupational detail. The final column of Table 8.2 shows the percentage of three-digit ties that would also be captured at the two-digit level of analysis, but now excluding those which are diagonals at the two-digit level. In this framework, we see an increase in the number of three-digit ties that are also detected at the two-digit level—but we still see that a very substantial

**Table 8.3** Replication of networked occupations by ISCO level of definition (Romania census data, 2002)

| | More aggregated level replicated at the more refined level (%) | Refined level replicated at aggregated level (%) | Refined level replicated at aggregated level (excluding diagonals) (%) |
|---|---|---|---|
| 3 and 4 digits | 81.6 | 78.2 | 82.9 |
| 2 and 3 digits | 98.5 | 57.9 | 71.5 |
| 2 and 4 digits | 97.5 | 53.1 | 62.3 |

proportion of pairs of occupations would not be identified in an analysis at the two-digit level (usually between 30 and 50%, but as many as 69% for Cuba). In summary, Table 8.2 indicates how a quantity of networked occupations that would be hidden at one level of occupational detail are visible at another.

Table 8.3 shows similar results from data for Romania in 2002. It shows that the use of different levels of occupational detail has consequences for results. Whilst nearly all two-digit connections are replicated by at least some three-digit combinations, only 82% of those at the three-digit level are replicated at four-digit level. In this example, analysis at the more aggregate level may even suggest relationships that do not in fact exist after more suitable, fine-grained measures are used. Like Table 8.2, Table 8.3 shows that quite a large number of combinations are not evident at the more aggregated level but would be detected at the more disaggregate level.

Lastly, Fig. 8.9 shows images of networked occupations revealed by analysis at either the three-digit or four-digit level for Romania in 2002. Here the criteria for connections (amongst married or cohabiting couples) was that the combination must occur at least once in every 10,000 relationships, and it must occur at least twice as often as would be expected if the distribution of connections were random. For simplicity, the presentation for both levels is given where nodes reflect occupations at the three-digit level, but the connections between them are defined by whether links occurred between the nodes at the two respective levels of

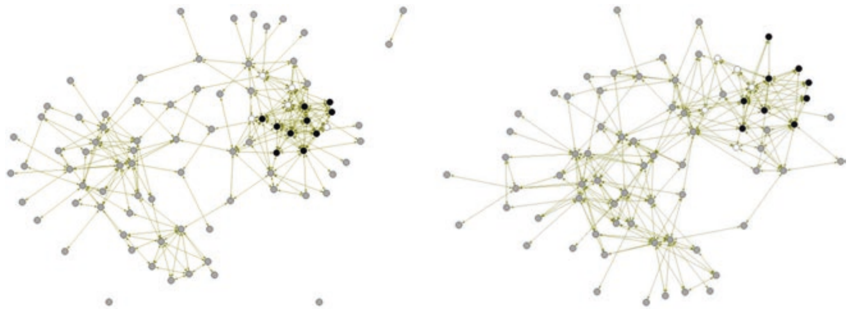


**Fig. 8.9** Romania 2002 network at three-digit ISCO level, analysed at three-digit (left) and four-digit (right) level. Source: Analysis of census data on husband-wife occupational connections accessed from IPUMS-I

detail.[4] Whilst the structures have similarities, there appears to be greater cohesion both amongst and between professionals (shaded black) and managers (white) when calculating connections at the three-digit level. For example, the network to the left shows connections between all ten professional minor groups, whilst only nine of these groups, with fewer mutual ties, are included in the network to the right. Moreover, managers appear, on the latter network, to bridge connections from professionals to other occupations, but this is not evident in the three-digit analysis. In this case, failure to use fine-grained occupational details would suggest that the most advantaged occupations are more strongly connected than might in fact be the case.

In summary, network analysis on social connections between occupations benefits from thought into what constitutes both a node and a tie—the story can readily be changed depending on choices that are made. When defining ties, it makes sense to report clearly upon the criteria used, to use similar criteria to other studies, and ideally to try out and compare different criteria in preliminary sensitivity analysis. When defining occupational nodes, a more refined occupational scheme will generally improve the chances of identifying consequential connections. Moreover, it seems likely that various 'non-standard' social mechanisms are more likely to be identified at the more disaggregated level (and elided when aggregations are used). Because of this, the impact of the level of detail is likely to play out in a complex, non-linear manner, rather than being a straightforward or predictable minor attenuation. Usually, therefore, it is compelling to work at the more disaggregated level of detail whenever reasonably possible.

### 8.5.3 Agency

The construction of sociograms for networked occupations raises some theoretical questions. For some writers, an SNA approach is usually associated with a form of agency on the part of the node, namely, to influence, or be influenced by, the structure of the network (e.g. Scott 2017).[5] Treating occupations as a 'node' within a network involves examining a collective grouping, but there is no sense of leadership or direction by one or more members of that group, so there is no obvious sense of agency inherent within occupational networks.

Ignoring agency is not problematic, but does raise some points of departure from conventional approaches to network analysis. For instance, many network studies explore who the most central actors are within the network—'centrality' is captured by various measures of the role and potential influence of actors within the network (Borgatti et al. 2013; de Nooy et al. 2011). In most contexts, measures of centrality are useful tools for identifying positions within the network—for instance, when studying employees linked by friendship, we could identify the most popular; those with the best access to all network members; or those with access to the most senior individuals. Within networked occupations, these concepts are largely irrelevant. To know that, for instance, teachers have the most ties, or can gain access to all other occupations with the fewest intermediaries, may not be substantially important. The former pattern might reflect artefactual features of how networked occupations and occupational units are defined. The links via intermediaries are probably inconsequential since the individuals behind the data have social ties that are not meaningfully affected by the aggregate profile of ties of the occupation as a whole.

## 8.6 Pseudo-Diagonals, Subsidiary Dimensions, and 'Catnets'

We have previously described 'pseudo-diagonals' as specific pairs of occupations that have above-average volumes of social interactions, due to specific, readily identified social mechanisms (e.g. Sect. 6.7). In the social interaction distance tradition, several mechanisms have been associated with pseudo-diagonal combinations (Sect. 6.7.2), the most prominent of which is 'situs', or shared workplace environment. In the SNA tradition, similar social mechanisms have sometimes been identified, but labelled differently. For instance, the idea of situs has a natural comparison to the concepts of 'propinquity' and 'foci'. In the network analysis literatures, 'propinquity' describes when social interactions are driven by shared membership of wider institutions and/or spaces. For instance, Marvin (1918) highlighted membership of the same occupation, and Wellman (1979) membership of residential communities, as factors that encourage interactions. Feld (1981) discusses the importance of 'foci' in forming social networks, namely, the physical spaces that individuals inhabit. Feld argues that individuals within the same foci are more likely to interact, that the size of the space will influence this tendency, and that individuals who inhabit multiple foci will have higher volumes of interactions (1981, p. 1026).

In the SID tradition, pseudo-diagonals are not the only consequence of conceiving that different social mechanisms can drive social interaction patterns. The dimension reduction tools of SID usually define multiple dimensional structures in the propensity to interact—for instance, the first dimension is usually presumed to reflect social stratification, but subsidiary dimensions might represent orthogonal forces such as of gender segregation or an urban/rural divide (see also Sect. 6.7 and Chap. 11). In the social network analysis tradition, a separate concept has been used to try to disentangle the influence of different social mechanisms on overall network structures—the concept of 'categories of networks' or 'catnets' (White 1992).

White (1992) argues that homophily in general is shaped by a wide range of interpersonal relationships. These 'catnets' can involve different dimensions of our lives—such as our occupation, home town, age, ethnicity, or favourite sports team. The connections brokered through different catnets will not necessarily overlap, and it is possible for different catnets to be more influential at different points in time for the same person. University students, for instance, could have catnets related to those living in the same residences, those on the same course, those who are members of the same societies, and to their friends from their home town.

White (1992) argues that people are more likely to form a connection to someone who is in multiple 'catnets', for instance, a student might be more likely to know people within their residences who are on the same course and sports team, and especially if on both. In terms of social connections between occupations, patterns in the most over-connected combinations might suggest similarities in multiple aspects of the lives amongst their incumbents.

Starting with census data for a purposive range of countries with comparable occupational data (Cuba in 2002, France in 1999, the Philippines in 2000, Portugal in 2001, South Africa in 2001, Thailand in 2000, and Venezuela in 2000), Fig. 8.10 shows the network of those occupational connections (at ISCO-88 three-digit level) that occurred, across countries, at least three times as frequently as expected (and occurred at least once in every 15,000 marriages). These can be thought of as very strong connections between occupations, because they appear to be replicated globally (across different countries), and they are very strongly connected in (at least three of) the countries in which the tie is observed.[6]

The networked occupations shown in Fig. 8.10 are labelled by their ISCO-88 three-digit code (see ILO 2010). ISCO-88 major groups 1–3 refer broadly to managerial, professional, and associate professional/technical occupations, so the network suggests a bonding of professional occupations at the centre of the sociogram, indicated by the black-shaded nodes. The specific codes suggest some pattern of clustering by industry, with educationalists to the left (e.g. 232, 233, 235) and business professionals (including business managers, shaded white) to the right (e.g. 123, 121, 122, 241). It also seems that a connection between these different groups of professionals may be 'brokered' by professional occupations in the legal, computing and health sectors (e.g. 213, 242, 222). On the

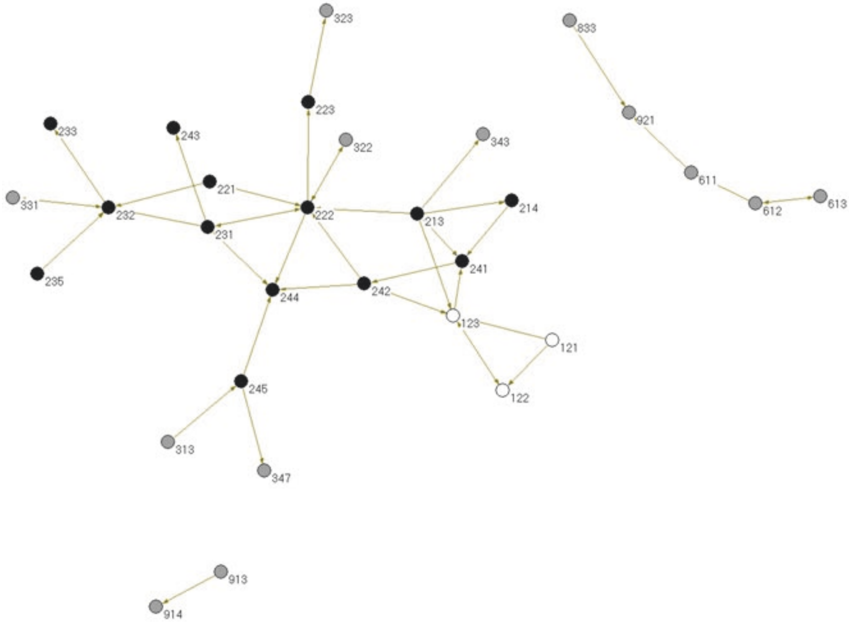



**Fig. 8.10** Networked occupations (ISCO-88 three-digit) that occur at least three times as often as anticipated in at least three countries. Source: Analysis of IPUMS-I datasets on husband-wife occupational combinations for Cuba 2002, France 1999, the Philippines 2000, Portugal 2001, South Africa 2001, Thailand 2000, and Venezuela 2000

whole, these patterns suggest that educational links might comprise one catnet that is a major force behind social connections within advantaged occupations. However, the 'brokering' occupations might reflect the influence of institutional catnets and propinquity (because these are occupations that can span different sectors and thus facilitate connections, perhaps through situs, at different institutions).

Many other linkages might reflect workplace ties. Connections between professional and associate professional or technical occupations might link computing professionals with administrative workers (213–343), school teachers with teaching assistants (232–331), nursing and midwifery professionals with associate professionals (223–323), health professionals with modern health professionals (222–322), and writers/creative artists to both artistic, entertainment or sports performers, and also optical equipment engineers (245–313/347). Aside from the latter tie, these are all instances of occupations which frequently work together within the same organisations. In addition, the two smaller components in Fig. 8.10 are similarly likely to reflect shared foci. One links domestics and cleaners with building caretakers and window cleaners (913–914). The other links a cluster of jobs in agriculture, including farmers (611, 612, 613), agricultural workers (921), and agricultural (and other) plant operatives (833). There is a suggestion that propinquity tends to be the main characteristic of those connections that bridge professional occupations across sectors, whilst educational backgrounds remain a good candidate for explaining most other patterns of differences in interaction patterns.

There are also occupations which do not appear to have strong ties to any other occupations. Table 8.4 shows those occupations which do not have any connections, in any country, above 1.5 times the expected rate.[7] First, the armed forces have no strong ties, for men or for women. It is possible that the diverse circumstances of those in the armed forces mean that their social connections may be spread evenly across a range of positions, without any particularly unusual associations.

Next, many of the occupational unit groups without strong ties are composite categories (labelled 'n.e.c'). These categories tend to involve a heterogeneous group of occupations, some of which may be ill-defined or not widely recognised. It is possible that jobs in these categories lack consistent identities or circumstances, but it is also plausible that there are

**Table 8.4** Occupations with no ties to any other occupation that occur more than 1.5 times as often as would be expected if ties were random

| | |
|---|---|
| Male | Armed forces; charcoal makers and related workers; concessionaires and loggers; other supervisors n.e.c.; sales and service elementary occupations n.e.c.; metal, machinery and related workers n.e.c.; other craft and related trades n.e.c.; armed forces; other associate professionals n.e.c. |
| Female | Armed forces; religious professionals; special education teaching associate professionals; other associate professionals n.e.c.; other craft and related trades n.e.c.; customer service clerks n.e.c. |

Source: As Fig. 8.10. 'n.e.c.' = 'not elsewhere classified' strong networking relationships amongst some of the specific occupations that underlie these categories, but that those patterns are elided by amalgamation.

There are four other occupations which fail to produce strong ties. Two are males who work in the forestry sector ('charcoal makers' and 'loggers'). Similarly to the armed forces, we suspect these occupations recruit incumbents from a range of circumstances and localities, perhaps on a short-term basis, meaning the groups do not form a distinctive profile in any of the categories that are usually influential. Two female roles without strong ties are religious workers and special education teaching associate professionals. Again, both categories seem likely to recruit from heterogeneous circumstances, so, again, none of their networks are sufficiently over-influential to generate over-represented ties to other occupations.

Indeed, some interaction patterns seem to reflect an influence that is close to but not exactly that of propinquity. They might reflect orientations, or behavioural propensities, of individuals. For example, it is plausible that shared values and attitudes might be fostered in educational institutions and/or workplaces (Christakis and Fowler 2010) and might account for empirical patterns of interactions that apply to broad occupational sectors rather than specific workplaces or professions. Shared 'orientations' might perhaps be found in healthcare work (a desire to assist others), in education (a belief in helping children), or the creative sector (a sense of expressionism). Other propensities might be linked to more practical aspects of occupations (for instance, the shift patterns of healthcare and protective services workers). In this example, the suggestion is of a category of networks associated with personal orientations and propensities, noting that the origins of those values may lie with educational and/or workplace characteristics.

In summary, those occupations that do not have unusually common social connections to other occupations seem to be characterised by heterogeneity. Inspection of those occupations that are defined as 'networked' suggests the interplay of several different categories to the network connections (or catnets). These include a general influence that might be connected to educational background, and more specific patterns that seem to reflect propinquity and foci, including connections that seem to be more common when occupations have a more well-known social identity. However, whilst the social network literatures tend to express how social circumstances lead to the formation of social connections, in the case of occupational patterns, we should also be careful to remember that the social connections of individuals could lead in turn to their occupational situation, for instance, by fostering recruitment or aspirations (e.g. Mouw 2003; Flap and Boxman 1999).

## 8.7 Practical Issues in Undertaking a Network Analysis of Occupations with Social Survey Data

There are several accessible texts on social network analysis which cover the common technical terms, and analytical methods, that are associated with the tradition (e.g. Scott 2017; Borgatti et al. 2013; de Nooy et al. 2011; Scott and Carrington 2011). The data used in a network analysis of occupational structure is the same as that used in a SID analysis (see Chap. 6). The key requirement is data on pairs of individuals who share a social connection and for whom the occupation is known. Examples include records on spouses' occupations from national censuses; individual and parental occupation from social surveys; or the occupation of respondents and their best friends from friendship surveys. Similar data can sometimes be obtained from administrative and by-product datasets.

Figure 8.11 shows illustrative software code (in Stata) that could be used to convert data from a large-scale social survey into a format suitable for an SNA analysis of networked occupations. It uses data from the 2011 Portuguese census (Minnesota Population Center 2015).

As in Fig. 6.1, Fig. 8.11 begins with a 'preliminary' section which is used to specify where relevant files are located for the specific computer in use. Segment (i) of Fig. 8.11 then illustrates the construction of a dataset featuring ego-alter pairs with data on both occupations. The code is actually the same as was used in Fig. 6.1. In this example, we select male cases with a valid occupation listed for ego and spouse, which creates a dataset where each row features a male occupation ('occ') and a female

```
****************************************************************
****************************************************************
**** Preliminary: Specifying locations/names for data and metadata files
global path1 "C:\camsis\countries\portugal\data\2011\" /* IPUMS-I downloaded dat file and do file */
global file4 "C:\data\resources\isco\labels\isco08_labels_2.do" /* value labels for ISCO-08 (www.camsis.stir.ac.uk) */
global path9 "c:\temp\" /* for temporary file storage */
****************************************************************


**** (i) Open source data from IPUMS-I: Portugal 2011

do $path1\ipumsi_00054.do /* downloaded from ipumsi: all Portugal 2011 sample  */
             /*  with sex, occupation of ego and their spouse  ('attach characteristics') */
tab1 occ occ_sp /* this is occupation of ego and alter, 3-digit ISCO-08, valid codes 11-962 */
keep if sex==1 & sex_sp==2  & occ >= 11 & occ <= 962 & occ_sp >= 11 & occ_sp <= 962
codebook occ occ_sp, compact /* 69k both-working heterosexual couples, 125 occ units  */
/* Acknowledgement:
  Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 6.4
     [Machine-readable database]. Minneapolis: University of Minnesota, 2015.
   The author wishes to acknowledge the statistical office that provided the underlying data
    making this research possible: National Institute of Statistics, Portugal.
*/
sav $path9\file1.dta, replace /* a temporary copy of the husband-wife microdata */
****************************************************************

**** (ii) Identify 'networked occupations' by calculating when occupational combinations are
**          over-represented by a certain threshold

use $path9\file1.dta, clear
rename occ hocc  /* standard label for male partner's occupation */
rename occ_sp wocc  /* standard label for female partner's occupation */
gen freq = 1
collapse (count) freq, by(hocc wocc) /* data is now in frequency table format */
egen tot=sum(freq) /* total cases across data */
egen nhocc=sum(freq), by(hocc) /* totals in male occupations */
egen nwocc=sum(freq), by(wocc) /* totals in female occupations */
gen phocc=nhocc/tot /* proportion of males in the job */
gen pwocc=nwocc/tot /* proportion of females in the job */
gen ewocc=pwocc*nhocc /* expected number in the h-w combination if connections were random */
gen value=freq/ewocc /* surplus between observations and occurrences ('representation ratio') */
gen prop=freq/tot /* proportion the combination occurs */
gen staner = sqrt((prop)*(1 - prop) / tot) /*  creates a standard error */
gen prop_min=prop-(1.96*staner) /* lower  95% CI  */
gen prop_max=prop+(1.96*staner) /* upper 95% CI  */
gen prop_exp=ewocc/tot /* expected proportion of combination  */
gen val_min=prop_min/prop_exp /* surplus of combinations, at lower level */
gen val_max=prop_max/prop_exp /* surplus of combinations, at higher level */

***label variables
label variable tot "total number in sample"
label variable nhocc "total number of males in occupation"
label variable nwocc "total number of females in occupation"
label variable phocc "percentage of men in occupation"
label variable pwocc "percentage of women in occupation"
label variable ewocc "expected number of partnerships"
label variable prop "Observed proportion of all ties"
label variable prop_exp "Expected proportion of all ties"
label variable prop_min "Lower observed proportion of all ties"
label variable prop_max  "Upper proportion of all ties"
label variable value "Observed value of representation ratio"
label variable val_min "Lower bound of observed value of representation ratio"
label variable val_max "Higher bound of observed value of representation ratio"

sav $path9\file2.dta, replace   /* temporary copy of the data file */
/*
* The same calculations can be generated directly using an online command file designed for this purpose:
 do http://www.camsis.stir.ac.uk/sonocs/do/pajek.do
*/
```

(*continued*)

```
****************************************************************
**** (iii) Exporting data from Stata, selecting ties according to the threshold approach
use $path9\file2.dta, clear
/* Selected threshold: combination probably occurs at least twice as often as would expect if connections
      were random  (i.e. lower bound of confidence interval for representation ratio exceeds 2) */
sum if val_min>=2 /* checks the data for those cases that make the selected threshold */
keep if val_min>=2 /* drops cases which do not meet the selected threshold */
keep hocc wocc freq /* reduces data to core edgelist content */
sav $path9\portugal_t1.dta, replace /* exports data in Stata format (keeping labels etc) */
outsheet  using"$path9\portugal_t1.txt", comma nonames nolabel replace
/* also exports the data as a text file (text file is convenient if other software is also to be used) */
****************************************************************

*** (iv) Exporting data from Stata, selecting ties according to the popularity approach
use $path9\file2.dta, clear
/* Selected threshold: combination is one of the three most common for the occupation, and
   it occurs at least 5 times in the data */
gsort +hocc -val_min /* sorts the data by occupation and descending order of the threshold */
bysort hocc: gen num=_n /* within occupations gives a rank to each ego-alter occupational permutation */
sum if num <= 3 & freq >= 5
keep if num <= 3   & freq >= 5  /* drops all combinations not within the 3 highest connections within occupations */
keep hocc wocc freq /* reduces data to core edgelist content */
sav $path9\portugal_t2.dta, replace
outsheet  using"$path9\portugal_t2.txt", comma nonames nolabel replace /* exports the data as a text file */
****************************************************

**** (v) Install Stata's 'nwcommands' library to define network structure
* (illustrated below is generic code that installs this extension library)
capture mkdir $path9\stata
capture mkdir $path9\stata\ado
adopath + "$path9\stata\ado"  /* code to ensure have somewhere suitable for local installation */
net from http://www.nwcommands.org
net set ado "$path9\stata\ado"
net install nwcommands-ado
****************************************************************

***** (vi) Use the 'nwcommands' extension to chart the network structures

* Example (1): If combination occurs at least twice as often as expected by chance
use $path9\portugal_t1.dta, clear
summarize
* Use NWcommands to define the network structure:
capture nwset, clear /* remove existing networks from memory if relevant */
nwset hocc wocc freq, name(occ1) edgelist undirected
nwsummarize
* Use NWcommands to show a sociogram of the structure:
nwplot(occ1), title("PT 2001: RR >= 2")   lab labelopt(mlabsize(tiny)) layout(mdsclassical) ///
      scatteropt(mfcolor(gs13) mlcolor(gs7) msymbol(circle)) scheme(s1mono)

* Example (2): Most popular combinations, regardless of RR value, so long as at least 5 instantiations
use $path9\portugal_t2.dta, clear
summarize
* Use NWcommands to define the network structure:
capture nwset, clear /* remove existing networks from memory if relevant */
nwset hocc wocc freq, name(occ1) edgelist undirected
nwsummarize
* Use NWcommands to show a sociogram of the structure (with slightly layout settings):
nwplot(occ1), title("PT 2001: Popularity threshold (up to 3 most popular)")  ///
   lab labelopt(mlabsize(tiny)) layout(mdsclassical) ///
      scatteropt(mfcolor(green*0.5) mlcolor(gs7) msymbol(circle)) scheme(s1mono)
****************************************************************

*** EOF
```

**Fig. 8.11** Software code (Stata format) for identifying and graphing 'networked occupations' occupation ('occ_sp'). For convenience, the syntax also deletes all cases where both partners are in the same job (this option impacts upon later results such as the values of the 'representation ratio' for each pair of occupations).

Section (ii) of Fig. 8.11 begins by converting the data from a list of ego-alter occupational combinations into 'table' format, a structure that will generate an 'edge list' where each row of the data describes a different permutation of occupations plus the number of cases for that permutation. For convenience, the two occupations are named as 'hocc' and 'wocc' (these are arbitrary names, based upon 'husband's occupation' and 'wife's occupation', but a small code file in Stata format that can automate this process is also available online[8] and requires the same naming conventions to be used). The product of the conversion, after the line that begins 'collapse …', is a dataset with just three variables: ego job ('hocc'), alter job ('wocc'), and the frequency of ties ('freq'). Network data is sometimes originally available in this ('edge list') format.

The remaining code in Section (ii) then calculates statistics that summarise the relative over-representation of the ties (the statistics that were described in Chap. 7). The total number of cases in the dataset is identified and then the number of people in each ego and alter occupation. The proportion of alters in each job is calculated and by multiplying this by the number of egos in each occupation, we can derive the expected frequency of ties between the occupations if the social connections between occupations were distributed randomly. Then, by dividing the observed frequency by this expected number, we create the 'representation ratio'.

The representation ratio gives an exact ratio between observed and expected numbers, but, as discussed in Chap. 7, there is a good case for calculating uncertainty statistics around this value, particularly because some occupational combinations may be represented by relatively few cases. Accordingly, a number of lines of the code are used to calculate 'standard error' statistics for the proportions that define the representation ratio, from which lower and upper limits to a 95% confidence interval for the representation ratio can be derived. Typically, it will be suitable to focus only on those over-represented ties for which the lower bound of the confidence interval exceeds the relevant threshold.

Sections (iii) and (iv) of Fig. 8.11 both export relevant data on the network for use in a network software tool. In general with network analysis, we summarise all records within a given network, so the key decision with occupational networks concerns which of the occupational combinations we will include in the analysis. Section (iii) is one instantiation of the 'threshold' approach to selection (it pulls out only those ties for which the lower bound on the confidence interval for the representation ratio is more than 2, and saves a database of only those ties). Section (iv) is an alternative instantiation—it illustrates selecting cases using the 'popularity' method, that is, by selecting only the most popular tie for each different occupation. In applied research, we would typically consider various alternative selection criteria and undertake multiple network analyses for each plausible dataset. In addition, although it is not illustrated in Fig. 8.11, we might also export other contextual data about either the occupations, or the tie, to the derived datasets, and this might be used in the network analysis package (e.g. we might colour code a sociogram based on the selected ties, with shading defined according to, say, the CAMSIS score for the occupation).

Sections (v) and (vi) of Fig. 8.11 concern using a software tool to visualise the networks that have been derived. We have put examples of relevant software code for doing this in three different packages—Stata, R, and Pajek—on the CAMSIS webpages. The code in section (v) installs the extension package 'nwcommands' in Stata (Grund 2014), and the code in section (vi) illustrates two examples that generate sociograms using that package (there are many variations to the presentation that we might alternatively have used). For elaboration of relevant examples both using Stata and using other packages, see the CAMSIS webpages.

## Notes

1. Data on the social connections between occupations might be deliberately treated as undirected ties, for instance, if there is no important distinction between the profile of ego and of alter ties (for instance, data on friend one and friend two), and alternatively if there is a meaningful direction between the ties, but there is felt to be no benefit to incorporating it in the analytical summary.

2. Based on Current Population Survey estimates as reported by http://kff.org/other/state-indicator/distribution-by-raceethnicity/ [accessed 1/12/16].
3. We use the 'macroclass' scheme associated with Weeden and Grusky's (2012) microclass scheme.
4. Therefore, a tie between two minor groups in the right diagram means at least one of the four-digit unit groups within the first three-digit minor group has a tie to at least one of the four-digit unit groups within the other three-digit minor group.
5. There are other counterexamples of non-agentic network nodes. For example, semantic networks explore the network properties of words with texts (Steyvers and Tenenbaum 2005); analyses of consumption data have treated the products of consumption as network nodes (e.g. Krebs 2008); Bellotti and Mora (2016) analyse data from a consumer survey by operationalising survey response variables as nodes.
6. Across the seven nations, the occupational combinations that are highlighted are at least three times over-represented in 74% of cases, and at least twice as over-represented in 92% of cases.
7. Not all occupations were included in each country. Additionally, combinations are only identified if they occur at least once in every 15,000 between-occupation marriages—some occupations are involved in combinations that are over-represented, but do not occur sufficiently frequently to satisfy the 1-in-15,000 criteria.
8. See www.camsis.stir.ac.uk/sonocs/.

## Bibliography

Bellotti, E., & Mora, E. (2016). Networks of Practices in Critical Consumption. *Journal of Consumer Culture, 16*(3), 718–760.

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analysing Social Networks*. London: Sage.

Christakis, N., & Fowler, J. (2010). *Connected: The Amazing Power of Social Networks and How They Shape Our Lives*. London: Harper Press.

Crossley, N. (2015). *Networks of Sound, Style and Subversion: The Punk and Post-Punk Worlds of Manchester, London, Liverpool and Sheffield, 1975–1980*. Manchester: Manchester University Press.

de Nooy, W., Mrvar, A., & Batagelj, V. (2011). *Exploratory Social Network Analysis with Pajek*. Cambridge: Cambridge University Press.

Feld, S. (1981). The Focused Organization of Social Ties. *American Journal of Sociology, 86*(5), 1015–1035.

Flap, H., & Boxman, E. (1999). Getting a Job as a Manager. In R. T. A. J. Leenders & S. M. Gabbay (Eds.), *Corporate Social Capital and Liability* (pp. 197–216). Boston: Springer US.

Ganzeboom, H. B. G. (2005). On the Cost of Being Crude: A Comparison of Detailed and Coarse Occupational Coding. In J. H. P. Hoffmeyer-Zlotnik & J. Harkness (Eds.), *Methodological Aspects in Cross-National Research* (pp. 241–257). Mannheim: ZUMA, Nachrichten Spezial.

Giuffre, K. (2013). *Communities and Networks: Using Social Network Analysis to Rethink Urban and Community Studies*. Cambridge: Polity Press.

Griffiths, D., & Lambert, P. S. (2012). Dimensions and Boundaries: Comparative Analysis of Occupational Structures Using Social Network and Social Interaction Distance Analysis. *Sociological Research Online, 17*(2), 5.

Griffiths, D., & Lambert, P. S. (2013). The American Occupational Marriage Structure, 1980 and 2010. *Procedia—Social and Behavioural Sciences, 100*, 21–39.

Grund, T. U. (2014). *nwcommands: Software Tools for the Statistical Modeling of Network Data in Stata*. Retrieved March 1, 2017, from http://nwcommands.org.

Hanneman, R. A., & Riddle, M. (2011). Concepts and Measures for Basic Network Analysis. In J. Scott & P. J. Carrington (Eds.), *The Sage Handbook of Social Network Analysis* (pp. 340–369). London: Sage.

Kadushin, C. (2012). *Understanding Social Networks: Theories, Concepts, and Findings*. Oxford: Oxford University Press.

Knoke, D. (2012). *Economic Networks*. Cambridge: Polity Press.

Knoke, D., & Yang, S. (2008). *Social Network Analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Krebs, V. (2008). *A Network of Books About Recent US Politics Sold by the Online Bookseller Amazon.com*. Retrieved March 1, 2017, from http://www.orgnet.com/divided.html.

Lin, N. (1999). Social Networks and Status Attainment. *Annual Review of Sociology, 25*, 467–488.

Lin, N., Ensel, W. M., & Vaughn, J. C. (1981). Social Resources, Strength of Ties and Occupational Status Attainment. *American Sociological Review, 46*(4), 393–405.

Marin, A., & Wellman, B. (2011). Social Network Analysis: An Introduction. In J. Scott & P. J. Carrington (Eds.), *The Sage Handbook of Social Network Analysis* (pp. 11–25). London: Sage.

Marsden, P. V. (2011). Survey Methods for Network Data. In J. Scott & P. J. Carrington (Eds.), *The Sage Handbook of Social Network Analysis* (pp. 370–388). London: Sage.

Marvin, D. M. (1918). Occupational Propinquity as a Factor in Marriage Selection. *Journal of the American Statistical Association, 16*(123), 131–150.

Minnesota Population Center. (2015). *Integrated Public Use Microdata Series, International: Version 6.4 [Machine-Readable Database]*. Minneapolis, MN: University of Minnesota.

Mouw, T. (2003). Social Capital and Finding a Job: Do Contacts Matter? *American Sociological Review, 68*(6), 868–898.

Scott, J. (2017). *Social Network Analysis: A Handbook* (4th ed.). London: Sage.

Scott, J., & Carrington, P. J. (Eds.). (2011). *The Sage Handbook of Social Network Analysis*. London: Sage.

Steyvers, M., & Tenenbaum, J. B. (2005). The Large Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science, 29*(1), 41–78.

Weeden, K. A., & Grusky, D. B. (2012). The Three Worlds of Inequality. *American Journal of Sociology, 117*(6), 1723–1785.

Wellman, B. (1979). The Community Question. *American Journal of Sociology, 84*, 1201–1231.

Wellman, B., & Berkowitz, S. D. (Eds.). (1988). *Social Structures: A Network Approach*. Cambridge: Cambridge University Press.

White, H. C. (1992). *Identity and Control: A Structural Theory of Social Action*. Princeton, NJ: Princeton University Press.

Wilkinson, R. G., & Pickett, K. E. (2009). *The Spirit Level: Why More Equal Societies Almost Always do Better*. London: Allen Lane.

# 9

# Occupational-Level Residuals and Distributional Parameters

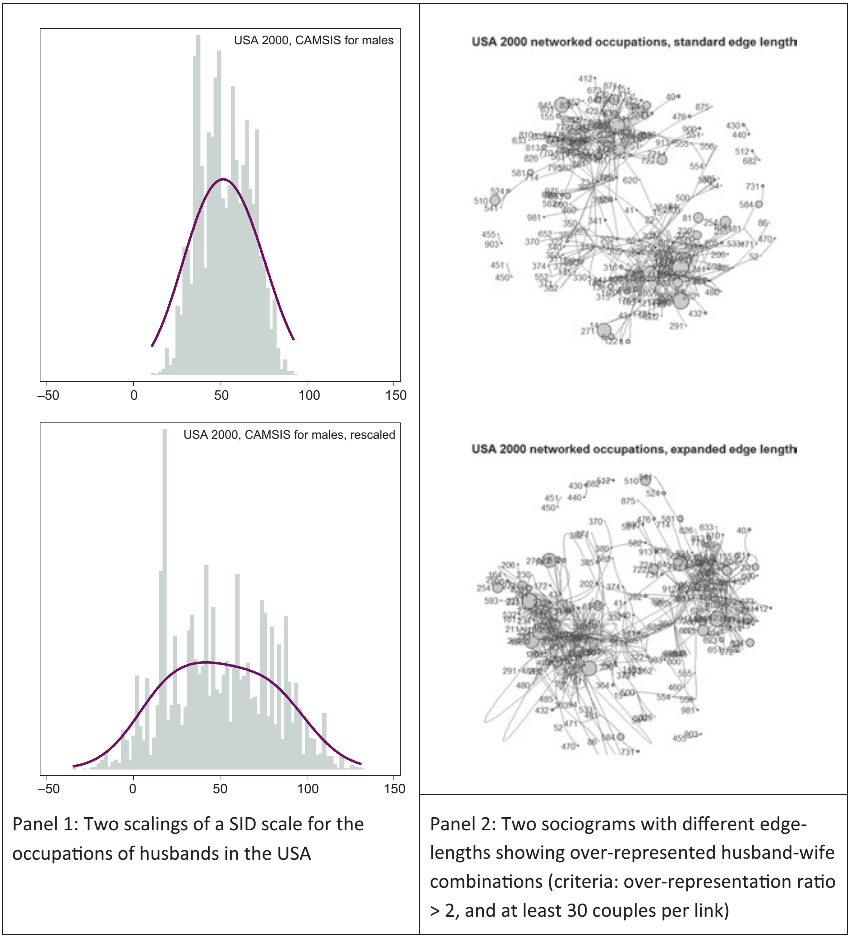
## 9.1 Introduction

In this chapter, we cover two further options in the statistical analysis of social interactions between occupations. One concerns the extent to which distributional parameters about the wider social structure can be used to provide alternative summary statistics concerned with social distance and occupational inequalities (Sect. 9.2). The other concerns scenarios where it can be useful to use 'random effects' models to explore occupation-to-occupation variations in relevant outcomes (Sect. 9.3).

## 9.2 Rescaling the Social Resin Using Relevant Distributional Parameters

Our previous descriptions used the metaphor the 'social resin' to describe how the social relations between people serve to bind together occupations of similar positions in the stratification structure. This resin helps to forge the structure of social inequality itself, and we claim that it is a mechanism at the heart of many important aspects of our social experiences. Most of our descriptions have involved the placement of social positions within a


P. Lambert, D. Griffiths, *Social Inequalities and Occupational Stratification*,
https://doi.org/10.1057/978-1-137-02253-0_9




Panel 1: Two scalings of a SID scale for the occupations of husbands in the USA

Panel 2: Two sociograms with different edge-lengths showing over-represented husband-wife combinations (criteria: over-representation ratio > 2, and at least 30 couples per link)

**Fig. 9.1** Depictions of the structure of social relations between occupations with different scaling values

structure defined by the social distance between occupations. We have drawn dimensional maps and sociograms that illustrate those placements, but the actual physical differences within those maps have in fact been standardised and have not usually been interpreted further. Figure 9.1 illustrates this, with plots of the structures of social relations between occupations for the USA in 2000, as defined through the distribution of the

CAMSIS scale scores (on the left), and through a network sociogram for the most over-represented husband-wife occupational combinations or 'networked occupations' (on the right). In both examples, the upper row shows the conventional scaling we recommend using for these approaches (the CAMSIS scores are scaled to mean 50, standard deviation 15; the sociograms are drawn with the 'optimal' distance between nodes and length of edges for visual examination as identified by a standard graphical algorithm). However, the lower row shows the same data with some rescaling to the positions—the CAMSIS scores are rescaled to mean 50, standard deviation 30; the sociogram is redrawn in such a way that the average edge lengths are forced to be longer. A change of scaling could change the emphasis of some results, though we would not ordinarily expect it to be hugely consequential to the depiction, as seems to be the pattern in Fig. 9.1 (much the same hierarchical ranking of occupations is evident in either scaling of the CAMSIS scales; much the same links between the same occupations are discernible in each sociogram). It is apparent, however, that the optimal means of scaling in both scenarios could be debated. Hitherto we have depicted the actual distances within the 'social resin' arbitrarily, but could there be a useful way to make them non-arbitrary?

Summary statistics that might be used as scaling parameters could come from various sources. Data on social connections themselves could provide summary parameters, for instance, on the relative density, depth, or immediacy of social relationships between occupations. Alternatively we could use other macro-level summary statistics for the same purpose, such as those national-level social indicators that are widely used in comparative research literatures. For example, we could obtain a Gini coefficient as an income inequality measure for the relevant society and use that to rescale, say, the standard deviation of the relevant CAMSIS scale (cf. Liao 2006). This example would have the effect of narrowing the range of CAMSIS scale scores in more equal countries and widening it in less equal countries (compare to Fig. 9.2). A wide range of summary statistics linked to social stratification and inequality are available that might be considered for this or a similar purpose (cf. Atkinson 2005; Atkinson and Brandolini 2006). Indeed, it seems to us that many studies that make use of social inequality measures (such as CAMSIS scales, ISEI scales, and social class measures) might gain from exploring whether the particular

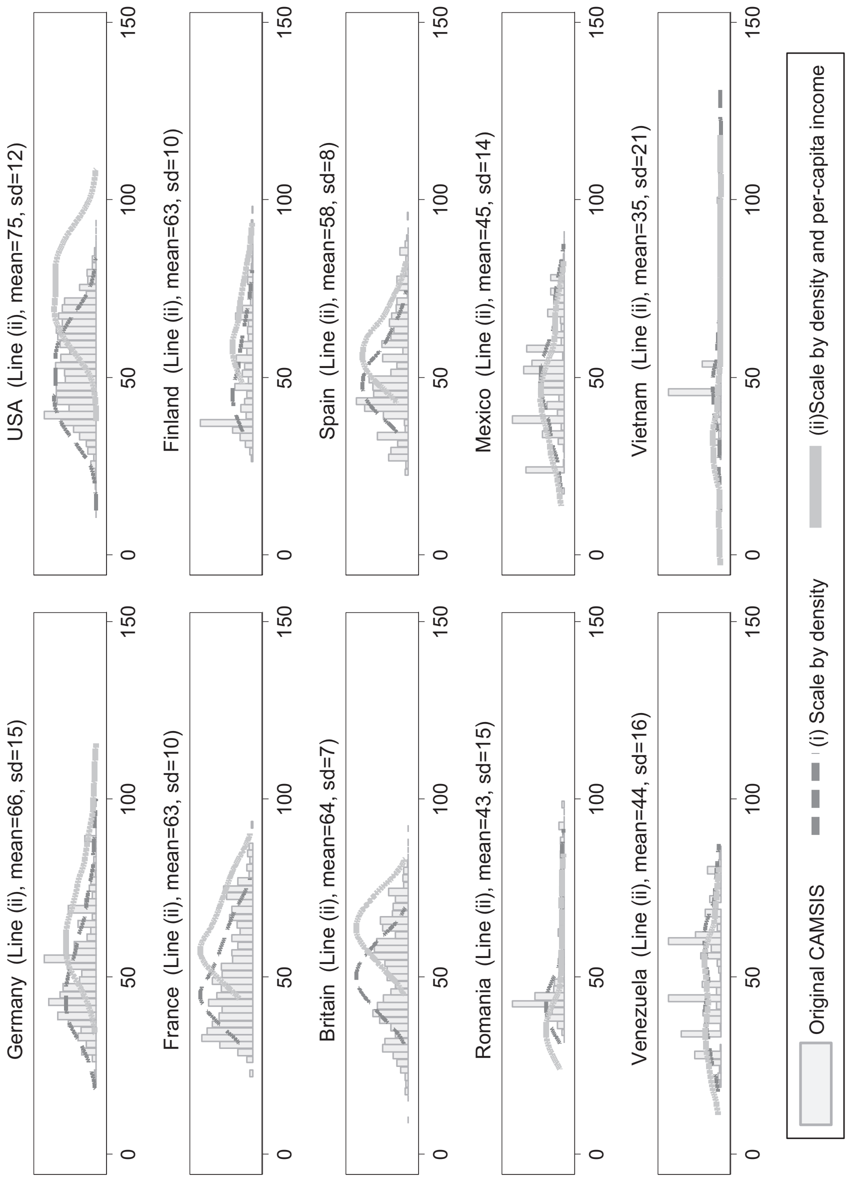



**Fig. 9.2** Re-parameterisations of CAMSIS scales according to country-specific features. Source: Microdata sources as Fig. 5.1

comparisons of interest might be enhanced by a well-justified rescaling on the basis of relevant macro-level indicators. However, hereafter we focus only on summary statistics that are themselves derived through data on social relationships and social connections.

### 9.2.1 Summary Statistics Based Upon SID and on Networked Occupations

Summary statistics based upon social connections and interactions could be expected to offer original evidence about the structuring of social relationships around occupations from one society to another, and the patterns obtained might not necessarily be equivalent to the patterns evident from other summary measures about society, such as income-based inequality measures. Several summary statistics are readily generated as a by-product of the calculations that we have already described under the social interaction distance approach (Chaps. 4, 5, and 6), or the popular algorithms of social network analysis (Chaps. 7 and 8), and we will discuss their relevance below.[1]

Turning first to social interaction distance patterns, calculations proceed by seeking to identify dimensional structures that help improve the prediction of empirical observations of social interactions between units (here occupations). In doing so, the relative empirical contribution of each dimension can be ranked and summarised, whilst overall model fit statistics, given the dimensions, can also be obtained. These summary statistics contrast to the scores given to specific occupations in the first dimension (the results that we have previously concentrated upon). In Table 9.1, the upper two panels show some illustrative examples of these statistics for microdata on male-female couples.

One complication is that conventional presentations use different terminology and different functional forms for the summary measures, depending on the nature of the technique used to calculate the dimensions. When using correspondence analysis, it is common to present the 'eigenvalue', 'singular value', or 'principal inertia' associated with a given dimension, in each case providing a measure that indicates the relative proportion of variance in the interaction patterns that can be attributed

**Table 9.1** Standard summary statistics from social interaction distance analysis and social network analysis approaches

| | Husband-wife occupational associations for data from … | | |
|---|---|---|---|
| | USA 2000 (excluding diagonals) | France 2006 (all adults) | France 2006 (excluding farming pseudo-diagonals) |
| Number of occupational units | 25 | 31 | 31 |
| Number of cases analysed | 610,004 | 1,909,822 | 1,889,142 |
| *Correspondence analysis* | | | |
| Total inertia (Cramer's V) | 0.15 (0.08) | 1.34 (0.21) | 0.78 (0.16) |
| Dimension 1 principal inertia (%) | 36.8 | 23.3 | 32.0 |
| Dimension 1 subjective interpretation | Stratification | Farming | Stratification |
| Dimension 2: inertia (%) and interpretation[a] | 16.8 (sales jobs) | 18.3 (stratification) | 12.7 (shopkeepers) |
| Dimension 3: inertia (%) and interpretation[a] | 13.3 (management) | 15.4 (farm work) | 8.6 (teaching) |
| Dimension 4: inertia (%) and interpretation[a] | 9.3 (transport) | 7.3 (shopkeepers) | 6.1 (public sector) |
| *RC-II association models* | | | |
| Dissimilarity index for: | | | |
| Dimension 2 model | 0.068 | 0.095 | 0.093 |
| Dimension 1 model | 0.084 | 0.125 | 0.115 |
| Independence model | 0.142 | 0.235 | 0.225 |
| Dimension 1 percentage of association | 29.8 | 62.3 | 61.4 |
| Dimension 1 interpretation[a] | Stratification | Farming | Stratification |
| Dimension 2 % association and interpretation[a] | 14.5 (management) | 16.2 (stratification) | 16.8 (?sector) |

(*continued*)

**Table 9.1** (continued)

| | Husband-wife occupational associations for data from … | | |
|---|---|---|---|
| | USA 2000 (excluding diagonals) | France 2006 (all adults) | France 2006 (excluding farming pseudo-diagonals) |
| *Social network analysis for 'networked occupations'* | | | |
| Number of over-represented ties | 36 | 119 | 116 |
| Number of couples being analysed | 104,678 | 538,810 | 515,435 |
| Density (proportion of possible ties that are networked) | 0.060 | 0.128 | 0.125 |
| Index of group degree centralization (variation in centrality for male occs) | 0.688 | 0.598 | 0.597 |
| Indicator of variation: mean gap in CAMSIS scores for networked jobs | 5.02 | 5.32 | 5.33 |

Source: Analysis of census data on occupations of cohabiting couples obtained from IPUMS-I (MPC 2015)

[a]Subjective interpretations of the fields of work that appear to drive the differentiation between high and low scores on the respective dimension

to each dimension. (Recall that the CAMSIS scale itself is usually calculated by extracting the most important dimension—that with the largest principal inertia—from the correspondence analysis solution.) It is also common to calculate a measure known as the 'total inertia', which summarises the overall relationship between rows and columns in the table being analysed through correspondence analysis—in broad terms this indicates the extent to which the row distributions vary across different columns or vice versa and is in fact directly related to the well-known association statistic Cramer's V (see Greenacre and Blasius 1994).

Alternatively, statistics with similar interpretations can be calculated based upon log-linear association models, but conventionally these use a different terminology and different calculations (e.g. Wong 2010). In this framework, the relative explanatory contribution of the dimensions can be captured by model fit comparisons between related models: the relative influence of each dimension upon the pattern of association is summarised through the proportionate decline in model deviance (using the 'log-likelihood chi-square' statistic) that results from adding each respective dimension in comparison to the independence model (e.g. Wong 2010, p. 101). Overall fit can be indicated with summary coefficients based upon the independence model likelihood—the 'dissimilarity index', for instance, is commonly reported as an indicator of the proportion of empirical patterns in the table that represent divergence from independence.

In the top two panels of Table 9.1, summary statistics are given from a social interaction distance analysis of husband-wife occupational associations in two societies when (for convenience) a relatively low number of occupational units are analysed. From the correspondence analysis, we see a lower total inertia (expressed in standardised form through the Cramer's V statistic) in the USA compared to France—i.e. husband-wife occupational distributions as a whole are more closely interlinked in France. Within countries, however, we see a relatively higher dimension 1 principal inertia in the USA than in France, which suggests that in that society, the dimension of stratification provides relatively more of an explanation of the association between spouses' jobs, when compared to any other factors. By comparison, in France relatively more of the marriage associations related to occupations can be linked to subsidiary dimensional structures (see also Chap. 11). The interpretation of the statistics from the association model results (panel 2 of Table 9.1) is more challenging given the way that the statistics are derived, but the same conclusion emerges. For the association model results, the USA's smaller dissimilarity index tells us that husband-wife occupational distributions are less strongly associated with each other in that country than France (the dissimilarity index is often interpreted as reflecting how many cases would have to be switched around within the table to achieve a pattern of independence between husbands and wives). On the face of it, the dimensional summary statistics for the association model results seem to tell a different story to the correspondence analysis percentages because the values reported here are larger in France than in the USA. However, the nuance of interpretation in this instance is that these percentages tell us about how much of the explained association can be attributed to each structure in each society. In addition, the table incorporates (columns 2 and 3) comparisons between two different analyses for France, one with all of the population included and a second where the so-called farming pseudo-diagonals (i.e. husband-wife combinations when both are in jobs related to farming) are excluded. This comparison reveals that substantially different dimensional structures can be identified depending upon whether or not small numbers of influential combinations are included in an analysis.

There are some operational challenges to using these statistics to make comparisons between societies. As in the production of other aggregate-level statistics about occupations, such as on gender segregation (cf. Charles and Grusky 2004, p. 146), the values obtained are shaped by the distribution of cases from occupation to occupation, but there is no consistent way in which they can adjust for peculiarities of the occupational taxonomies from one society to another. For instance, if in one society all occupations in teaching are coded into the same occupational unit group, but in another there are 12 different categories encompassing different teaching specialisms, then clearly the distributional statistics that summarise the overall occupational patterns in each society could have incomparabilities. In such scenarios, it seems impossible to resolve whether the differences that we see between the USA and France in Table 9.1 are due to genuine differences between the two countries in how social relationships are organised around occupations, or whether they are simply a product of the different occupational classification schemes in each society. Put differently, the statistics might only provide effective comparisons if the underlying occupational unit group schemes were equivalent in each society, a scenario that is rarely likely to be possible. A further difficulty concerns the way in which 'diagonal' or 'pseudo-diagonal' combinations of jobs are treated (see Sect. 6.8). As illustrated in Table 9.1, whether or not occupations are excluded from the analysis on the basis of being regarded as diagonals or pseudo-diagonals can result in some changes to the model summary statistics.

A number of summary statistics can also be generated when we define 'networked occupations' or when we summarise patterns in networked occupations through the descriptive tools of social network analysis. Indeed, whilst network analysis methods are well known for their graphical summaries, there are a wide range of standard statistics that can usefully inform us about the character of a network (e.g. Knoke and Yang 2008, c4). However, not all of the commonly used network statistics make good sense in the example of summarising social connections between occupations. In Table 9.1, for example, we list a 'density' statistic (which summarises how many of the possible connections between nodes are represented in the network), and a 'centrality' statistic, which is related to whether nodes have connections to many different occupations, and how much that varies between nodes. On the face of it, these values might provide interesting information on differences between societies, but the comparison breaks down because the links that we are describing were themselves defined in terms of the number of cases that connect occupations (i.e. we defined ties by a criteria that means it is impossible for all ties to occur simultaneously). The range of values that network summary statistics could take is therefore constrained in a complex way that will be related to the distribution of individuals in occupations; in summary we would not advocate using these values further due to lack of comparability between contexts.

Other simple macrodata about the distribution of networked occupations might be usefully generated, however. In Table 9.1, we report the total number of 'overrepresented ties' (i.e. pairs of networked occupations), which gives us tangible information about the social structure of occupations for the specific society, and we report the mean gap in CAMSIS scores between those pairs of occupations that are defined as 'networked occupations', which provides a simple characterisation of the association pattern (for instance, the value in France is slightly higher than that for the USA, suggesting that unusually common occupational connections in that country are marginally more diverse in their character than in the USA). Both of these statistics offer some information about the character of the society in itself, and might reasonably be used as scaling parameters in depicting other aspects of the network structure. For instance, we might consider setting the length of ties to be proportional to the mean gap in CAMSIS scores between ties as a way of conveying the closer association in one society compared to another.

In summary, the various statistics given in Table 9.1 offer some additional characterisations of the social interaction structure that might be useful in making comparisons between societies. However, they have an important limitation because the values of the statistics can be influenced by the occupational coding frames, and by operational decisions about inclusion or exclusion of cases. In our observation, such factors don't usually have a major impact upon the core results of a SID or SNA analysis of occupational structure (e.g. the main dimensional pattern, or network structure interpretation), but they do impact more substantially upon supplementary results, such as the summary statistics shown in Table 9.1.

### 9.2.2 Exploiting Distributional Parameters

When calculating CAMSIS scores, we usually standardise the scores to a mean of 50 and standard deviation of 15 in a nationally representative sample. Accordingly, in any given society, the score given to a certain occupation is indicative of its placement relative to the national structure. Standardisation of measures within a society is common practice in comparative research, yet there may be times when it is less desirable. In absolute terms, both the mean and the standard deviation of a stratification measure could legitimately be different from society to society: one country might on average be more socially advantaged than another; and/or the dispersion of social inequality within one society might be greater or less than that of another. Indeed, both of these scenarios are self-evident in stratification research—there is tremendous variation between societies in levels of social advantage (e.g. Atkinson 2005; OECD 2004); there is also considerable variation between societies in the scale of internal inequalities (esp. Wilkinson and Pickett 2009). Should depictions of social inequality such as CAMSIS scales seek to represent these wider differences between societies?

A conservative approach is to conceive of CAMSIS measures as focussed strictly upon internal differentiations within a society. We might, for example, run regression models over a range of societies and examine the coefficients for the effect of CAMSIS scale score for people upon an outcome of interest. As one example, Jarman et al. (2012) report the relationship between CAMSIS scale scores and occupational gender segregation patterns in a variety of countries—this provides a legitimate comparison of the connection between stratification and gender segregation from one country to another.

A different mode of comparison could be obtained however if we rescale CAMSIS scores according to additional parameters held about the society. We could rescale using a mean parameter, a dispersion parameter, or both. A plausible dispersion parameter is the intra-cluster correlation of occupations in terms of their social relationships (see Sect. 9.3). A mean parameter might be based upon the markers of inertia and homogamy described above in Table 9.1, or it might be based upon externally generated statistics. Figure 9.2 then shows indicative histograms for CAMSIS scores for 10 societies in their original units then again with alternative scalings (the first scaling only reflects dispersion, based on an ICC statistic; the second scaling reflects both dispersion, using an ICC statistics, and mean variation, using a measure of per capita income (in 2004) taken from the CIA World Factbook[2]).

From these figures, we can potentially put a different interpretation upon social inequality from society to society. In the USA, for instance, the density adjustment suggests a widening of the distribution to exacerbate the gap between more and less advantaged positions, whilst the per capita income adjustment suggests a movement of the distribution to the right (relatively more people in relatively more advantaged occupations). Both of these measures make sense intuitively—another way to think of this would be that looking at the second scale for the USA, a low score on that scale would mean an occupational position that is disadvantaged internally, say 2 standard deviations below the mean, but is in a medium relative position when compared to several countries, and is towards the top of the distribution when compared to Vietnam and Romania. The most advantaged of all, according to this data, are those with the highest CAMSIS scores in the USA and Germany (typically, professors and medical doctors), but also those with the highest CAMSIS scores in Vietnam (including politicians), reflecting the higher dispersion in that society. The most disadvantaged are those with the lowest CAMSIS scores in Vietnam (farm labourers and rubbish collectors), but also those in similar positions in Venezuela and Romania. These descriptions are certainly plausible global portrayals of social inequality.

In such examples, data from the analysis of social interactions can potentially be used 'twice over' to help us study and understand social inequality. Firstly, it can help us to characterise the internal distribution of inequality in any particular society (as also in Chaps. 4, 5, 6, 7, and 8), but secondly it can also provide standardisation parameters which could be used to rescale the representation of inequality and provide promising comparative analytical tools.

## 9.3 Exploring Occupational-Level Residuals with Random Effects

A rather different device for using occupational data to inform characterisations about the 'social resin' can be achieved by using multilevel models with data on occupations. The popular social statistics tradition of multilevel modelling (e.g. Rabe-Hesketh and Skrondal 2012; Hox 2010; Gelman and Hill 2007; Goldstein 2003) is concerned with analysing data when records are grouped (or 'nested' or 'clustered') into different categories of a structure of interest. A common example is in educational research, where the data may comprise the grades of individual pupils, but pupils are clustered into different classes. Multilevel models generically can be thought of as any sort of statistical model that features parameters or adjustments that reflect the clustering of cases. 'Random effects' multilevel models, on which we concentrate here, involve estimating how the model error terms can be decomposed, between components that are related to the cluster, and others that reflect individual differences. If we think of individual level records as 'clustered' into occupations, random effects models can give us a range of useful data about occupations and influences that are related to them (e.g. Bol and Weeden 2015; Mills 2007).

A basic formulation of a 'two-level' multilevel model is shown in Eq. (9.1), where '*j*' indexes a 'higher level' or 'level 2' unit (such as a school, or occupation), and '*i*' indexes individual cases at the 'lower level' or 'level

1' unit (such as individual pupils, or employees). The formulation resembles that of other regression models, where the outcome $Y$ is predicted as a function of values of a number of nominated explanatory variables (indicated by the $\beta X$ term). There is then an additional gap between the model predictions ($\beta X$) and the outcome, the 'error', which in the random effects model, is allowed for by two (or more) further terms, connected to the different 'levels'. In Eq. (9.1), there are two error terms, an individual error represented by '$\varepsilon$' and a cluster-specific error represented by '$\mu$'. The term '$\mu_j$' can be through of intuitively as an additional term that is applied to individual cases if they belong to the relevant unit j.

$$Y_{ij} = \beta X_{ij} + \mu_j + \varepsilon_{ij} \tag{9.1}$$

The way in which the $\mu_j$ parameters in (9.1) will be calculated is open to two possibilities. A specific coefficient could be calculated for each higher-level unit (e.g. each of the different occupations). If this happens, we say that we calculate a 'fixed parameter' for each unit '$j$', and this amounts to what is often called the 'fixed effects' hierarchical model (cf. Allison 2009). Alternatively, a 'random effects' means of estimating model (9.1) is often used. Rather than calculate parameters for each cluster unit, the random effects analysis proceeds by modelling the overall ('random') variation that is associated with empirical differences from cluster to cluster—in practice, this is achieved by estimating a variance parameter for the $\mu_j$ terms.[3] It is also worth stressing that the errors represented by '$\mu_j$' can be thought of as residual variation that is net of the patterns captured by the average effects of the other predictor variables '$\beta X$'—that is, in the multilevel modelling scenario, we will usually be talking about the residual character of different clusters (e.g. occupations), net of overall population-level patterns that are captured in the '$\beta X$' formulation.

Random effects models for data where records are clustered into occupations offer an appealing way to explore the 'multilevel' influences of both occupational and individual level factors upon an outcome (cf. Snijders and Bosker 2012). Nevertheless, some discussions of multilevel modelling suggest that random effects models are only appropriate if there is a relatively large number of cluster units, and they can be regarded as a sample from a wider population of clusters (e.g. Rabe-Hesketh and Skrondal 2012, pp. 93–97; Snijders and Bosker 2012). Working with occupational data, a researcher may have anywhere between 5 and 500 categories, and these clusters (i.e. different occupations) would normally be a complete list of categories (rather than a sample of occupations from a wider distribution). However, there are different positions in the methodological literature, and several authors recommend using a random effects modelling approach even for data that features only a small number of clusters and/or for categories that might represent the full set of available categories (cf. Schmidt-Catran and Fairbrother 2016; Gelman and Hill 2007, pp. 244–246). Hitherto, random effects models have been constructively used for occupational data in a handful of applications (e.g. Bol and Weeden 2015; Lui and Grusky 2013; Mills 2007), although the approach is not yet particularly common. By contrast the 'fixed effects' model for detailed occupational categories is also used occasionally (e.g. Stansfeld et al. 2011), and the use of 'fixed effects' is standard practice when analysis concerns the influence of a small number of aggregate occupational units, such as dummy variables representing social class categories. However, we have stressed elsewhere that broad aggregations of occupations risk ignoring interesting heterogeneity. With many different occupational categories, the fixed effects approach has limitations because it is not possible to identify additional parameters that summarise the influence of other occupational-level variables, whilst there is also some risk that inferences about specific occupations have low sample power (since some occupations are likely to be represented by only a small number of cases within the relevant dataset). Inferences about occupations that are based upon the random effects model, on the other hand, benefit from the property of 'shrinkage'. This means that statistics at the occupational level can be calculated in a way that is influenced both by the individual cases within the occupation, and by data on the wider distribution of occupation-to-occupation variations (e.g. Mills 2007).

The specification shown in (9.1) can be extended in some potentially interesting ways. When random effects are used, the formulation of (9.1) represents a version of the model known as the 'random intercepts' model, but this can be extended to the 'random coefficients' or 'random slopes' multilevel model, which allows the occupational-level error variation to itself be related to the values of other explanatory variables (in a similar style to an interaction effect). Random coefficients models are easier to conceptualise when the '$\beta X$' term is broken down into distinctive $X$ variables—in (9.2), we show a scenario involving three measured explanatory variables ($X_1$, $X_2$, and $X_3$), alongside an indicative explanatory variable '$X_0$' that serves to illustrate the constant term (i.e. it takes the value 1 for every case in the dataset).[4] Model (9.2) therefore allows for the impact of the explanatory variables $X_1$ and $X_3$, as well as the intercept term, to vary from occupation to occupation.

$$\begin{aligned} Y_{ij} = {} & \beta_0 X_{0ij} + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3ij} + X_{0ij}\mu_{0j} + X_{1ij}\mu_{1j} \\ & + X_{3ij}\mu_{3j} + X_{0ij}\varepsilon_{ij} \end{aligned} \tag{9.2}$$

There are at least three benefits to modelling occupations with random effects. First, the models will generate useful statistical parameters about the magnitude of occupational effects that would not otherwise be recorded—for instance, 'intra-cluster correlation' statistics can describe the relative magnitude of occupational and individual patterns in the error terms, and assessments of 'random coefficients' parameters provide a systematic means of testing the relationship between occupational-level variations and other explanatory factors. Second, random effects models provide tools that allow us to subsequently calculate refined estimates about the properties of each occupation and to explore occupational structure through those terms. Lastly, the statistical fit of models with hierarchical random effects for occupations is ordinarily superior to that of models without these effects, and improved model fit should in turn lead to improved model parameters for other terms, such as more appropriate standard errors. Indeed, there is an interesting question of whether statistical models which use occupation-based measures, such as social class measures, should *ordinarily* include random effects for individual occupations—in other scenarios, it is common to argue that plausible random effects terms should always be added in this situation, for instance, to ensure appropriate standard errors for higher-level (occupation-based) explanatory variables (e.g. Schmidt-Catran and Fairbrother 2016; Snijders and Bosker 2012). Some of these attractions are generic across application areas, but some, as discussed below, are specifically relevant to analysis of social distance and social contact patterns involving occupations.

### 9.3.1 Illustrating Random Effects Models for Occupations: Occupations and Social Contact Patterns

Table 9.2 shows results from a statistical model with random effects for occupations. The data, from the UK's British Household Panel Survey (University of Essex 2010), refers to individuals who are in work and who have spouses who are also in work. The outcome is a measure based upon self-reported frequency of contact with the three people that respondents nominated as their closest friends (a score is derived based on answers to a series of relevant questions found on the BHPS; the higher the score, the less frequent were reported contacts). When the 'social isolation' score is explored through regression models (e.g. model (2)), we see significant average net effects associated with gender (females on average have more regular contact with their closest friends), age (a curvilinear pattern through the life course), educational qualifications (those with degrees or diplomas have on average less regular contact), and occupational advantage (more advantaged jobs, both of the individual and of their spouse, are independently associated with less frequent contact with close friends). All of these patterns seem plausible, though in total they only account for about 13% of the variance in the outcome (suggesting that most of the forces that shape contact with friends are not included effectively in the model).

Fitting random effects for occupations, as shown in models (3) to (7) of Table 9.2, allows us to explore variation in responses that is not captured by the regression predictor variables but that might be linked to occupational inequalities. Across the table we see that all of the models that use random effects improve upon the fit compared to corresponding models without them (evident in reduced log-likelihood and 'BIC' statis-

**Table 9.2** Example of statistical models for a social distance outcome, using random effects for occupational units

| Predictors of index score for level of contact with three closest friends (higher values = less frequent contact) | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Linear model | | Random effects for own job | | | Random effects for spouse's job | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Intercept | 13.47 (0.02) | 7.22 (0.18) | 13.56 (0.08) | 7.55 (0.20) | 7.61 (0.22) | 13.29 (0.08) | 7.59 (0.20) |
| Female | | −0.780 (0.034) | | −0.691 (0.043) | −0.724 (0.043) | | −0.842 (0.043) |
| Age in years (coeff * 100) | | 4.89 (0.70) | | 4.82 (0.71) | 4.40 (0.75) | | 4.59 (0.71) |
| Age-squared (coeff * 1000) | | −0.299 (0.071) | | −0.293 (0.072) | −0.262 (0.074) | | −0.268 (0.072) |
| Has degree or diploma | | 0.695 (0.038) | | 0.638 (0.039) | 0.645 (0.039) | | 0.673 (0.038) |
| Own job CAMSIS (coeff * 10) | | 0.498 (0.014) | | 0.467 (0.024) | 0.482 (0.023) | | 0.482 (0.014) |
| Spouse's job CAMSIS (coeff * 10) | | 0.421 (0.013) | | 0.403 (0.013) | 0.402 (0.013) | | 0.393 (0.021) |
| Level 1 variance | 13.87 (0.09) | 12.11 (0.08) | 12.38 (0.08) | 11.90 (0.08) | 11.81 (0.08) | 12.65 (0.09) | 11.95 (0.08) |
| Level 2 variance (intercept variance) | | | 2.13 (0.19) | 0.320 (0.056) | 1.40 (0.28) | 1.73 (0.16) | 0.235 (0.035) |
| Level 2 variance with age | | | | | 0.000 (0.000) | | |

(*continued*)

**Table 9.2** (continued)

Predictors of index score for level of contact with three closest friends (higher values = less frequent contact)

| | Linear model | | Random effects for own job | | | Random effects for spouse's job | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Level 2 age-intercept covariance | | | | | −0.029 (0.006) | | |
| Level 2 ICC | – | – | 0.147 | 0.026 | – | 0.120 | 0.019 |
| Log-likelihood | −119,343 | −116,424 | −117,295 | −116,219 | −116,164 | −117,734 | −116,286 |
| BIC | 238,707 | 232,934 | 234,622 | 232,534 | 232,445 | 235,499 | 232,667 |
| Level 1 *r*-2 | | 0.13 | | 0.04 | | | 0.06 |
| *N* | 43,655 | 43,655 | 43,655 | 43,655 | 43,655 | 43,655 | 43,655 |

tics), and they also lead to small changes—improvements—in the parameter estimates and standard errors calculated for the other explanatory variables. Indeed, the 'Level 2 ICC' that is reported for the random intercepts models shows us that a modest proportion of variation in the outcome can be linked to occupations alone (i.e. 12–15% for the models with no explanatory variables; 2–3% net of the influence of the explanatory variables). There is also an unusual extension to Table 9.2, insofar as we have estimated random effects models both for clustering into the respondents' own occupations (models 3, 4 and 5), and also for clustering in the occupations of the respondents' spouses (model 6 and 7). In this latter scenario, we are asking whether there is evidence that levels of contact are empirically related to the job held by the spouse. This is confirmed to be the case (even after controlling for the average social advantage associated with the occupation, as captured by the CAMSIS measures). These models allow us to test and confirm the hypothesis that the spouse's job matters, which in this case provides a further insight into the social processes linked to variations in social contact.[5] Lastly, model (5) of Table 9.1 allows for a random coefficient specification for the influence of the respondent's job. In this model, the way in which age effects the outcome is allowed to differ from occupation to occupation. This interaction is empirically validated, evident from the improvement in model fit between models (4) and (5). The interpretation is that the impact of age upon social contacts tends to be greater in some occupations than in others (we could subsequently use the model residuals to identify which occupations, though this information isn't evident from Table 9.2 alone).

Figure 9.3 illustrates another contribution that we can get from fitting random effects for occupations, namely, the capacity to explore specific occupational residuals. Figure 9.3 shows what in random effects modelling is usually known as a 'caterpillar plot', namely, a depiction of the occupational residuals and their estimated standard errors arranged in rank order. The 'occupational residuals' refer to estimated values for the terms '$\mu_j$' described earlier—that is, they are indicative of how much above or below the average that particular occupation is, in terms of outcome (net of explicitly modelled explanatory variables). We are analysing influences upon self-reported volume of contact with the three people

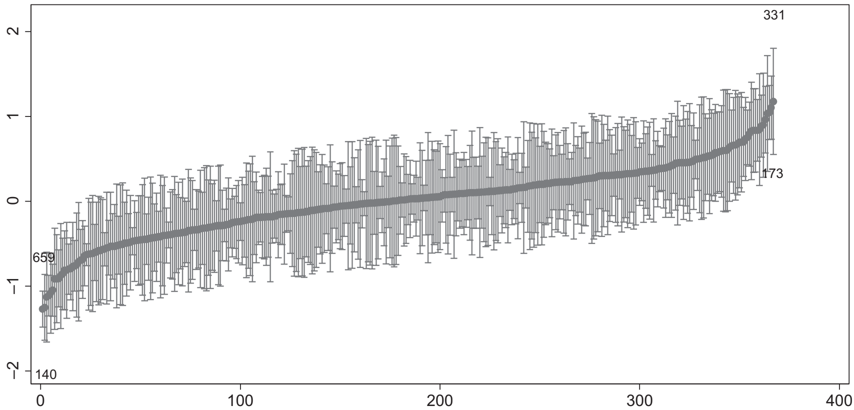



**Fig. 9.3** Job-level residuals (predicting level of contact with closest friends). Source: UK BHPS. *N* = 40k records for cohabiting adults in the UK, 1991–2008

listed as the respondent's three closest friends, using the same model as described in Table 9.2 (the graph uses model (4) in that table). The caterpillar plot, showing the average residuals associated with occupations net of the overall patterns captured by the explanatory variables, illustrates that several occupations as the extremes of the graph have markedly different patterns in the outcome. Those towards the top end, all other things being equal, have individuals who report greater distance from and less contact with their closest friends, and those towards the bottom end are those that stand out as having disproportionately more contact with their closest friends.

The residuals shown (known as the 'Empirical Bayes' residuals in random effects frameworks) are generally thought to provide robust, appropriate indicators about the specific occupations. In this example, we can helpfully look up those occupations with the highest and lowest residuals. The five that have the lowest contact compared to what would be expected by the model are 'Personnel and training managers', 'Hotel and accommodation managers', 'Aircraft flight officers', 'Rail transport inspectors and guards', and 'Farm workers'. All of these are, arguably, occupations that geographically constrain their incumbents, perhaps sundering existing personal ties. On the other hand, the five occupations that have more frequent contact than expected by the model are 'Transport managers',

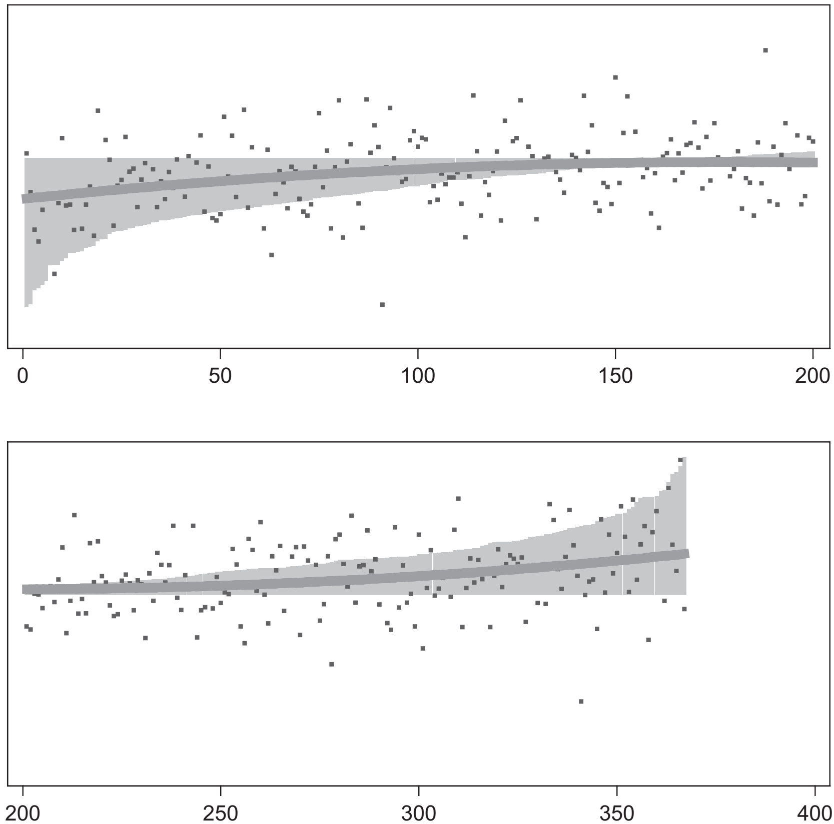



**Fig. 9.4** Job-level residuals (bars), and spouses' job residuals (points) for level of contact with up to three closest friends. Source: UK BHPS. *N* = 40k records for cohabiting adults in the UK, 1991–2008

'Pharmacists', 'Shoe repairers, leather cutters and sewers', 'Other childcare and related [e.g. childminders]', and 'Counterhands/catering assistants'. These are all arguably examples of occupations that often recruit locally or are embedded within local communities.

Figure 9.4 shows the same residuals in a different format, with bars for the height of the residuals (ignoring the standard error calculation), and points that depict the corresponding residuals in the jobs for spouses rather than for the individual. It is noticeable that there is only a weak correlation between the residuals for own job and the residuals for the spouse's job—this suggests that different processes are typically at work when the spouse's job, rather than own job, impacts upon social contact patterns.

### 9.3.2 Illustrating Random Effects Models for Occupations: Aspects of Social Inequality in Britain

Tables 9.3 and 9.4 summarise some other interesting findings that can be drawn from statistical models with random effects for occupations. Still using UK data, we are now looking at particular occupational titles (using the administrative categories of the UK's Standard Occupational Classification 1990—see OPCS 1990) that stand out as having unusually high or low 'random effects', across an illustrative selection of different models.

The random intercepts model (used in Table 9.3) serves to represent those patterns of variation in the outcome that link to occupations but aren't already captured by the regression model predictors. In these examples, we have models predicting the outcome (a score for subjective health, a measure of personal income, and a score for the amount of housework undertaken), with controls for a range of predictor variables (listed in Table 9.3). The distribution of 'random intercepts' therefore tells us about the extent to which occupations stand out as being above or below average on the outcome, net of controls for the explanatory factors in the model (whilst the 'shrinkage' property of the random effects estimates helps make appropriate inferences for those occupations that are represented by lower numbers of cases). For example, we see that 'shelf fillers' have disproportionately better health than average, net of explanatory factors. This pattern might reflect that this occupation features physical activity albeit in a relatively safe environment.[6] In fact, many more interesting insights on occupational inequalities might be available from these random effects estimates—for example, we have only highlighted the extreme residuals, when we could alternatively have presented all occupations, or perhaps have compared a handful of the same illustrative occupations across all models. Indeed, not only does Table 9.3 suggest

**Table 9.3** Selected occupations in the UK with extreme 'random intercepts' for occupations

| | |
|---|---|
| *… (1) Extreme intercepts in predicting health (net of controls for age, gender, educational level, marital status, occupation-based CAMSIS score and their interactions) …* | |
| **More healthy:** Financial and office managers; engineering technicians; scientific technicians; personnel officers; filing, computer clerks; stores, despatch clerks; armed forces, NCOs and other ranks; educational assistants; construction operatives; shelf fillers; other miscellaneous occupations | **Less healthy:** Publicans, innkeepers; managers and proprietors in services; university and polytechnic teachers; architects; social workers, probation officers; authors, writers, journalists; sewing machinists; face-trained coalmining workers; chefs, cooks, hotel supervisors; care assistants and attendants; launderers, dry cleaners, pressers; market and street traders; telephone salespersons; kitchen porters, hands; cleaners, domestics |
| **Wives' jobs with most healthy husbands:** Civil service executive officers; library assistants; nursery nurses; other personal and protective services; retail cash desk and check-out operators; other farming occupations; other miscellaneous occupations | **Wives' jobs with least healthy husbands:** Publicans, innkeepers; social workers, probation officers; organisation and methods and work s; sewing machinists; security guards; care assistants; packers, bottlers, canners, fillers; counterhands, catering assistants |
| **Husbands' jobs with most healthy wives:** Managers in mining and energy; bank managers; other financial and office managers; transport managers; farm owners and managers; other managers and administrators; other scientific technicians; builders, building contractors; carpenters and joiners; school caretakers; construction operatives; farmworkers; other miscellaneous occupations | **Husbands' jobs with least healthy wives:** Software engineers; artists, commercial artists; warehousemen/women; computer engineers; moulders, core makers, die casters; fishmongers, poultry dressers; bar staff; care assistants; plastic process operatives, moulder; bus and coach drivers; transport machinery operative |

(*continued*)

**Table 9.3** (continued)

| | |
|---|---|
| *… (2) Extreme intercepts in predicting income (same controls as model (1)) …* | |
| **Higher income:** Managers, large companies; managers in mining and energy; financial managers; purchasing managers; bank mangers; police officers (sergeant and below) | **Lower income:** Higher and further education teachers; other teaching professionals; clergy; professional athletes, sports officials; waiters, waitresses; bar staff; playgroup leaders; other childcare occupations; sales assistants; cash desk and check-out operators; collectors and credit agents; market and street traders; messengers, couriers; kitchen porters; shelf fillers; cleaners, domestics |
| *… (3) Extreme intercepts in predicting volume of household chores undertaken (same controls as model (1)) …* | |
| **Chores (more housework) (3%):** Social workers, probation officers; nurses; face-trained coalmining workers; chefs, cooks, hotel supervisors; hospital ward assistants; other childcare occupations; kitchen porters | **Chores (less housework) (3%):** Farm owners and managers; garage managers and proprietors; roofers, slaters, tilers, sheeters; plasterers; vehicle body repairers, panel beaters; coach painters; mechanical plant drivers; farmworkers (livestock) |

Source: Analysis of the BHPS, 1991–2008

some interesting patterns of variation in outcomes that are related to individuals' own occupations, but we also see some potential for exploring the influence of the occupations held by other relevant individuals—husbands who are married to a publican, for instance, apparently have worse health that would otherwise be predicted, but the wives of financial managers seem to do better than might otherwise be expected given their individual circumstances.

Table 9.4 explores 'random coefficients' or 'random slopes' patterns for the same individual-level processes. As they are used here, random coefficients allow for the scenario where the particular occupation itself mediates the effect of the incorporated explanatory variables. Such terms are often particularly interesting from a substantive viewpoint, since they resemble 'interaction effects' between a variable and the occupational patterns. For example, Table 9.4 summarises those occupations which, according to the random coefficients model, had unusually high or

**Table 9.4** Selected occupations in the UK with extreme 'random slopes' for occupations

| | |
|---|---|
| *… (4) Variations in the influence of age upon health …* | |
| **Markedly lesser negative effect of age upon health:** Marketing and sales managers; financial managers; other managers and administrators; filing, computer clerks; clerks; secretaries, personal assistants; other construction trades; care assistants; educational assistants | **Markedly greater negative effect of age upon health:** Hotel and accommodation managers; social workers, probation officers; nurses; face-trained coalmining workers; chefs, cooks, hotel supervisors; waiters, waitresses; food, drink and tobacco process operatives |
| *… (5) Variations in the influence of age, gender, and education upon income …* | |
| **Markedly greater positive effect of age upon income:** University and polytechnic teachers; special education teachers; waiters, waitresses; messengers, couriers; other occupations in sales and services | **Markedly lesser positive effect of age upon income:** Property and estate managers; authors, writers, journalists; artists, commercial artists; typists; roofers, slaters, tilers, sheeters; electricians; nursery nurses; merchandisers; weighers, graders, sorters |
| **Markedly greater negative effect of being female upon income:** Professional athletes, sports officials; playgroup leaders; other childcare occupations; sales assistants; collectors and credit agents; market and street traders; messengers, couriers; cleaners, domestics | **Markedly lesser negative effect of being female upon income:** Managers, large companies; managers in mining and energy; organisation and methods managers; computer systems managers; bank managers; other associate professionals; police officers (sergeant and below) |
| **Markedly greater positive effect of education upon income:** Local government officers; production and works managers; financial managers; primary school teachers; social workers, probation officers; nurses; actors, entertainers, stage managers; metalworkers and fitters; motor mechanics, auto engineers; wholesale sales reps; other miscellaneous occupations | **Markedly lesser positive effect of education upon income:** Managers and proprietors in service; clerks; warehousemen/women; secretaries; construction trades; shoe repairers; bakers; cash desk and check-out operators; assemblers/line workers (electrical); postal workers; kitchen porters; shelf fillers; cleaners, domestics |

(*continued*)

**Table 9.4** (continued)

| | |
|---|---|
| *… (6) Variations in the influence of gender upon volume of household chores undertaken …* | |
| **Markedly greater positive effect of being female upon volume of chores:** Company secretaries; farm owners and managers; garage managers and proprietors; other secretaries; roofers, slaters, tilers, sheeters; plasterers; mechanical plant drivers and operators; farmworkers; | **Markedly lesser positive effect of being female upon volume of chores:** Personnel managers ; restaurant and catering managers; university teachers; higher and further education teachers; school inspectors and advisors; social workers, probation officers; storekeepers, warehousemen/women; face-trained coalmining workers; customs and excise officers; hospital ward assistants; ambulance staff; care assistants and attendants old; bus and coach drivers |

Source: Analysis of the BHPS, 1991–2008. Models as Table 9.3 with additional 'random slope' terms

unusually low values of the relevant regression parameter. In the upper-left panel of Table 9.4, we see that 'educational assistants' are one of a number of jobs that are found to have a relatively smaller negative effect of age upon health (i.e. a smaller gap in health between older and younger individuals); in the lower-level panel, 'university teachers' are one example of an occupation where the impact of being female upon the volume of household chores undertaken is not as great as average. There are numerous other apparently interesting trends—older waitresses seem to be less healthy than younger ones, but earn more money; education doesn't matter much for cleaners' wages, but gender does; clerks don't age too badly, and educational background is less of a barrier. Such interpretations obviously need to be followed up with further evaluation and critical reflection, but it seems obvious to us that applying random effects models to individual processes and their relations with occupations raise fruitful descriptive opportunities.

### 9.3.3 Variance in Social Relations by Occupations Across Societies

The examples above included some scenarios where the processes being described were themselves a facet of the 'social resin', in the sense of the relationship between social inequality, occupations, and social connections between people. However, the models are specific to the particular data context—it is unlikely, for instance, that precisely the same occupations would be identified in a reanalysis from another time point or country, nor that other model parameters would stay largely consistent through different application areas. A more generic means through which occupation-level random effects can be used to describe features of the social resin is to undertake a comparative analysis of different societies with similar levels of occupational details, and report differences in occupational-level summary statistics that are related to the circumstances of social contacts such as friends or family. Figure 9.5 does this for data on the occupations of spouses, using the European Social Survey.

The analysis shown in Fig. 9.5 summarises the influence of the spouse's job (and educational qualifications) on the individual's outcome (the occupational, educational, or health score). On the left of Fig. 9.5, conventional model fit statistics ($r^2$ values) are shown: the bars show the increase in variance explained by incorporating a direct measure of spouse's position (either the CAMSIS score of the spouse's job or an education-based score) when predicting the outcomes (respondent's occupational score, respondent's education, and respondent's health). These figures provide data on variation from country to country in the average influence of spouse's circumstances—they effectively provide an index of homophily with regard to the strength of the spouse's influence. Generally speaking, large correlations are seen with spouses for the occupational and educational measures, but there is negligible additional effect of spouse's occupational score upon health outcomes (net of the effect of own occupational score). Nevertheless, in line with other studies (e.g. Smits 2003; Smits et al. 1999), we see substantial national variation in measured spousal influences, suggesting considerable differences in the extent of husband-wife occupational homophily. Broadly speaking, the average correlations between occupation or education of one spouse and another suggest low correlations in Scandinavian countries, medium to low correlations in most other northern and western European countries, and higher correlations in many of the southern and eastern European countries. These general trends are reasonably similar to those seen in other national reviews of homogamy (e.g. Smits et al. 1999).[7]

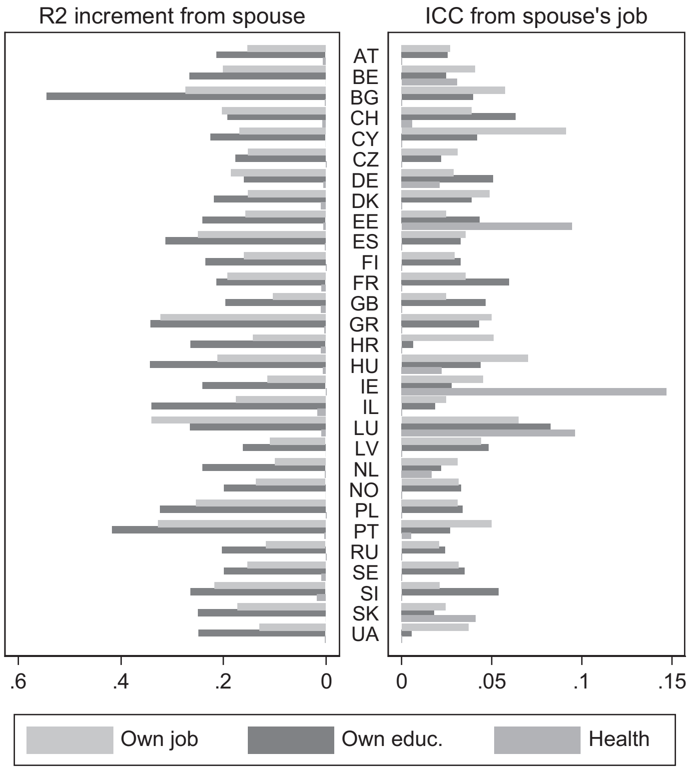



**Fig. 9.5** National-level variations in the influence of the spouse's occupation. Notes: Accumulated data from rounds 1 (2002) to 5 (2010) of the European Social Survey (coverage per year varies by country). Minimum sample size per country of 500 cases (with valid data on own and spouse's occupation). ICCs refer to variance component associated with the spouses occupation (four-digit ISCO88) when predicting the outcome (with controls for gender, age, and own and spouse's occupational CAMSIS score). $R^2$ increment refers to change in $r^2$ between model predicting the outcome with controls for gender and age and if relevant own occupational CAMSIS score, and model with addition of explanatory variable for spouse's occupational CAMSIS or educational score

On the right of Fig. 9.5, we show the relative proportion of error variance in the outcome that can be associated with clustering of responses around the spouse's occupation (net of the influences captured by the model explanatory variables). This makes for an alternative and interesting statistical summary of the relative influence of spouses. The earlier correlations between own and spouses jobs (or qualifications) are contingent upon the functional form of the occupation-based classification used (here the linear correlation between CAMSIS scale scores), but the ICC statistics might capture the 'upper limit' of spousal influence through occupations—that is, these values might proxy for additional differences associated with occupations, including those that are not well represented by the non-linear features of occupational associations (such as disproportionate patterns linked to sectoral affinities). The right panel of Fig. 9.5 suggests that the evidence on national variations in these additional occupational structures will reveal different patterns to those seen from more traditional methods (such as the left panel). That is, by analysing influences of occupational detail more rigorously, we may well draw different conclusions about social variations in the interplay between social connections and social inequality.

## Notes

1. We don't give examples, but we could also construct 'post-hoc' summary statistics about the social interaction patterns for a similar purpose (for instance, the correlation between the derived SID scale and a direct measure of, say, income).
2. Data from https://www.cia.gov/library/publications/the-world-factbook/rankorder/2004rank.html (Accessed 1 August 2013).
3. Generally speaking, 'fixed effects' models are best applied to scenarios where interest focusses clearly upon determinants of variations in an outcome given the categories of the higher-level unit (e.g. Allison 2009), whereas random effects models are better suited to describing social processes that feature a combination of higher-level and lower-level mechanisms (e.g. Rabe-Hesketh and Skrondal 2012, p. 92).
4. For completeness, the formulation also shows how the intercept term is allowed to vary in its impact from cluster to cluster. This term features in the simpler 'random intercepts' model (9.1) but is not normally written explicitly.
5. It is also feasible to fit a 'cross-classified' model with random effects for both own job and spouse's job, but this is not shown in the table.

6. This pattern might also arise if the occupation requires physical fitness as an entry criteron. However, there are few barriers to entry to this relatively disadvantaged occupation, so we suspect that it is plausible that this occupation is genuinely good for people's health (net of other factors).
7. A convenience of these statistics, which will also apply to the intra-cluster correlation statistics for occupations in the right panel of Fig. 9.5, is a natural comparability in scaling for the influence of linear measures and random effects terms for occupations. This contrast with the challenges of presenting comparable statistics based upon non-linear occupation-based class measure (cf. Smits 2003; Smits et al. 1999; Luijkx 1994, c6).

# Bibliography

Allison, P. D. (2009). *Fixed Effects Regression Models*. Thousand Oaks, CA: Sage.

Atkinson, A. B. (2005). Social Indicators, Policy, and Measuring Progress. In S. Svallfors (Ed.), *Analyzing Inequality* (pp. 83–107). Stanford, CA: Stanford University Press.

Atkinson, A. B., & Brandolini, A. (2006). The Panel-of-Countries Approach to Explaining Income Inequality: An Interdisciplinary Research Agenda. In S. L. Morgan, D. B. Grusky, & G. S. Fields (Eds.), *Mobility and Inequality* (pp. 400–448). Stanford, CA: Stanford University Press.

Bol, T., & Weeden, K. A. (2015). Occupational Closure and Wage Inequality in Germany and the United Kingdom. *European Sociological Review, 31*(3), 354–369.

Charles, M., & Grusky, D. B. (2004). *Occupational Ghettos: The Worldwide Segregation of Women and Men*. Stanford, CA: Stanford University Press.

Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Goldstein, H. (2003). *Multilevel Statistical Models* (3rd ed.). London: Arnold.

Greenacre, M., & Blasius, J. (1994). *Correspondence Analysis in the Social Sciences: Recent Developments and Applications*. London: Academic Press.

Hox, J. (2010). *Multilevel Analysis* (2nd ed.). London: Routledge.

Jarman, J., Blackburn, R. M., & Racko, G. (2012). The Dimensions of Occupational Gender Segregation in Industrial Countries. *Sociology, 46*(6), 1003–1019.

Knoke, D., & Yang, S. (2008). *Social Network Analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Liao, T. F. (2006). Measuring and Analyzing Class Inequality with the Gini Index Informed by Model-Based Clustering. *Sociological Methodology, 36*(1), 201–224.
Lui, Y., & Grusky, D. B. (2013). The Payoff to Skill in the Third Industrial Revolution. *American Journal of Sociology, 118*(5), 1330–1374.
Luijkx, R. (1994). *Comparative Loglinear Analyses of Social Mobility and Heterogamy*. Tilburg: Tilburg University Press.
Mills, C. (2007). Unequal Jobs: Job Quality and job Satisfaction. In P. McGovern, S. Hill, C. Mills, & M. White (Eds.), *Market, Class and Employment* (pp. 233–282). Oxford: Oxford University Press.
Minnesota Population Center. (2015). *Integrated Public Use Microdata Series, International: Version 6.4 [Machine-Readable Database]*. Minneapolis, MN: University of Minnesota.
OECD. (2004). *OECD Handbook for Internationally Comparative Educational Statistics*. Paris: Organisation for Economic Cooperation and Development.
OPCS. (1990). *Standard Occupational Classification, Volume 1: Structure and Definition of Major, Minor and Unit Groups*. London: Office for Population Censuses and Surveys.
Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata, Third Edition (Two-Volume Set)*. College Station, TX: Stata Press.
Schmidt-Catran, A. W., & Fairbrother, M. (2016). The Random Effects in Multilevel Models: Getting Them Wrong and Getting Them Right. *European Sociological Review, 32*(1), 23–38.
Smits, J. (2003). Social Closure Among the Higher Educated: Trends in Educational Homogamy in 55 Countries. *Social Science Research, 32*(2), 251–277.
Smits, J., Ultee, W., & Lammers, J. (1999). Occupational Homogamy in Eight Countries of the European Union, 1975–89. *Acta Sociologica, 42*(1), 55–68.
Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (2nd ed.). London: Sage.
Stansfeld, S. A., Rasul, F., Head, J., & Singleton, N. (2011). Occupation and Mental Health in a National UK Survey. *Social Psychiatry and Psychiatric Epidemiology, 46*(2), 101–110.
University of Essex, & Institute for Social and Economic Research. (2010, July). *British Household Panel Survey: Waves 1–18, 1991–2009 [Computer File]* (7th ed.). Colchester, Essex: UK Data Archive [distributor], SN: 5151.
Wilkinson, R. G., & Pickett, K. E. (2009). *The Spirit Level: Why More Equal Societies Almost Always do Better*. London: Allen Lane.
Wong, R. S. K. (2010). *Association Models*. Los Angeles: Sage.

# 10

# Social Interactions and Educational Inequality

## 10.1 Introduction

Hitherto we have focussed upon occupational positions, but in this chapter we explore the extent to which data on educational experiences can be introduced into the analyses of social interactions and social stratification. We explore whether we might get a different and more compelling result if we used data on educational experiences in addition to (or instead of) occupational data in the analysis of social interactions.

Educational experiences are certainly important in theory. As far back as 1693, Locke's treatise *Some Thoughts Concerning Education* asserted that 'nine parts out of ten' of a person's character is formed through their education. In this and many further accounts, educational experience is portrayed as the source of important 'qualities of consequence' at the individual level, qualities that shape important further aspects of individual lives (cf. Mirowsky and Ross 2003).

Some of the qualities that are fostered by education are directly related to occupational roles. For example, formal qualifications can serve as an indicator of capabilities relevant to employment, and some qualifications act as a 'licence to practice' in certain trades and professions—not least the


P. Lambert, D. Griffiths, *Social Inequalities and Occupational Stratification*,
https://doi.org/10.1057/978-1-137-02253-0_10

advanced technical skills that are linked by some to an emergent 'technical middle class' in wealthy societies (e.g. Savage et al. 2013). However, educational experiences also develop more generic qualities—for example, behavioural patterns, social skills, social networks, and the acculturation of values and orientations—that can influence occupational outcomes. Powerful examples of these indirect mechanisms include Khan's (2011) depiction of how elite schools provide the most privileged children with the orientations and social networks that sustain elite positions in contemporary societies; and Willis's (1977) account of how oppositional or disadvantaged educational participation sets up some individuals for a lifetime of economic exploitation and social alienation. Moreover, many qualities that are developed through educational experiences might foster inequalities that cut across occupational categories and help to explain within-occupational variations in income and wealth accumulation (cf. Laurison and Friedman 2016)—one example is that differential social networks developed during education might lead to within-occupational divisions in social capital and network support (e.g. Field 2015; Schuller et al. 2004).

Data on educational experiences are routinely included within the sort of large-scale social survey datasets that we have used when analysing social connections and occupations—census datasets, labour force surveys, and other social surveys. Several of these resources also feature data on the educational experiences of the most immediate social contacts of the respondent: many hold information on the qualifications held by the spouse; some feature data on the qualifications held by other household sharers; and a few surveys also ask questions of the respondents about the qualifications held by other significant individuals from outside the household, such as parents or friends. In general, data on educational experiences is also of relatively high quality, in the sense that there is a substantial literature, and agreed good-practice guidance, on suitable ways of recording and storing data on education (e.g. Connelly et al. 2016; Hoffmeyer-Zlotnik and Warner 2014; Schneider 2008). However, there remain important challenges in using data on educational experiences appropriately (see Sect. 10.2).

In general, people with more advantaged educational experiences also have more advantaged positions in the occupational structure, but the relationship is far from deterministic. For example, Fig. 10.1 visualises the connection between highest level of educational qualification held

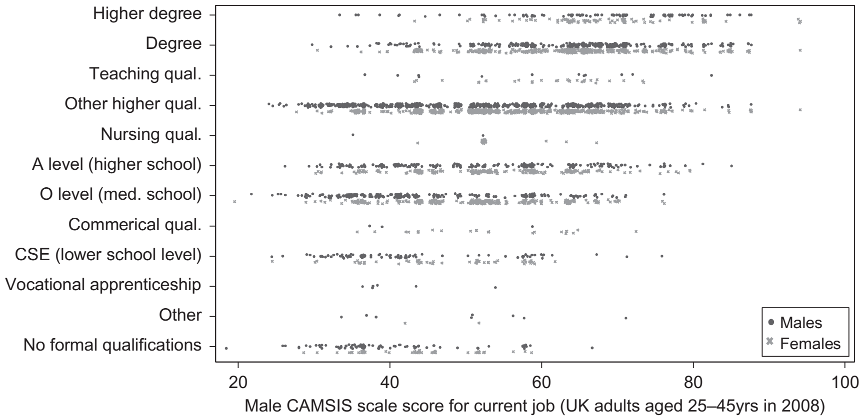



**Fig. 10.1** Relationships between highest educational qualifications and occupational advantage for UK adults in 2008. Source: Data from the BHPS (University of Essex 2010), *N* = 3817

and relative occupational advantage as measured by the male CAMSIS scale score for current occupation, for individuals in the UK. It shows a pattern of correlation, but also makes clear that there are plenty of people who diverge from the average pattern (the overall correlation is 0.55 for men and 0.49 for women, using the square root of the bivariate regression coefficient of determination). In principle, educational and occupational positions reflect two different dimensions of inequality, so datasets that measure both could provide fruitful opportunities for empirical evaluation of the relative influence of each (e.g. Bukodi and Goldthorpe 2013). However, this reasoning will only hold if we make the strong assumption that the measures of occupation and education capture their concepts without error, which is questionable in practice.

## 10.2 Educational Expansion and Reform and Heterogeneous Educational Experiences

Two important problems arise when studying data on educational experiences and its relationship to social stratification. First, there is often greater heterogeneity in educational experiences than is captured by standard measures. Second, in most countries and throughout recent history, that has been a tendency towards educational expansion over time and periodic reforms to educational systems, which means that different birth cohorts typically experience a very different distribution of educational experiences.

### 10.2.1 Recognising Heterogeneity in Educational Experiences

Empirical evaluations demonstrate that highly specific aspects of educational experiences, such as in terms of particular institutions attended, specific courses taken, and fine-grained differences in performance within courses, can all influence how education impacts upon individuals (e.g. Raffe and Croxford 2015; Savage et al. 2015, c7; Hallsten 2013; van de Werfhorst and Luijkx 2010; van de Werfhorst et al. 2001).[1] By contrast, in many survey data sources, there are only a small number of measured educational categories in which individuals could be located. Even recommended standard educational measures such as the 'ISCED 2011' or 'CASMIN' measures (e.g. Hoffmeyer-Zlotnik and Warner 2014) have relatively few categories in their standard versions (nine categories in both cases), and in comparative research, it is common to reduce measures that were originally more detailed into a simpler taxonomy that may reflect the 'lowest common denominator'—for example, the four categories of the harmonised measure of educational qualifications that is available in most IPUMS-I datasets.[2] If educational experience is much more fine-grained than is routinely measured in standard datasets, there is an obvious risk that empirical analysis underestimates the relative importance of education.

Below, we try to take some account of this possibility by performing some analyses of social interaction and social inequality patterns that use measures of education that are—to a limited extent—more fine-grained than normal. This is achieved by cross-classifying selected measures related to education, and comparing more and less aggregated versions of certain measures as are available in some data from the UK. However, the bulk of our examples still use relatively 'broad brush' measures. Accordingly, our own results might still fail to fully measure and respond to fine-grained heterogeneities in educational experiences.

### 10.2.2 Educational Expansion and Reform

A second major complication in using data on education is that in most countries, major features of educational systems have changed substantially through time. In general, formal educational systems have expanded, steadily and substantially, over time, but in most countries they have also been infrequently reformed and restructured in other important ways in addition (e.g. Connelly et al. 2016). Both expansion and restructuring mean that there can be considerable differences in the distribution of educational experiences between different birth cohorts. This relationship is particularly important when studying social interaction patterns because social interactions are also influenced by birth cohort; ignoring the cohort dependency of educational credentials could inappropriately introduce birth year itself as a structuring element to social interaction patterns involving education.[3]

The pattern of educational expansion is dramatic: in most countries, for at least the last two centuries, every decennial birth cohort has experienced a substantial increase in some form of educational participation levels when compared to their forebears (e.g. Simon and Boggs 1995). Expansion often results from deliberate strategies, such as social policies that are intended to increase formal educational participation in targeted areas (e.g. Arum et al. 2007; Halsey 1988). However, the persistent and gradational character of educational expansion suggests that there are other social forces driving expansion beyond policy reform alone (e.g. Brown and Scase 1995). Figure 10.2 illustrates selected elements of educational expansion in modern history. An important feature is that education typically expands to saturation points from lower levels of academic attainment upwards. This means that in earlier periods, participation in lower levels of academic education differentiates substantially between individuals, but in later years a similar differentiation would only be evident from examining differences at a higher academic level (see Fig. 10.2). In Europe and North America in the first half of the twentieth century, for instance, inequalities in post-compulsory educational participation accounted for most social inequalities of educational experience. In recent decades, however, social inequalities have been increasingly concentrated towards the specific courses and institutions attended rather than levels of

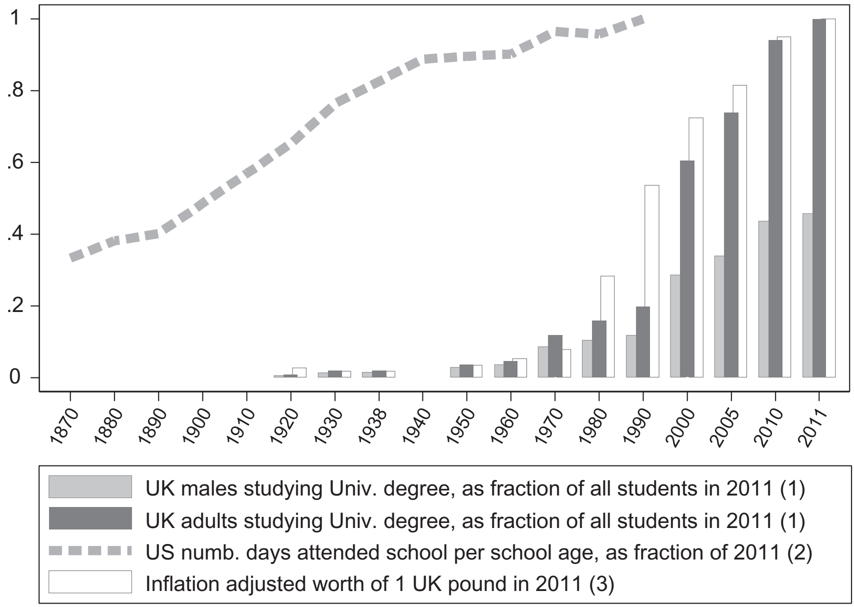



**Fig. 10.2** Indicators of educational expansion, 1870–2011. Sources: (1) from Bolton (2012, p. 21), (2) from Simon and Boggs (1995, p. 210) (derived), (3) from Bank of England Inflation Calculator, http://www.bankofengland.co.uk/education/Pages/resources/inflationtools/calculator/ (accessed 1/1/16)

participation in themselves (e.g. Savage et al. 2015, c7; van de Werfhorst and Luijkx 2010; van de Werfhorst et al. 2001).

It we are analysing data that spans different birth cohorts, it is usually compelling to 'control' for educational expansion. This might be achieved by disaggregating analysis by birth cohorts, by allowing for relevant interaction effects (for the effect of education by cohort), or by statistical adjustment for the relative distribution of educational experiences within birth cohorts (e.g. Buis 2010; Wong 2010). Nevertheless, many empirical studies do not control for expansion. As is illustrated in Fig. 10.2, the overall rate of educational expansion has followed a remarkably similar trajectory to that of monetary inflation, yet whilst in empirical research we would seldom think of comparing income figures through time without adjusting for inflation, it is salutary to recognise that social science researchers routinely do the equivalent for educational qualifications data.

Educational expansion has also been distributed in socially unequal ways. Across countries, trends in educational expansion have often favoured women, in the sense that there has usually been a greater acceleration in women's participation and qualification levels in comparison to men (e.g. Hout and DiPrete 2006). In terms of ethnic inequalities, in many countries periods of educational expansion correspond to increasing diversification in the educational attainment profiles of minority groups, with some making relative improvements, but others increasingly isolated or excluded (e.g. Modood 2005; Portes and Rumbaut 2001). The overall influence of social stratification background during periods of expansion has however been a 'persistent inequality' (Shavit and Blossfeld 1993)—that is, alongside dramatic educational expansion, the impact of social origins upon salient inequalities of participation and attainment has been largely stable (e.g. Ermisch et al. 2012, [c2]; Pfeffer 2008; Shavit et al. 2007; Shavit and Blossfeld 1993).

Educational expansion is however only half of the story in understanding change in educational experiences over time. Most countries also experience significant adjustments to the administrative arrangements of their educational institutions at certain points in time. For instance, reforms might lead to sudden adjustments to assessment methods or certification standards, or even to the reclassification of certain types of activity in terms of their formal educational level or institutional affiliation. Likewise, new social policies may be introduced that seek to change educational participation experiences for certain social groups (e.g. Lupton et al. 2009). A prominent example can be seen in national training regimes for nurses. A century ago, trained nurses across the globe ordinarily received extensive formal training that was delivered within the healthcare environment or in affiliated specialist colleges; but in many countries, at discrete points over the last century, the training of nurses was moved into university settings. In this example, nurses may be characterised by sharp differences in educational credentials that are dependent upon when they received their training and in which country, but it is unlikely that the actual work, lifestyles, or circumstances of nurses from different educational cohorts is as substantially different. Alongside such examples, however, it is wise to remember that not every aspect of educational provision has changed greatly through time—in the UK, for example, the educational trajectories of those entering many traditional professions (such as medical doctors, lawyers, religious professionals, and higher education teachers) have barely altered in 300 years!

## 10.3 Analysing Educational Inequality Through Social Interaction Distance and Network Approaches

Mindful of the heterogeneity, expansion and reform of educational experiences, could combining educational information with social interaction data better inform us about the social structure of inequality? First, we could look at the social interactions between people as a function of their educational experiences alone and examine the structure of educational and social inequality revealed. In the methodologies covered in Chaps. 4, 5, 6, 7, and 8, this would entail constructing CAMSIS scales, or using SNA depictions, for units that measure educational rather than occupational positions. Second, we could analyse units that constitute occupations in combination with data on educational experiences—that is, performing CAMSIS and SNA analysis on units formed by cross-classifying educational criteria and the occupation.

### 10.3.1 Summarising the Social Interactions Between the Holders of Educational Qualifications

The core patterns evident from analysing social interactions between educational categories are interesting and are consistent with interpretations of social inequality based upon other social interaction patterns. There have been some previous exercises that have used social interaction distance models to explore educational structures on the basis of friend-friend or husband-wife association patterns (e.g. Lambert 2012; Ultee and Luijkx 1990),[4] and all have suggested a dimensional structure influencing social interaction patterns, which could be interpreted as the relative stratification advantage associated with the qualification level.

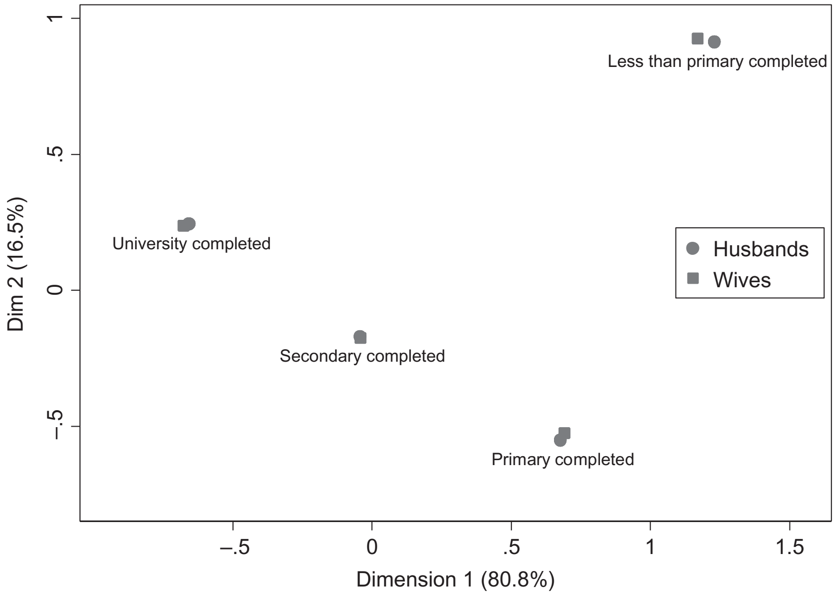



**Fig. 10.3** A SID solution for US educational homogamy. Source: USA 2010 census, accessed from IPUMS-I. SID uses 'edattan' measure, allowing for three dimensions

However, because most studies work with only a small number of different educational categories, it follows that the evidence of differences between categories is relatively less discriminating than would be obtained for a comparable exercise using occupations.

As examples, Fig. 10.3 shows the two-dimensional structure for educational categories from a SID analysis of US data, and Fig. 10.4 summarises the first-dimensional structure from nine further examples, six applied to the four categories of an IPUMS-I measure in six different countries, and three applied to UK data (which has been coded into different schemes and also shows, for comparison, how other stratification indicators are associated with the educational levels). Of course, the first-dimensional structure could reasonably be given other interpretations than that of 'stratification advantage', for instance, the structure is also consistent with a ranking of the skill or difficulty associated with the qualification, or with the 'human capital returns' linked to the level. Both

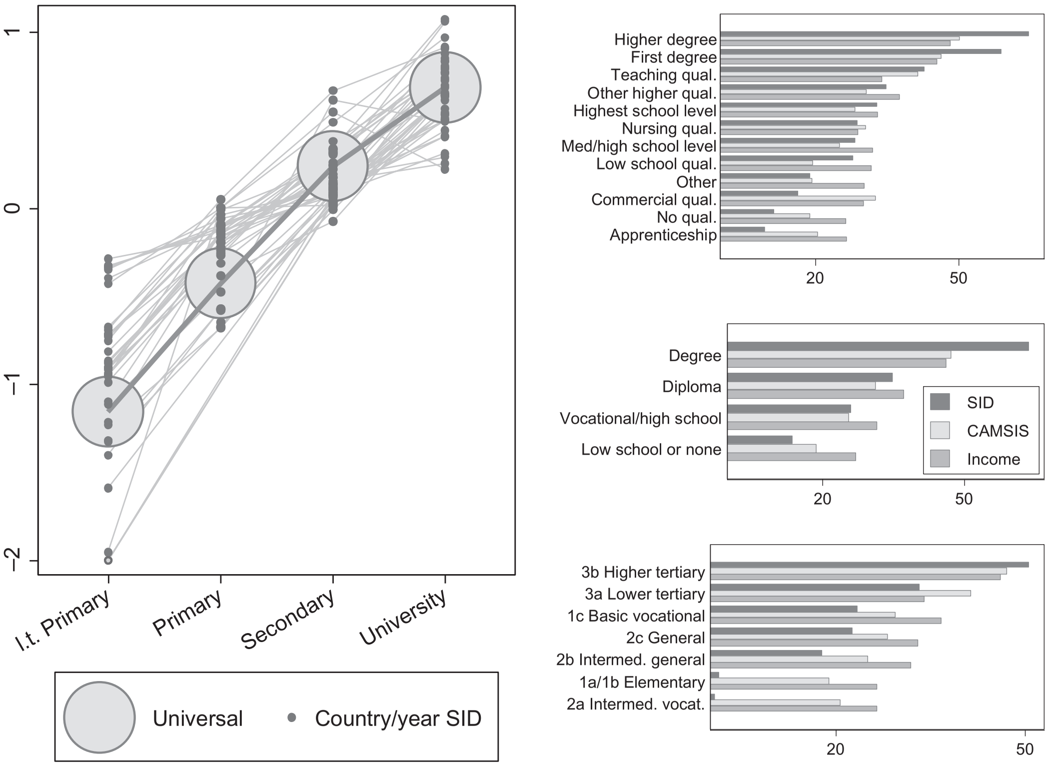



**Fig. 10.4** Illustrating SID scores in 'dimension 1' for educational homogamy across selected contexts. Source: Left panel uses IPUMS-I data for m-f cohabitees (scores for male educational levels). Data shown for: US, 1960–2010; MX, 1970–2010; FR, 1962–2006; HE, 1971–2001; HU, 1970–2001; ES, 1991–2001; CH, 1970–2000; PT, 1981–2001; IE, 1971–2006 (lines link data from same country-year; scores <−2 are cropped at −2). Right panel uses UK data from BHPS m-f cohabitees (1991–2008), showing SID scores for educational levels plus mean CAMSIS/income for adults with those qualifications

in the examples shown (Figs. 10.3 and 10.4) and in other literature, however, second and subsidiary dimensions to the social interaction structure involving educational measures do not seem to reflect empirically or theoretically important social differences. For example, the second dimension structure of Fig. 10.3 could perhaps be related to regional or age-cohort differences in educational combinations, but it seems to be of limited empirical importance and ambiguous interpretation. It is also interesting to note that there is little compelling evidence of difference between the scores allocated to educational levels for males and females. Using the same interpretation as applied to the SID analysis of occupations, a lack of gender difference would suggest that the social consequences of different educational levels of attainment are largely the same for men and women.

One attraction of the SID approach might be that it generates a 'score' for educational categories that might be conveniently used in other statistical applications. The dimension scores might also be used as a device for making appropriate comparisons between qualifications from different taxonomies and/or in different social contexts (Lambert 2012). For example, the left panel of Fig. 10.4 suggests modest variation between different country-year combinations in the relative positioning given to the same categories; the right panel indicates the scaling of measures under different taxonomies, which are given a certain comparability by their use in the same social interaction distance scaling procedure. However, there are also plenty of other plausible ways of assigning scale scores to educational categories if this is required (cf. Buis 2010), and when the number of educational categories recorded is relatively small, it may in any case be just as convenient to describe the specific educational circumstances (and estimate parameters) for each individual category.

A key issue in assessing the role of educational qualifications is cohort change in the distribution of qualifications. The SID score approach could generate different SID scores for categories of the same qualification for different birth cohorts. Figure 10.5 summarises data of this nature: it shows data for the UK that was collected in Labour Force Surveys between 2001 and 2010, and the SID score for the educational qualification category (arithmetically standardised) when the SID analysis is run within birth cohorts.

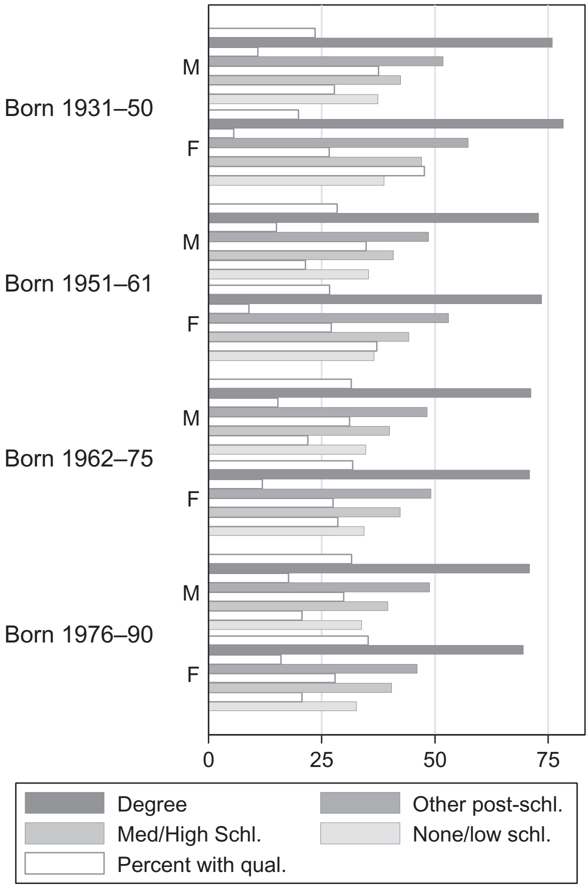



**Fig. 10.5** SID scores for UK education categories, by birth cohort and gender. Source: UK Labour Force Survey 2001–2010, *N* ~ 200k couples. Analysis of highest educational-level data for cohabiting couples (by year of birth of male)

Figure 10.5 is a little complicated to look at because if includes several comparisons. We divided the sample into four birth cohorts that were amply represented in the data and that coincide, broadly speaking, with four different phases of educational expansion and reform in the UK. From the SID scores (the shaded columns), we can see that the allocated scores for educational categories are broadly comparable between genders and between birth cohorts. However, we can also see some changes in the scores given to categories by birth cohorts. For example, the relative score for the degree-level qualification category declines slightly through the more recent cohorts, and the distancing between scores within cohorts declines somewhat as well. Perhaps confusingly, the standardised score of the lower-level qualification categories also declines through the more recent cohorts—that might seem odd at first sight, but it does make sense when we take into account the changing proportions of people in the different categories between cohorts (evident from the clear bars): the relative advantage of the most advantaged category decreases through time, as more people achieve that level; at the same time, the relative disadvantage of the least advantaged category is exacerbated through time, as fewer people hold that category. These patterns of cohort change for the SID scores make theoretical sense and illustrate another way that social interaction data can help reveal the distribution of social inequality. To reiterate however, there are many other summary statistics about educational profiles that we could have calculated for the different categories that would have shown similar trends, so the added value of using the SID method to understand educational inequalities is not as dramatic as it might be in the case of occupations.

It might also be useful to explore the patterns of ties between different educational qualifications in the same style that we used network techniques to explore social ties between occupations. In this approach, a link between qualification categories would be defined as an above-average propensity to interact socially (see Chap. 7). Figures 10.6 and 10.7 give two examples, for data from the UK Labour Force Survey (with four educational categories) and using the UK's British Household Panel Study (with 12 educational categories). The sociograms in these examples paint a familiar picture: educational categories that are more similar in their level are generally more likely to have social connections between them. The main pattern seems to be a hierarchical ranking of educational categories in terms of academic requirements. In the case of the analysis of occupations, network sociograms offer us the chance of identifying unanticipated connections between occupational categories that might otherwise be overlooked. There is some limited evidence that comparable insights can emerge when undertaking similar analysis with data on educational qualifications, although the smaller number of categories means

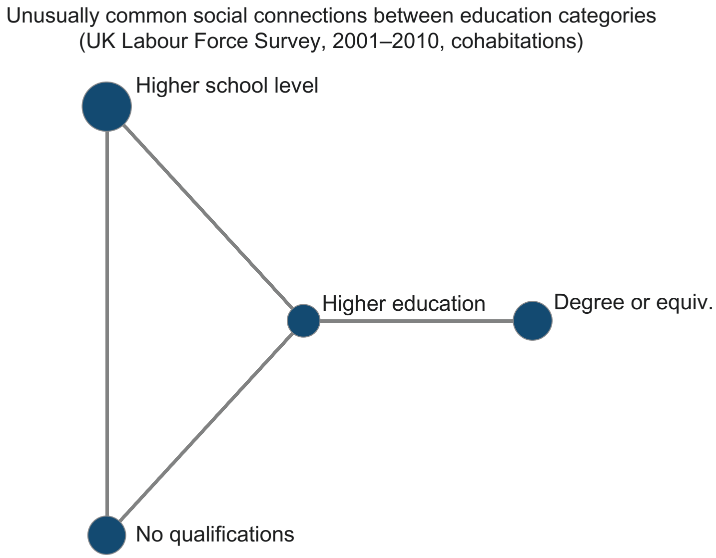



**Fig. 10.6** Network diagram depiction of social associations between four educational categories

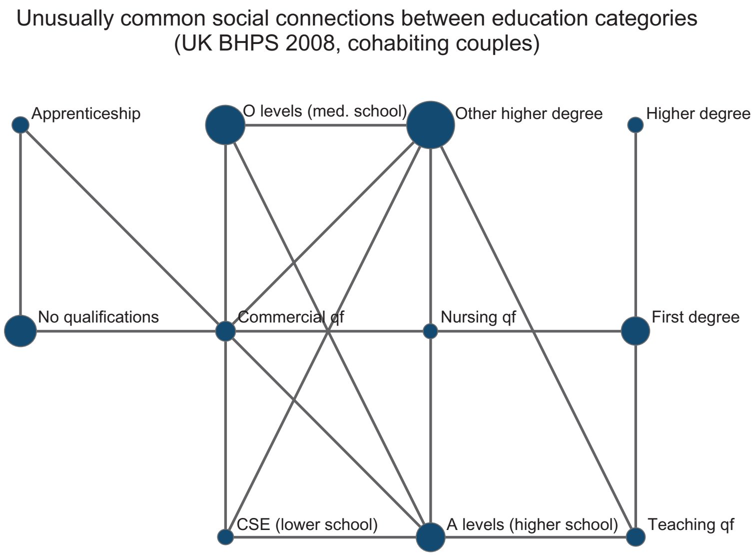



**Fig. 10.7** Network diagram depiction of social associations between 11 educational categories

that there are fewer opportunities to uncover unanticipated results. For example, there is some suggestion in Fig. 10.7 of a proximity between different vocational qualifications (e.g. 'apprenticeship' and 'commercial qualifications') and perhaps some contribution to understanding the large but ambiguous category of 'Other higher degree' (its many connections suggest that it is a heterogeneous category but that it is not academically elitist, because it is not disproportionately connected to the two most academically prestigious categories; in fact, the BHPS documentation confirms that the label refers to a range of post-school qualifications that, on the whole, occupy intermediate positions in conventional hierarchies of attainment).

In general, a useful contribution from analysing social interactions between categories can be in 'sorting out' the positions associated with a wide range of different categories—estimating the relative social position of a category, on the basis of data about the social interactions of its incumbents. In the case of data on educational levels, we normally have a small number of relatively broad-brush indicators of education, which lessens the added value of exploratory techniques. However, there are some scenarios when more complex information is available, such as on the type of institution(s) attended, specific courses taken, or number or grade point of qualifications. In Fig. 10.8, we show one example of the possible contribution of SID tools in this scenario. The figure characterises the first-dimensional structure revealed by using a social interaction distance analysis for data on cohabiting couples in terms of up to 72 different educational positions. The positions have been defined, for the UK's BHPS survey, by cross-classifying the 12-category measure of 'highest qualification held', against a three-category measure of the type of school attended, and a two-category measure of whether a further education institution was attended. In Britain it is very plausible that the average social circumstances of people who share the same educational level, but have different experiences of school and college, may be different. The SID analysis seems to confirm this: on the right-hand side, we show the dimension 1 scores for the BHPS education categories overall, but on the left-hand side, we show the complicated variations in dimension 1 scores for the different combinations of school, college and educational level. There are some examples of fairly substantial variations within educational

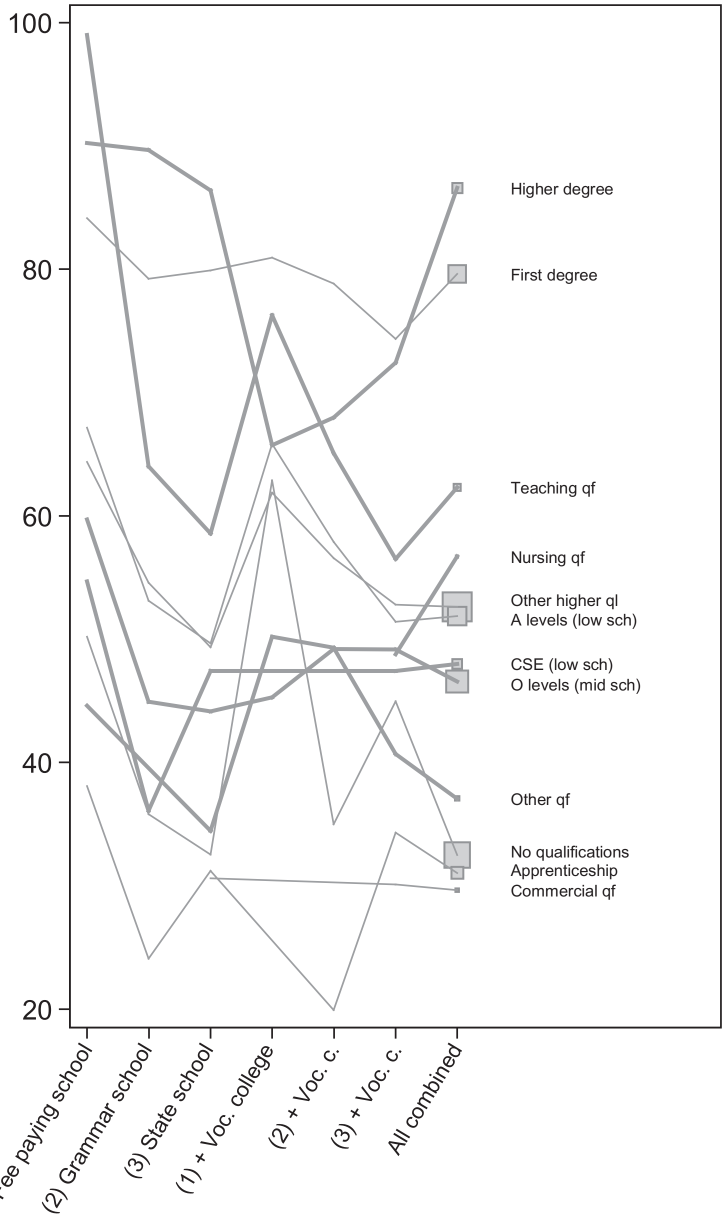



**Fig. 10.8** SID scores for combinations of educational circumstances. Source: British Household Panel Study, pooled data 1991–2008

levels—for example, the positive impact of a higher degree is much greater for people who went to fee-paying schools and who did not go to a technical college; the apparent disadvantage of having no qualifications seems considerably greater for those who attended a grammar school (a state school that was selective on the basis of ability) than for all others. Indeed, the fluctuations away from the horizontal lines in Fig. 10.8 point to some interesting variations in the circumstances of individuals according to different elements of their educational experiences. In this application, the SID analysis of disaggregated educational categories does seem to reveal something about social inequalities that would have been masked by a more broad-brush approach. However, this scenario using the BHPS survey data from the UK is somewhat unusual since we are able to combine large quantities of data on social interaction patterns, with relatively rich data on the educational experiences of both an 'ego' and an 'alter'. Although the methodology is promising, there may not be that many other data situations where comparable exploratory analysis is constructive in this way.

### 10.3.2 Interaction Structure in the Cross-Classification of Occupational and Educational Categories

A further possibility is to assign individuals to categories defined by the combination of occupational and educational positions and then explore the social interactions between people in the relevant positions. We have tried this in a number of scenarios and have presented some relevant results previously (Griffiths and Lambert 2012).

It is easy to imagine that the occupational-educational combination might constitute a coherent unit of analysis when studying social stratification and social interactions. University graduates might cluster more tightly together within occupations (perhaps graduates in non-graduate or mixed jobs—say working at call centres—will tend only to interact with co-workers of a similar educational background). Similar links might cut across occupations (e.g. perhaps there will be connections between graduates across diverse occupations which require similar educational backgrounds—such as the example of university friends staying in contact after moving to different occupations). In either case, the occupation-education combined position could be more revealing than a measure of occupation or education alone.

There are however some operational issues in defining combined educational-occupational positions, because many occupational units are educationally homogenous—that is, only some occupational unit groups have moderate numbers of cases with different levels of educational qualifications, and some instances of occupation-education combinations should not realistically be possible at all (for instance, a medical doctor without a university-level qualification). By this reasoning, it could make sense in some situations if only certain occupational categories are split by educational level. In CAMSIS analysis of occupational units alone, a working convention is to analyse separate occupations only if they are represented in the data by at least 30 cases, and a similar criteria could be applied here. Given the structural relations between occupations and educational positions, it could also be logical to only disaggregate occupational and educational combinations if there is data that shows a moderate level of educational heterogeneity within the occupation—say that at least 15% of the occupation must be in the non-modal educational category. In our analysis below, we applied only the first criteria: we disaggregated occupation-by-education combinations whenever possible, but we restricted attention to combinations represented by at least 30 cases and excluded any combinations represented by fewer cases.

The key pattern associated with the analysis of combined occupation-education units, that seems to be consistent in the many different evaluations that we have considered, is summarised in Table 10.1 and Figs. 10.9 and 10.10. When we analyse social interactions between the incumbents of positions defined by the combination of educational and occupational circumstances, it seems that the educational levels 'dominate' the dimensional solution or network structure. In the SID solutions, summarised in Table 10.1 and Fig. 10.9 for an analysis using nine occupational categories and four education categories, one end of the first dimension structure is dominated by the highest qualifications, and the other end by the lowest qualifications, with little crossover between the approaches. When we expand the analysis to use more detailed measures of occupation and/or education, moreover, the same pattern persists (Fig. 10.10)—educational

**Table 10.1** Summary of the SID solution for the analysis of combined categories of occupation and education (UK Labour Force Survey, marriages, 2001–2010)

| Occupation (major group) | | Degree | Other post-school | Medium/high school | None/low school |
|---|---|---|---|---|---|
| | SID score in dimension 1 {*Z*-score * 15 + 50} | | | | |
| | Occupation only | Occupation * Education combinations | | | |
| 1. | 54.9 | 62 | 52 | 47 | 45 |
| 2. | 65.6 | 69 | 54 | 51 | 49 |
| 3. | 55.7 | 63 | 53 | 49 | 47 |
| 4. | 51.1 | 60 | 51 | 46 | 44 |
| 5. | 43.7 | 54 | 46 | 42 | 40 |
| 6. | 46.1 | 58 | 49 | 45 | 42 |
| 7. | 44.9 | 57 | 49 | 43 | 42 |
| 8. | 38.7 | 51 | 43 | 38 | 38 |
| 9. | 37.6 | 52 | 44 | 38 | 37 |

Source: UK Labour Force Survey 2001–2010. SID scores for SOC2000 major groups/education

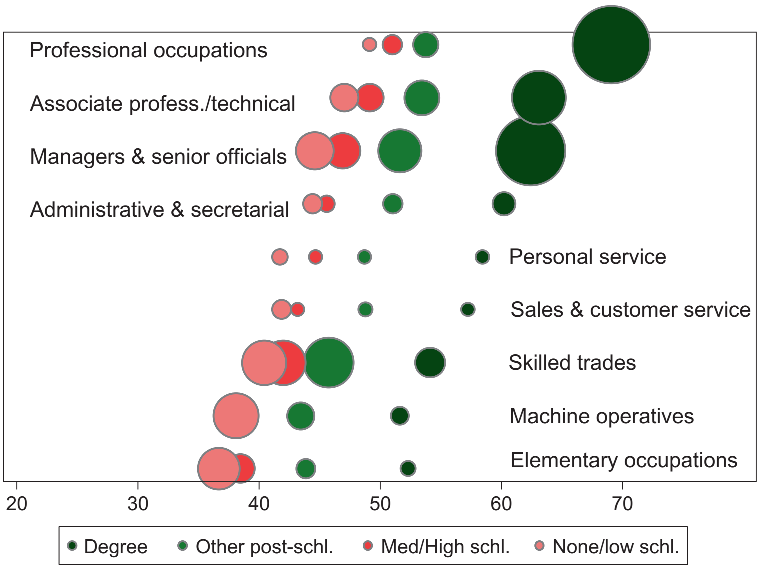



**Fig. 10.9** Visualisation of the SID solution for the analysis of combined categories of occupation and education. Source: As Table 10.1

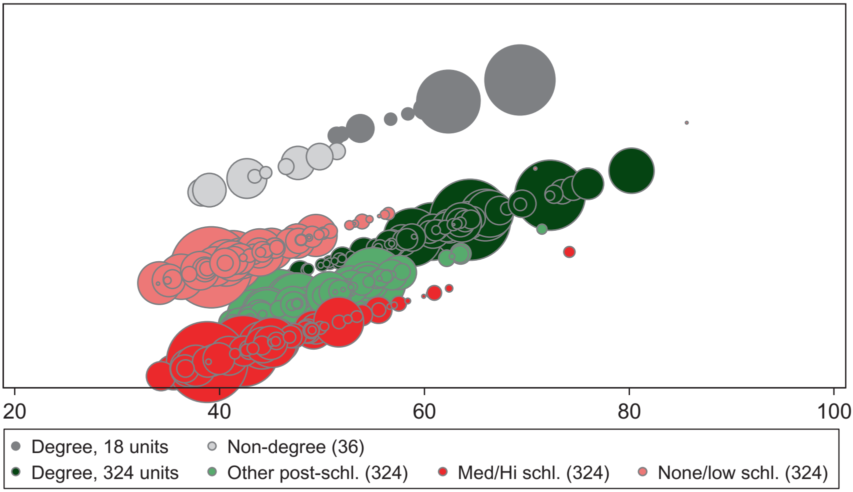



**Fig. 10.10** Visualisation of SID solution for analysis of combined categories of occupation-education, using more and less detailed measures of occupation and of education. Source: As Table 10.1

position is the stronger determinant of dimensional location, with only modest overlap between occupational positions net of educational circumstances.

The apparent dominance of education could be substantively significant, but it is important to recognise the gradation between occupations that seems to occur within educational levels as well. These patterns suggest that for a given social category (occupation-by-education combination), social interactions are more strongly shaped by educational background than occupational position. However, within educational positions, there is gradation in social interaction patterns that is related to occupations in the standard way. The limited examples of overlap suggest that there are only a few scenarios where a more advantaged job combined with a less advantaged educational background 'outranks' a more advantaged educational background.

This pattern seems to hold across a wide range of societies and data scenarios. In Fig. 10.11, we summarise data from evaluations that we have undertaken using census datasets from nine countries and a range of time periods for which we have convenient access to data on husband-wife pairs. The figure shows the mean scores in the first dimension of the SID solution for all the different occupation-education categories, split by the education level, and gives some information on the distribution around that mean. The key point is the consistently limited overlap between distributions: the separation seems overwhelmingly to lie with educational circumstances rather than occupational ones.

Similar stories are revealed when analysing the spread of social interaction patterns through network sociograms. Figure 10.12 shows data on occupation-education combinations that have disproportionately common social interactions between them. The network depiction offers some useful additional contributions. First, we could now conveniently identify and reflect upon the key ties that seem to connect across education levels. Second, the SNA depiction offers a different leverage on the comparison between different societies. Whereas the SID approach lends itself to aggregate statistical summaries of average differences (cf. Fig. 10.11), the network sociogram makes it feasible to visualise and compare differences in the interaction patterns from one society to another. In Fig. 10.12, for instance, when we compare the USA and Philippines, we see a pattern of a greater number of links between graduates and non-graduates in the USA, which might perhaps be interpreted as the impact of educational expansion (it is possible that, because of educational expansion, graduates are distributed more broadly across the occupational structure, so that it is more common for them to interact with non-graduates than it would be in societies with fewer graduates).

Could these patterns be read as evidence that education matters much more to the long-term social structure of inequality than occupational position? The answer, for us, is to return to the disparities between educational and occupational circumstances as were evident in Fig. 10.1. Clearly, there are some people in circumstances linked to their measured occupations that are at odds with their measured educational positions, but whether these differences are important social patterns, or the result of measurement error (exacerbated by the broad-brush level of educational detail), is an open question. If these patterns were genuine, we might think of scenarios where the same occupations contain separate tracks of employee—those with higher qualifications might have different lifestyles, orientations and social connections, and might ultimately

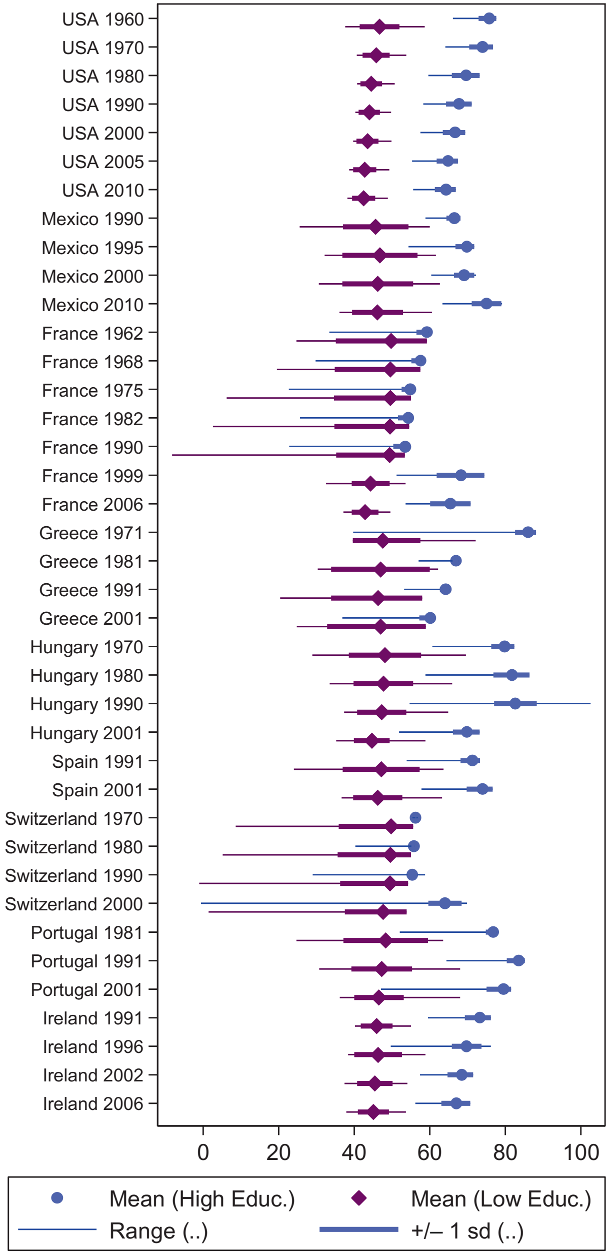
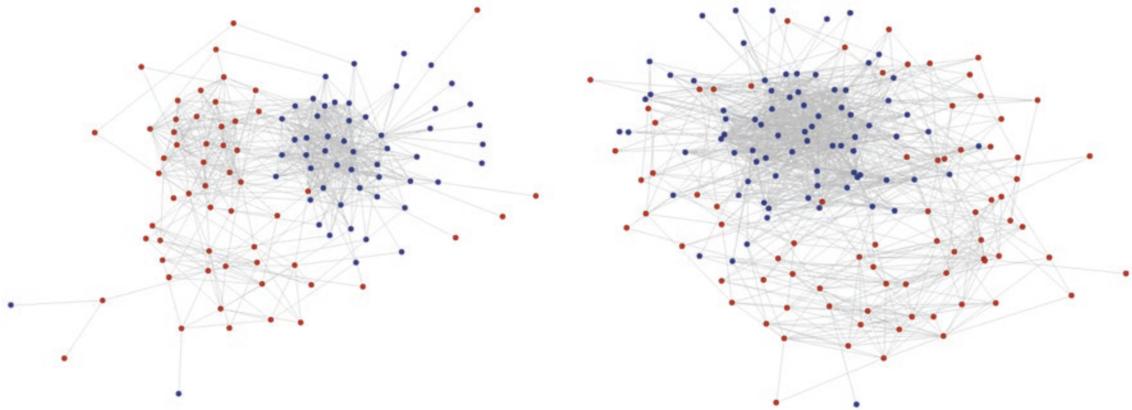



**Fig. 10.11** Summary of the SID solution for occupation-education categories across a range of societies and time periods. Source: As Fig. 10.4



Philippines 2000 USA 2000

**Fig. 10.12** Network sociograms depicting separation of social interactions between graduates and non-graduates within occupational categories for the USA and Philippines. Notes: Data from IPUMS-I. Nodes represent occupation-education combinations. Blue = graduates, Red = non-graduates. Ties represent disproportionately common social interaction patterns (two or more times more likely than chance) be heading towards a different life-course trajectory. Indeed, Stewart et al. (1980) emphasised this scenario in the case of male clerks in the UK in the mid-twentieth century. On the other hand, if measurement error were the bigger issue, we would be suggesting that the difference between people in the same occupations but in different educational categories is more a reflection of unimportant or artefactual issues of the measurement of education, rather than of meaningful differences in circumstances. Are those primary school teachers who don't have a university degree, for example, different from those that do in an important way, or does the difference emerge from unimportant exogenous patterns such as birth cohort? Particularly in the case of education, birth cohort may be behind many of these differences, and birth cohort also influences social interaction patterns (people tend to interact with others from similar birth cohorts).

Figure 10.13 summarises the first dimension of SID scores for education-occupation combined categories for adults from within the 1951–1961 birth cohort in the UK. The same division between educational categories is evident, but the severity is a less pronounced—there is a little more overlap between occupation and education categories.

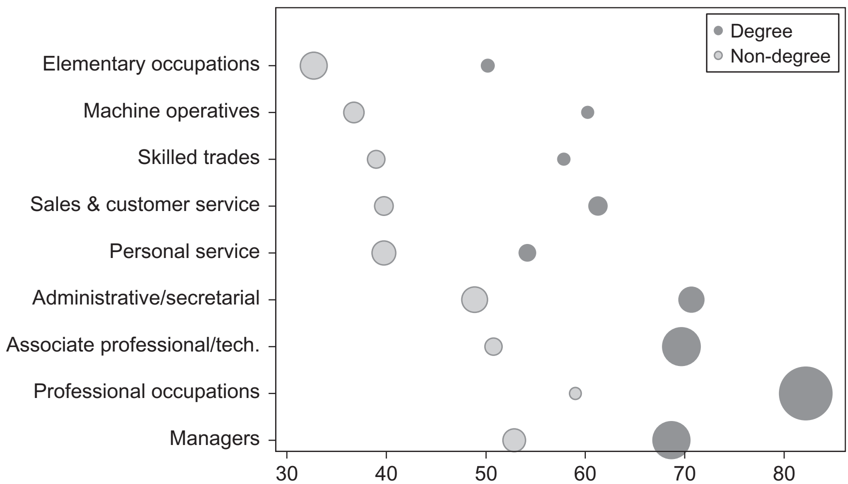



**Fig. 10.13** SID scores for combined occupation-education units, restricted to the 1951–1962 birth cohort. Source: UK LFS 2001–2010 (births 1951–1962)

Moreover, in the presentational style shown in Fig. 10.13, it is evident that there is considerable variation in estimated positions within educational qualification categories as well as within occupational categories. This pattern suggests two things. First, some but not all of the prominent patterns linked to education probably do reflect measurement error in education, since the scale of the pattern is diminished when we restrict analysis to a narrow range of ages. Second, the patterns of variation between education-occupation groups are such that it seems likely that neither education alone nor occupation alone would capture the full texture of differences in social positions: there are important differences within occupations by educational levels, and there are important differences within educational levels by occupations.

Following this logic we might then anticipate that the most important approach to understanding social stratification should be one that involves both educational and occupational positions in a combined social interaction distance scale. This is plausible, but we are not yet convinced that it is necessary. At a pragmatic level, it is a lot to ask that a measure should be based upon both educational and occupational data. With regard to analytical methods, moreover, it is readily possible to involve educational indicators in an analysis through separate measures, rather than in a manner that is integrated with occupational measures. Figure 10.14, for example, summarises a number of regression analyses in which, for cohabiting heterosexual couples, the male partner's individual characteristics are used to predict the female's CAMSIS score. The bars show a function of the variance explained when different aspects of the male characteristics are exploited as predictors. The first point to note is that in all the examples, the second model (light-shaded bar) is amongst the best performing. This uses a SID scale based on occupations, plus a separate set of dummy variables to capture the education effect—a conventional approach to separating measures of education and occupation. By contrast, there are no scenarios where the combined occupation-education scale scores are superior to any other alternative; although in predictive terms they do improve on the SID scale that only uses occupations, the gap is small and is not as much of an improvement as the conventional approach. Lastly, there are slight variations between age cohorts in the relative performances of the different models. The relative strengths

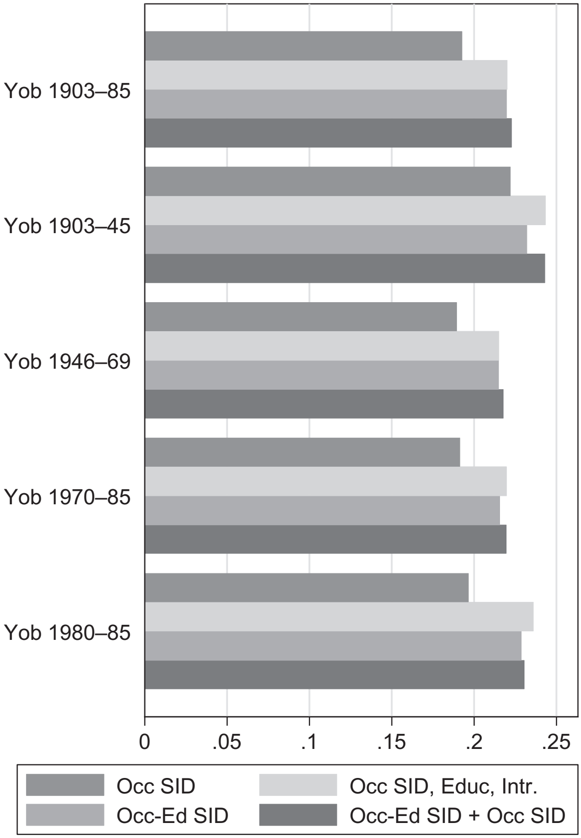



**Fig. 10.14** Summary of the performance of regression models predicting spouse's CAMSIS, for male LFS respondents by birth cohort. Source: UK LFS 2001–2010. Figure shown is R2 in predicting wife's stratification score

of the occupation-by-education SID measure seem less substantial for the older generation, and the gaps between different combined measures vary between cohorts. It seems plausible that the impact of accounting for education is greater amongst the younger cohorts, but since the pattern suggests that age cohort does mediate the relative influences of occupation and education, we suspect that a safer approach is the conventional strategy of using a SID scale for occupations, and separate measures for education levels.

### 10.3.3 School or University?

An enduring question concerns which sorts of educational divisions are the most important to take account of. Our analysis of social interactions provides some evidence on this issue. One hypothesis is that 'graduateness' is an especially critical division (the difference between those who have had an extended period of post-compulsory academic education and those who have not). The category of 'graduates' may be defined as those who have attained a 'university/college degree' or those who have 'ever attended university', including partial completions. However defined, the number of graduates is increasing across societies, and there is ample evidence that it is specifically the threshold of university which accounts for a variety of major life-course privileges such as of health and social and economic well-being (e.g. Mirowsky and Ross 2003; Attewell and Lavin 2007 argue that even a very brief 'intervention' of university educational support for non-traditional cohorts of students can successfully transform lives and the experiences of offspring).

Another educational division that is often thought to have considerable influence is the 'type' of school attended for all or some of a person's childhood. In the UK (e.g. Connelly et al. 2016), the majority of children attend state-provided non-selective schools, but around 1 in 20 attend privately funded schools. In addition, dependent in part upon the region within the UK in which they lived as children, many contemporary residents of the UK may also have attended academically or socially selective schools, such as 'grammar' schools (academically selective schools that were used widely until 1974), or religious schools (typically state-supported schools with a specific mandate to provide education only to children of a nominated religious background). Comparable differences of school type can be found in many nations, and there is evidence that divisions of school type are empirically associated with differences in other outcomes (e.g. Khan 2011). Nevertheless, school type is only occasionally measured in social surveys, and is largely invisible from the broad categories of attained educational qualifications analysed above. School type might in any case be relatively less important than is often thought, if the impact of school type upon later life outcomes acts substantially through attained educational qualifications. In the UK BHPS, for instance, our own analysis suggests that an eight-category measure of school type explains 7.9% of the variance in male log incomes net of age for adults aged 25 and above, but this diminishes to 2.3% once highest educational attainment (12 category measure) is added as a further control. Empirically, there are residual influences of school type upon outcomes that are not captured by existing measures (see also Laurison and Friedman 2016), but the scale of these is, arguably, relatively small.

With regard to social interaction data, within the UK private school attendance is probably not associated with as extreme a social divide as is sometimes imagined. Empirical data on marriage partners suggests that it is very common for marriages to cross this particular educational divide. Using the BHPS in 2011, for example, of 311 couples where one or both partners were privately educated, only 35 were homogamous in this way. The relative odds are still disproportionate in favour of homogamy,[5] but the point is that the social division of school type is not as stark as is often imagined. Instead, the clues provided by our analysis of social interaction patterns point towards university-level participation being a particularly important educational division, and a much more important division than that related to school type. When explored briefly above (Fig. 10.8), there was a suggestion that school type does have some role to play in influencing social positions, but the variations were on the whole within the range of the overall educational level. When differentiations were allowed between university-level categories and others, however (e.g. Figs. 10.9 and 10.10), social interaction patterns were quite sharply divided. Social interaction data therefore seems to lean towards the relative importance of university- rather than school-level experience for contemporary populations as a whole.

## 10.4 Educational and Occupational Change

Substantial recent educational expansion may also have altered the way that occupational inequalities operate—that is, the social circumstances of the incumbents of occupations may have been altered due to their changing educational compositions through time. We already have evidence that the overall structure of occupational inequalities probably does not change substantially over time or between countries, for instance, in the relative stability of CAMSIS scales across societies (Chap. 5). However, we may be able to evaluate this pattern more thoroughly by analysing change between birth cohorts in the joint position of individuals in occupational and educational terms. Figure 10.15 summarises a pertinent result. It explores the empirical relationship between educational and occupational change for two different birth cohorts in the UK (births 1951–1961 contrasted with births 1976–1990). The horizontal axis represents the extent to which the estimated SID score for the same occupation has increased or decreased between the cohorts: an increase, for example, would suggest that the average profile of the social interactions of the incumbents of the occupation has become more advantaged (perhaps due to occupational expansion or contraction, or genuine compositional changes in the circumstances of the

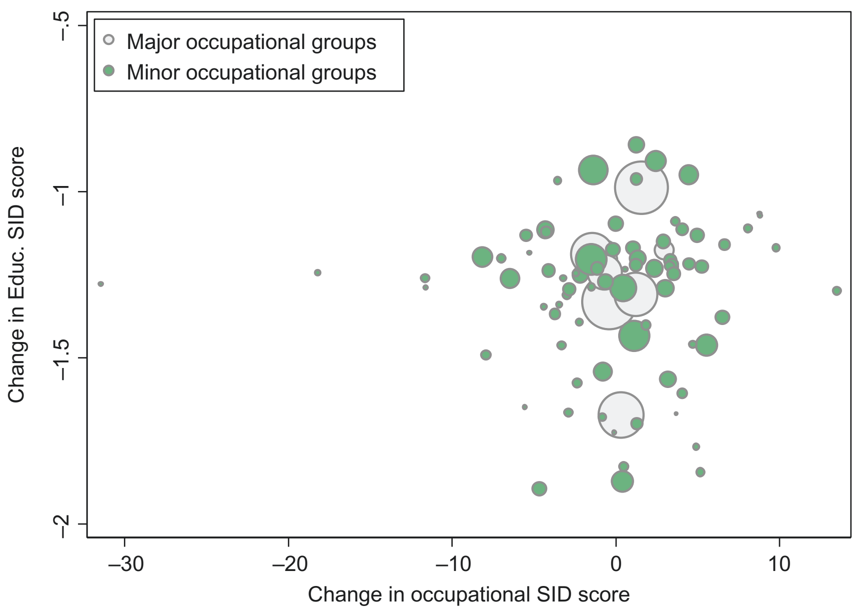



**Fig. 10.15** Cohort change in SID scores for occupations and change in the mean SID scores for education for the incumbents of occupations. Source: UK Labour Force Survey, 2001–2010, cohabiting couples. 'Change' refers to difference between 1976–1990 birth cohort and 1951–1961 birth cohort

occupation). The vertical axis represents average change in the educational scores of the incumbents of the occupation, calculated on the basis of the SID scores for educational categories that are assigned to each relevant individual. If educational expansion through time had a major impact on the underlying stratification structure of occupations, we would expect a structured relationship between the two change scores. As it is, however, we see no strong pattern of difference, a null result that seems to be consistent with the view that educational expansion proceeds largely exogenously of occupational change.

## 10.5 Summary

In this chapter, we have reviewed the way that educational measures influence the story told when analysing occupational and educational inequalities with data on social interactions. There are some promising opportunities for studying social interaction patterns in relation to educational categories or indeed in relation to categories defined by the combination of occupational and educational position. However, there are significant complications connected in particular to educational expansion and reform. There is some suggestion from comparative patterns that as educational systems expand, the disparities in social interaction patterns between educational levels diminish (e.g. Figs. 10.11 and 10.12). Equally, it might be more useful to analyse direct measures of education, rather than measures that are cross-classified with occupations, in recent periods after educational expansion (Fig. 10.14). These patterns ought to be considered in the context of long-term stabilities in the social interaction patterns linked to occupations—it seems likely that educational expansion and change proceeds at a faster pace than occupational change in relation to the stratification structure. Such patterns lead us on balance to the view that in most scenarios a more traditional approach, which uses separate indicators of occupational circumstances and education, and prioritises data on occupations rather than on educational experiences, is a safer strategy, associated with less risk of incorrectly conflating birth cohort change with depictions of social stratification.

## Notes

1. That is, many divisions in educational experience, such as the specific school attended, or exact degree class obtained, are rarely measured in survey instruments. On most social surveys, for example, most readers of this book would belong to the same category of educational experience and attainment as the prominent conservative politicians David Cameron and Boris Johnston (i.e. holding university-level qualifications), but few will feel that they have had the same educational experience (Cameron and Johnston's education is often characterised as being elitist and exceptionally privileged).
2. Indeed, since categorical measures of educational qualifications are difficult to process consistently over countries or across time, there is a good case for representing educational differentiation through gradational rather than categorical measures (cf. Buis 2010; Prandy et al. 2004).
3. Occupational distributions do also evolve over time, but the scale of occupational change is much less substantial, and its pace less rapid, than that of educational expansion and restructuring. In addition, individuals themselves may evolve in their occupations in times of social change (whereas educational measures for individuals usually remain 'stuck' at the qualifications obtained at a young age).
4. Many other studies have used comparable statistical approaches to scale educational categories according to their relationship with occupational categories for the same individuals (e.g. Wong 2010; Clogg and Shihadeh 1994; Duncan-Jones 1972).
5. Re-expressing the results of the same data, we could say that whilst 3.8% of the female partners of all cohabiting males had attended private school, 3.2% of cohabiting males who had not attended a private school were nevertheless living with a privately educated partner, compared to 21.6% of those males who were privately educated.

## Bibliography

Arum, R., Gamoran, A., & Shavit, Y. (2007). More Inclusion Than Diversion: Expansion, Differentiation and Market Structure in Higher Education. In Y. Shavit, R. Arum, & A. Gamoran (Eds.), *Stratification in Higher Education: A Comparative Study* (pp. 1–35). Stanford, CA: Stanford University Press.

Attewell, P., & Lavin, D. E. (2007). *Passing the Torch: Does Higher Education for the Disadvantaged Pay Off Across the Generations?* New York: Russell Sage Foundation.

Bolton, P. (2012). *Education: Historical Statistics*. London: House of Commons Library, Standard Note SN04252. Retrieved January 1, 2016, from http://researchbriefings.files.parliament.uk/documents/SN04252/SN04252.pdf.

Brown, P., & Scase, R. (1995). *Higher Education and Corporate Realities: Class, Culture and the Decline of Graduate Careers*. London: UCL Press.

Buis, M. L. (2010). *Inequality of Educational Outcome and Inequality of Educational Opportunity in the Netherlands During the 20th Century*. Amsterdam: VU University Amsterdam.

Bukodi, E., & Goldthorpe, J. H. (2013). Decomposing 'Social Origins': The Effects of Parents' Class, Status and Education on the Educational Attainment of Their Children. *European Sociological Review, 29*(5), 1024–1039.

Clogg, C. C., & Shihadeh, E. S. (1994). *Statistical Models for Ordinal Variables*. Thousand Oaks, CA: Sage.

Connelly, R., Gayle, V., & Lambert, P. S. (2016). A Review of Educational Attainment Measures for Social Survey Research. *Methodological Innovations Online*. https://doi.org/10.1177/2059799116638001.

Duncan-Jones, P. (1972). Social Mobility, Canonical Scoring and Occupational Classification. In K. Hope (Ed.), *The Analysis of Social Mobility* (pp. 191–210). Oxford: Oxford University Press.

Ermisch, J., Jantti, M., & Smeeding, T. (Eds.). (2012). *From Parents to Children: The Intergenerational Transmission of Advantage*. New York: Russell Sage Foundation.

Field, J. (2015). Social Ties, Agency and Change: Education and Social Capital in Adult Life. In Y. Li (Ed.), *Handbook of Research Methods and Applications in Social Capital* (pp. 292–306). Cheltenham: Edward Elgar.

Griffiths, D., & Lambert, P. S. (2012). Dimensions and Boundaries: Comparative Analysis of Occupational Structures Using Social Network and Social Interaction Distance Analysis. *Sociological Research Online, 17*(2), 5.

Hallsten, M. (2013). The Class-Origin Wage Gap: Heterogeneity in Education and Variations Across Market Segments. *British Journal of Sociology, 64*(4), 662–690.

Halsey, A. H. (1988). Higher Education. In A. H. Halsey (Ed.), *British Social Trends Since 1900* (pp. 268–296). London: Macmillan.

Hoffmeyer-Zlotnik, J. H. P., & Warner, U. (2014). *Harmonising Demographic and Socio-Economic Variables for Cross-National Comparative Survey Research*. Berlin: Springer.

Hout, M., & DiPrete, T. A. (2006). What We Have Learned: RC28s Contributions to Knowledge About Social Stratification. *Research Into Social Stratification and Mobility, 24*, 1–20.
Khan, S. (2011). *Privilege: The Making of an Adolescent Elite at St Paul's School*. Princeton, NJ: Princeton University Press.
Lambert, P. S. (2012). *Comparative Scaling of Educational Categories by Homogamy—Analysis of UK Data from the BHPS*. Technical Paper 2012-1 of the DAMES Node, Data Management Through e-Social Science. Stirling: University of Stirling. Retrieved from www.dames.org.uk/publications.html.
Laurison, D., & Friedman, S. (2016). The Class Pay Gap in Higher Professional and Managerial Occupations. *American Sociological Reiew, 81*(4), 668–695.
Lupton, R., Heath, N., & Salter, E. (2009). Education: New Labour's Top Priority. In J. Hills, T. Sefton, & K. Stewart (Eds.), *Towards a More Equal Society? Poverty, Inequality and Policy Since 1997* (pp. 71–90). Bristol: Policy Press.
Mirowsky, J., & Ross, C. E. (2003). *Education, Social Status and Health*. Hawthorne, NY: Aldine de Gruyter.
Modood, T. (2005). The Educational Attainments of Ethnic Minorities in Britain. In G. C. Loury, T. Modood, & S. M. Teles (Eds.), *Ethnicity, Social Mobility and Public Policy: Comparing the US and UK* (pp. 288–308). Cambridge: Cambridge University Press.
Pfeffer, F. T. (2008). Persistent Inequality in Educational Attainment and Its Institutional Context. *European Sociological Review, 24*(5), 534–565.
Portes, A., & Rumbaut, R. G. (2001). *Legacies: The Story of the Immigrant Second Generation*. Berkeley and Los Angeles: University of California Press.
Prandy, K., Unt, M., & Lambert, P. S. (2004). *Not by Degrees: Education and Social Reproduction in Twentieth-Century Britain*. Paper Presented at the ISA RC28 Research Committee on Social Stratification and Mobility Neuchatel, 7–9 May 2004.
Raffe, D., & Croxford, L. (2015). How Stable Is the Stratification of Higher Education in England and Scotland? *British Journal of Sociology of Education, 36*(2), 2015.
Savage, M., Cunningham, N., Devine, F., Friedman, S., Laurison, D., McKenzie, L., Miles, A., Snee, H., & Wakeling, P. (2015). *Social Class in the 21st Century*. London: Pelican Books.
Savage, M., Devine, F., Cunningham, N., Taylor, M., Li, Y., Hjellbrekke, J., Le Roux, B., Friedman, S., & Miles, A. (2013). A New Model of Social Class: Findings from the BBC's Great British Class Survey Experiment. *Sociology, 47*(2), 219–250.

Schneider, S. L. (2008). *The International Standard Classification of Education (ISCED-97). An Evaluation of Content and Criterion Validity for 15 European Countries*. Mannheim: MZES.

Schuller, T., Preston, J., Hammond, C., Basset-Grundy, A., & Brynner, J. (2004). *The Benefits of Learning: The Impact of Education on Health, Family Life and Social Capital*. London: Routledge.

Shavit, Y., Arum, R., & Gamoran, A. (Eds.). (2007). *Stratification in Higher Education: A Comparative Study*. Stanford, CA: Stanford University Press.

Shavit, Y., & Blossfeld, H. P. (Eds.). (1993). *Persistent Inequality: Changing Educational Attainment in Thirteen Countries*. Boulder, CO: Westview.

Simon, J. L., & Boggs, R. (1995). Trends in the Quantities of Education: USA and Elsewhere. In J. L. Simon (Ed.), *The State of Humanity* (pp. 208–223). Oxford: Blackwell.

Stewart, A., Prandy, K., & Blackburn, R. M. (1980). *Social Stratification and Occupations*. London: Macmillan.

Ultee, W. C., & Luijkx, R. (1990). Educational Heterogamy and Father-to-Son Occupational Mobility in 23 Industrial Nations: General Societal Openness or Compensatory Strategies of Reproduction. *European Sociological Review, 6*, 125–150.

University of Essex, & Institute for Social and Economic Research. (2010, July). *British Household Panel Survey: Waves 1–18, 1991–2009 [Computer File]* (7th ed.). Colchester, Essex: UK Data Archive [distributor], SN: 5151.

van de Werfhorst, H. G., de Graaf, N. D., & Kraaykamp, G. (2001). Intergenerational Resemblance in Field of Study in the Netherlands. *European Sociological Review, 17*(3), 275–293.

van de Werfhorst, H. G., & Luijkx, R. (2010). Educational Field of Study and Social Mobility: Disaggregating Social Origin and Education. *Sociology, 44*(4), 695–715.

Willis, P. (1977). *Learning to Labour: How Working Class Kids Get Working Class Jobs*. Westmead, Hants: Saxon House.

Wong, R. S. K. (2010). *Association Models*. Los Angeles: Sage.

# 11

# Exploiting Non-standard Dimension Scores and Network Structures in the Analysis of Social Interactions Between Occupations

## 11.1 Introduction

The SID analytical approach involving occupations summarises social distance patterns through statistical dimension scores (e.g. Chaps. 4, 5, and 6). Other approaches of network analysis can also be used to summarise similar data on social interactions (e.g. Chaps. 7 and 8). The usual focus is on the most important elements of the interaction structure—for example, the first and most important dimension of the SID solution, or the most influential patterns of network connections. Previously, we have also looked at social interaction data for large populations, such as countries. However, there are some interesting extensions in the options for calculating and exploring dimension scores and network structures. In this chapter, we discuss making more use of 'subsidiary' dimensions from the SID solution (Sect. 11.2), undertaking SID analysis on smaller population groups of particular interest (Sect. 11.3). and ways of using other network analysis tools to study sub-populations of special interest (Sect. 11.4).


P. Lambert, D. Griffiths, *Social Inequalities and Occupational Stratification*,
https://doi.org/10.1057/978-1-137-02253-0_11

## 11.2 Using 'Subsidiary Dimensions' from the SID Solution

In the social interaction distance analytical tradition, we have already mentioned that there are structures to social interactions between occupations that do not seem to reflect social stratification. 'Pseudo-diagonals' (see Sect. 6.7) are particular combinations of occupations that are disproportionately likely to have social connections between them, but for reasons other than their similarity in social stratification position—such as workplace contiguity or 'situs'. In addition, 'subsidiary dimensions' in the SID solution typically reflect more general supplementary patterns of social interactions that are not tied down to just a few specific occupations.[1] In Chap. 6, we commented that subsidiary dimensions sometimes reflect clearly interpretable structures, and that their properties can depend upon the level of occupational detail that is exploited (Sect. 6.5.2).

Although subsidiary dimensional patterns have not been extensively exploited in previous studies, there are at least three good reasons for studying them further. First, there could be exploratory, inductive value in calculating and examining them—particularly if they suggest social structures linked to occupations that we would not otherwise have anticipated (see Sect. 11.2.1). Second, subsidiary dimension structures are often connected to socio-demographic differences such as of gender, age, or ethnicity, so it is possible that statistical parameter estimates linked to socio-demographic measures could be spurious without suitable control for the social process that is reflected in the subsidiary dimensional structure (see Sect. 11.2.2). A third consideration is that if subsidiary dimensions reveal social structures in occupations that are genuinely 'exogenous' to the stratification structure that we are primarily interested in, then we might be able to use subsidiary dimensions to isolate these different mechanisms more thoroughly from an analysis that focusses upon social stratification (see Sect. 11.2.3).

### 11.2.1 Exploring Subsidiary Dimension Structures

There are some easily recognisable social processes that have often been identified as subsidiary dimensions in the SID approach—for instance, the influence of gender segregation in occupations (e.g. Chan and

Goldthorpe 2004), of industrial and institutional structure (e.g. Levine and Spadaro 1988; Blau and Duncan 1967), and divisions related to employment status (e.g. Laumann and Guttman 1966). The gender segregation dimension—typically running from male-dominated, through mixed, to female-dominated jobs[2]—is usually apparent whenever the social interaction data involves the occupations of men and women (e.g. Chan and Goldthorpe 2004; Prandy and Lambert 2003). Additionally, in societies with any substantial agricultural sector, another common subsidiary dimension structure is associated with the 'agriculturality' of the occupation (in societies with many jobs in agriculture, this dimension is sometimes the most important statistically—e.g. Lambert et al. 2013). Explanations for the social distances associated with these dimensions are usually obvious. Consider a male in a job that is male dominated and is linked to the agricultural sector (e.g. the job of gamekeeper). Here, two things about their spouse's occupation will be more probable than average: their spouse probably doesn't work in a male-dominated occupation, but is relatively more likely to be in a female-dominated or mixed occupation; and their spouse probably has a job that is commonly found in rural areas. Whilst some such structural patterns in interactions between occupations might be easy to anticipate *a priori*, in other scenarios it is plausible that dimension reduction techniques are the only tractable way to identify a parsimonious representation of subsidiary phenomena that would otherwise go unrecognised.

Figure 11.1 provides a representation of the occupations at the extremes of the first five dimensions of a SID solution for data on marriage partnerships in contemporary Britain. We consider that the first dimension reflects the core force of social stratification (i.e. this is the dimension that is usually focussed upon and that lies behind a CAMSIS scale). However, the appropriate interpretation of the subsidiary dimensions is more ambiguous. The second dimension, to us, seems to reflect an arts/science division in the composition of occupations—in the UK context, this is quite a plausible social division because it is ingrained in UK educational pathways (e.g. Playford and Gayle 2016), although it might not be a division that translates consistently to other nations. We would see the third dimension from Fig. 11.1 as that of 'agriculturality', a structure that is very commonly observed in subsidiary dimensions of a SID analyses. We

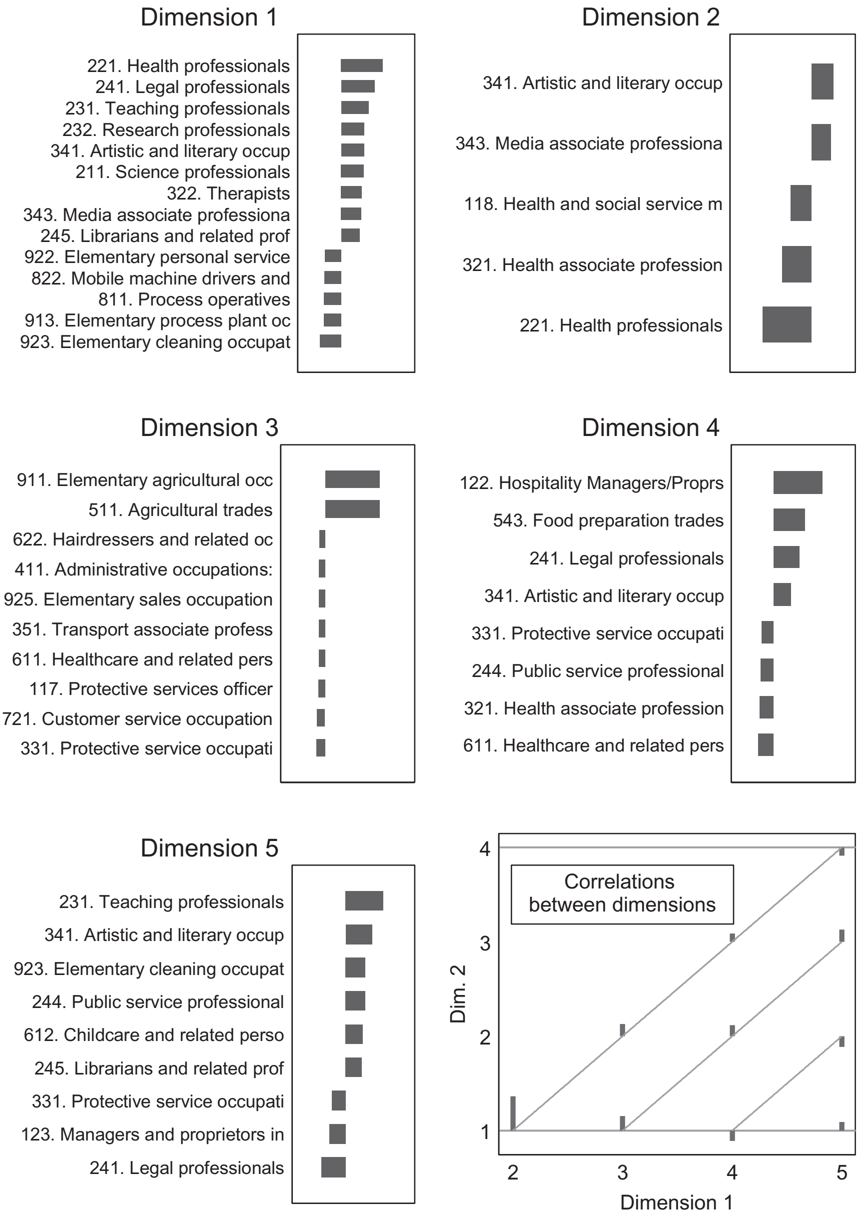



**Fig. 11.1** Outlying occupations amongst the five subsidiary dimensions in the SID solution for heterosexual couples, UK 2001–2010. Source: 128k married couples from UK Labour Force Survey, 2001–2010. Bars indicate selected highest positive or negative scores in each dimension

suspect that the fourth dimension reflects a public/private sector division in the UK. Lastly, the fifth dimension from Fig. 11.1 seems to be that of gender segregation in occupations, a scale that moves broadly from female-dominated, through mixed, to male-dominated occupations. In this example, it is perhaps only the structures of subsidiary dimensions 3 and 5, of agriculturality and gender segregation that would have been readily anticipated in advance. The other factors might well be plausible social patterns in occupations that are worth measuring and controlling for, but we might not have considered or identified them if we did not use the 'inductive' approach of the SID analysis.

It is worth remembering that the lower-order dimensions are calculated as statistically orthogonal to earlier dimensional structures, a property that has some implications for interpretation of the scores. If the scores in each dimension are then assigned to individuals on the basis of their occupations, they will usually have a fairly low correlation (the correlations amongst the subsidiary dimensions from Fig. 11.1 range in magnitude from 0.086 to 0.131).[3] At the same time, the various subsidiary dimensions might not correlate perfectly with other direct measures of the same concepts (such as gender segregation). This occurs because the subsidiary dimensions represent a structure after the influence of any higher-order dimensions has already been 'removed' (whereas direct measures of the same phenomena would not seek to discount any higher-order influences). This means that the subsidiary dimensions have an interesting but nuanced social meaning: they can be thought of as special representations of the social structure that they are associated with, but net of relevant higher-order structures. For example, we could think of the fifth dimension in Fig. 11.1 as a measure of gender segregation in occupations that has already been purged of those aspects of gender segregation that are linked to social stratification inequality (dimension 1), the arts/science division (dimension 2), agriculturality (dimension 3), and employment sector (dimension 4). This empirical measure is something subtly different to a direct measure of gender segregation and is worth attention in its own right.

### 11.2.2 Using Subsidiary Dimension Structures to Improve Inferences About Socio-demographic Inequalities

It is possible that subsidiary dimension structures may represent important aspects of social relations that should usually be controlled for in other analysis. In particular, subsidiary dimensional structures are often correlated to socio-demographic differences between individuals, so it might be important to control for them when studying socio-demographic inequalities. The empirical relevance of one common subsidiary dimension—gender segregation in occupations—is well recognised. Although it does not follow automatically, empirical statements about gender differences might be misleading if they don't factor in differences associated with gender segregation in occupations (e.g. Blackwell 2003; van der Lippe and van Dijk 2002).

Tables 11.1 and 11.2 show analyses that provide modest evidence on the relative benefit of modelling subsidiary dimension structures when analysing socio-demographic inequalities. The tables show results from two different

**Table 11.1** Influences on the probability of having a permanent job (UK, 2007)

| | Model 11.1a | (+ occ r.e.'s) | Model 11.1b | (+ occ r.e.'s) | Model 11.1c | (+ occ r.e.'s) |
|---|---|---|---|---|---|---|
| Female | 0.09 | –0.07 | 0.19 | –0.04 | 0.29 | 0.06 |
| Age in years | –0.04* | –0.03* | –0.04* | –0.03* | –0.04* | –0.03 |
| Education score | –0.05* | –0.03* | –0.02 | –0.02* | –0.02 | –0.02* |
| SID dimension 1 (CAMSIS) | | | –0.02* | –0.02* | –0.02* | –0.02* |
| SID dimension 2 | | | | | 0.00 | 0.00 |
| SID dimension 3 | | | | | –0.01* | –0.01 |
| SID dimension 4 | | | | | –0.02* | –0.01 |
| SID dimension 5 | | | | | –0.01* | –0.03* |
| N | 6129 | 6229 | 6129 | 6229 | 6129 | 6229 |
| Log-likelihood | –1134 | –1098 | –1121 | –1095 | –1105 | –1085 |
| Pseudo-r2 | 0.033 | | 0.044 | | 0.057 | |
| Occupation level ICC | | 11.1%* | | 9.1%* | | 4.9%* |

Source: Analysis of the UK BHPS, wave 18 (2007). 6229 adult respondents aged 30–75 with valid data on all relevant measures. *N* = 6229 but left-hand models use individual-level sampling weights. Numeric values of SID scores and education score are mean 50, SD 15. Light-grey columns refer to equivalent models with random effects for clustering into occupational minor groups

*Indicates statistically significant at 95% threshold

**Table 11.2** 'Ethnic penalties' as influences upon personal income amongst the economically active in the UK (2001–2010)

| | Model 11.2a | (+ occ r.e.'s) | Model 11.2b | (+ occ r.e.'s) | Model 11.2c | (+ occ r.e.'s) |
|---|---|---|---|---|---|---|
| Female | –0.63* | –0.42* | –0.68* | –0.42* | –0.63* | –0.42* |
| Age in years | 0.08* | 0.06* | 0.07* | 0.06* | 0.07* | 0.06* |
| Age-squared (coef*100) | –0.09* | –0.07* | –0.08* | –0.07* | –0.08* | –0.07* |
| Married/Cohabiting | –0.14* | –0.14* | –0.14* | –0.14* | –0.14* | –0.14* |
| Education score | 0.03* | 0.01* | 0.01* | 0.01* | 0.01* | 0.01* |
| Ethnic group: | | | | | | |
| Other white | 0.09* | 0.09* | 0.08* | 0.09* | 0.10* | 0.09* |
| Mixed | 0.03 | 0.05* | 0.02 | 0.05* | 0.03 | 0.05* |
| Indian | 0.04* | 0.01 | 0.01 | 0.01 | –0.01 | 0.01 |
| Pakistani | –0.28* | –0.16* | –0.26* | –0.16* | –0.22* | –0.16* |
| Bangladeshi | –0.63* | –0.31* | –0.54* | –0.31* | –0.41* | –0.31* |
| Chinese | –0.15* | –0.07* | –0.13* | –0.07* | –0.10* | –0.07* |
| Other Asian | –0.12* | –0.01 | –0.06* | –0.01 | –0.04 | –0.01 |
| Black Caribbean/African | –0.05* | 0.05* | –0.01 | 0.05* | 0.02 | 0.05* |
| Other | –0.10* | 0.00 | –0.09* | 0.00 | –0.03 | 0.00 |
| SID dimension 1 (CAMSIS) | | | 0.02* | 0.02* | 0.02* | 0.02* |
| SID dimension 2 | | | | | 0.001* | 0.001 |
| SID dimension 3 | | | | | –0.002* | –0.003* |
| SID dimension 4 | | | | | –0.006* | –0.004 |
| SID dimension 5 | | | | | –0.009* | –0.010* |
| r2 | 0.297 | | 0.387 | | 0.424 | |
| Occupation level ICC | | 31.8%* | | 19.5%* | | 15.9%* |

Notes: Analysis of the UK Labour Force Survey (pooled data from years 2001–2010). Y variable is log of income (inflation adjusted to 2010 equivalent). *N* = 108674 (aged 25–65 with valid data on all variables). Light-grey columns refer to equivalent models with random effects for clustering into occupational minor groups

*Indicates statistically significant at 95% threshold

statistical models that explore influences upon the probability of having a permanent job, and income levels, respectively, for adults in the UK. Of interest, here is whether there are any consequential differences in results when we add controls for measures of positions in the subsidiary dimensions of the SID solution—the difference in broad terms between models labelled 'b' which do not use subsidiary dimension scores (i.e. models 11.1b and 11.2b), and those labelled 'c' which do use them (i.e. models 11.1c and 11.2c). There are some modest patterns of difference, albeit more pronounced for influences upon holding a permanent job than for influences upon income. We see, for instance, a considerable change in the estimated influence of gender upon having a permanent job when we control for subsidiary dimension locations, and we see a modest change in the influence of dummy variables that might depict 'ethnic penalties' on income.

Our point generally is that controls for the structures that are depicted by the SID subsidiary dimensions are not usually estimated in comparable statistical studies, yet it is possible that they are relevant influences that ought to be controlled for. There are complexities of interpretation in using subsidiary dimensions, and there might sometimes be a risk of 'over-fitting' the data (for instance, if several subsidiary dimensions are built into a model and some of the dimensions hinge upon differences amongst a relatively small number of cases). Nevertheless, Tables 11.1 and 11.2 suggest that there are some scenarios where an analysis that does not control for SID subsidiary dimension scores might result in 'omitted variable bias' in parameter estimates.

The issue is particularly pertinent to those styles of statistical analysis that seek to summarise the net effect of a socio-demographic differences after relevant controls—a prominent example in the UK is research that seeks to estimate 'ethnic penalties' that constitute the net effect of ethnic group membership after controlling for other social inequality patterns (e.g. Khattab 2016, and cf. Table 11.2). In these situations, the use of appropriate 'control' measures is critical to the suitable interpretation of the final results, yet Table 11.2 suggests that we might reach different conclusions about ethnic penalties if we ignore controls for subsidiary dimension structures from the social interaction distance solution. In this way, the use of inductively derived subsidiary dimension structures might provide otherwise unanticipated plausible 'controls' that could substantially change our understanding of social inequalities.

### 11.2.3 Using Subsidiary Dimensions to Disentangle Separable Mechanisms from the Influence of Social Stratification

Perhaps the most important way in which subsidiary dimensional structures could improve our analysis of social inequality is if they can help us to 'disentangle' social patterns that are not about the forces or mechanisms of 'social stratification', from those that are. This ought to be possible, in principle, since the subsidiary dimension structures should represent relevant social patterns that are defined orthogonally to the stratification dimension.

As described in Chap. 3, there are many alternative occupation-based measures in common use in social science research. Our hypothesis is that interpretations of a SID measure of stratification should not be greatly perturbed by controls for subsidiary dimensions (since the first dimension was designed to be orthogonal to the subsidiary structures), whereas interpretations linked to other measures will be more substantially affected. Arguably, if interpretations of popular occupation-based measures of stratification are indeed perturbed by whether we control for subsidiary dimensions, this might be evidence that the measures conflate empirical influences (of stratification, and of the other dimensional mechanisms) in a manner that is not deliberate and might lead to spurious theoretical interpretations.

The empirical evidence provides some limited support for our hypothesis. Table 11.3 focusses upon five influential occupation-based measures of stratification: a SID measure, the ISEI scale, and the EGP class scheme represented by three commonly used variants (using 2, 7, and 11 categories). The first panel shows that the four subsidiary dimensions account, between them, for modest proportions of the variance in the respective stratification measures. On the one hand, this means that, without controls, each prospective measure of social stratification position might also represent some patterns linked to subsidiary dimensions (i.e. unintentionally). On the other hand, the pattern doesn't confirm our expectation that the SID scale should have a markedly weaker association to the subsidiary dimension structures; on the contrary, all of the measures have comparably modest associations to the combined subsidiary dimensions.

Table 11.3 then summarises the extent to which the results from statistical models that do and do not control for subsidiary dimensions differ in the interpretation given to the stratification measure. The models show the relative influence of the occupation-based measures upon a selection of individual-level indicators of factors that are known to have some correlation to social stratification position (measures of attitudes, educational levels, and material well-being). Similarly to the results reported in Sect. 11.2.2, there is a case for paying attention to the subsidiary dimension structures because, for each of the stratification measures, we can point to

**Table 11.3** Relation between stratification measures and subsidiary dimension scores, and variations in model parameters for the influence of occupation-based social classifications, with and without controls for the first four subsidiary dimensions of a SID structure

| | CAMSIS | ISEI | EGP2 | EGP7 | EGP-11 |
|---|---|---|---|---|---|
| R2 for dimensions 2–5 in predicting... | 0.187 | 0.235 | 0.084 | 0.144 | 0.129 |
| | 95% CI's after controls for subsidiary dimensions | | | | |
| | *95% CI's before controls for subsidiary dimensions* | | | | |
| 1. Subjective optimism | 19–27 | 17–23 | 45–65 | –55–93 | 52–87 |
| | *18–24* | *14–19* | *45–64* | *14–47* | *13–58* |
| 2. Liberal attitude to homosexuality | 22–29 | 16–22 | 52–72 | 5–150 | 65–98 |
| | *22–29* | *15–20* | *54–73* | *27–59* | *24–67* |
| 3. Education score | 30–32 | 23–25 | 67–74 | 24–84 | 111–124 |
| | *30–33* | *20–22* | *68–75* | *23–37* | *81–98* |
| 4. Household income quintile | 34–41 | 29–35 | 92–110 | –40–101 | 7–10 |
| | *33–39* | *29–34* | *98–115* | *13–41* | *13–17* |
| 5. Personal income quintile | 63–73 | 55–62 | 18–20 | –66–82 | 17–20 |
| | *58–66* | *49–55* | *17–20* | *18–48* | *25–30* |
| 6. Doesn't smoke regularly | 23–24 | 15–23 | 30–56 | 30–238 | 62–104 |
| | *26–36* | *16–24* | *38–63* | *65–105* | *0–56* |
| 7. Doesn't live in social housing | 55–77 | 38–53 | 11–15 | –17–16 | 12–18 |
| | *59–76* | *42–55* | *12–16* | *7–12* | *6–14* |

Source: Analysis based upon individual data from the BHPS, 2008, *N* ~ 7000 adults in current employment. Figures in lower panel are 95% confidence intervals around an estimated regression coefficient for the influence of an occupation-based measure (multiplied by 10 to simplify rounding). Models for outcomes 1–2 and 4–5 use ordered logit (five categories); model 3 uses linear regression; models 6–7 use logistic regression. Models feature controls for gender and age (quadratic) and whether individuals are cohabiting (except for model 4 with no controls, and model 3 without cohabiting status control). Four SID subsidiary dimensions used are those for dimensions 2–5 from LFS analysis as in Fig. 11.1.

small differences in estimated regression parameters for the measure, depending upon whether or not we have controlled for subsidiary dimensions. Some of these differences are enough to cross the boundaries of the confidence intervals to the point that would change the conclusions drawn—for instance, the majority of estimates for the influence of the EGP-11 dummy variable before and after controls have confidence intervals that don't overlap. That there is any difference at all suggests that direct measures of subsidiary dimensions do indeed have the potential to help disentangle stratification effects from other influences. At the same time however the patterns are not as compelling as they might hypothetically have been. The scale of the impact of controlling for subsidiary dimension structures is modest rather than dramatic. Furthermore, we do not particularly see a strong premium to the SID measure compared to the other occupation-based measures—its parameters are perturbed to a comparable extent as are those based on the other measures. This suggests that SID measures are not intrinsically any better at disentangling the influences of subsidiary structures as we had hypothesised: the case for using subsidiary dimension measures as ways of telling more refined stories about social stratification patterns is broadly comparable regardless of which occupation-based measures are being used.

## 11.3 Socio-demographic Variations in Social Interaction Distance Solutions

The hypotheses of 'universality' and 'specificity' in occupation-based measures contrast whether the same occupations may have the same ('universal') or different ('specific') relative social meanings in different societies (e.g. Lambert et al. 2008). The so-called Treiman constant (Hout and DiPrete 2006) reflects evidence generated by Treiman and others that occupational roles hold roughly the same relative social positions across countries and time points (e.g. Ganzeboom and Treiman 2003; Treiman 1977; Hodge et al. 1967), and is widely taken as evidence for 'universality'. Nevertheless, counter-evidence is occasionally presented (e.g. Lambert et al. 2008), and we have mentioned already that most SID analysis take a 'specific' approach, insofar as they estimate different SID scales for different countries and time periods (e.g. Chan 2010; Prandy and Jones 2001).

A special version of the 'specificity' hypothesis is the idea that it might be useful to construct different occupation-based measures for different sections of a society. This is occasionally undertaken for one population division, that of gender (see Sect. 6.8; Prandy 1986; Martin and Roberts 1984). However, there are many other socio-demographic differences where we might plausibly imagine that different structural relationships with occupations operate—for instance, for different ethnic groups, birth cohorts, regions, or for other important social groups. Social relationship make a particularly interesting example—it is potentially difficult to define 'specific' sub-populations, because social relationships will often cross between groups in a non-systematic way, yet it is very easy to imagine that social connections between occupations involve different patterns within different social groups (e.g. Alderson et al. 2007).

Although the construction of different SID scales for different social groups has an appealing logic, there are several operational limitations. Firstly, social groups are not socially discrete, so in order to find social interaction dimensions for members of a group (for instance, females), some inclusion criteria must be used. For instance, it is plausible to select only those social interactions in which both ego and alter are part of the social group of interest (e.g. women and their female friends), but this will often mean ignoring relevant records on other social interactions (e.g. women and their male friends). Alternatively, we can select all pairs of occupations when either the ego or alter is part of the group (e.g. friendships involving at least one woman), though this means that some individuals from outside the group contribute to the analysis. This is our approach below when summarising social interactions for different ethnic groups (Fig. 11.2 shows SID patterns for married couples when either husband or wife is part of the relevant ethnic group). We can also consider weighted combinations of the above (e.g. a smaller weight if only one member of the pair is part of the relevant group). In special cases, when the social group it part of the definition of the relationship, we can separate the groups by separating the rows or columns of the social interaction matrix (e.g. when studying social interactions between males and females). Secondly, the SID approach is likely to work best when large datasets can be used that allow us to disaggregate many different occupational categories, and the case for using detailed occupational units is at its strongest

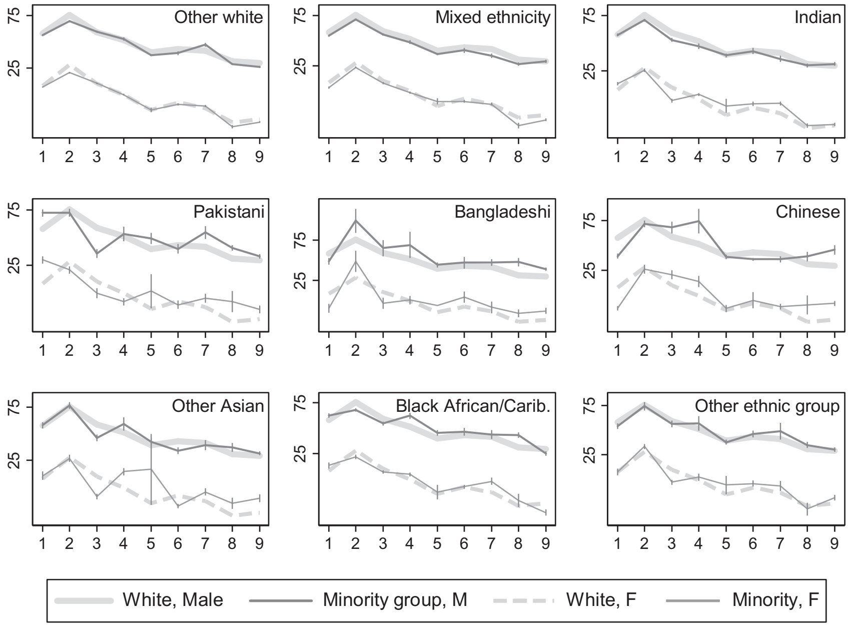



**Fig. 11.2** Depiction of different SID scales for different ethnic groups in the UK. Source: Analysis of UK Labour Force Survey, male-female couples, 2001–2010. Horizontal axis = UK SOC 2000 major occupational group. Vertical axis = CAMSIS scale specific to ethnic group and gender (female score shown is value—50 for ease of presentation). Vertical lines show approximated 95% confidence intervals for minority group score (values for White group are negligible)

when we are interested in specific socio-demographic groups (because socio-demographic characteristics such as gender, ethnicity, and age are all associated with unequal occupational distributions). However, only the largest datasets are likely to allow us to have a good representation of many different occupations across different socio-demographic groups. Many CAMSIS scales have been effective in characterising SID dimensions for the distribution of male and female jobs separately, but if we focus on smaller age groups or ethnic minorities, we can expect numerous occupational positions to be sparsely represented, leading to suboptimal aggregations, high margins of error, and inconclusive findings.

Despite the complications, there are some examples where it has proven feasible and interesting to generate SID scales for different social groups. During the construction of a CAMSIS scale for the USA in earlier research, we reported an exploratory SID analysis for 'White' and separately 'Black' Americans that led to separate scales that shared similar features but included a few differences that were substantively plausible in the context (Prandy et al. 2002). White bus drivers, for example, were given a relatively lower SID score (relative to other White males) than Black bus drivers—this seems plausible if we consider that this job is popularly associated with Black male workers in the USA, but Black males are generally found in less advantaged occupational positions across the country.

For the contemporary UK, Fig. 11.2 summarises a range of SID scores for different ethnic groups from the UK. There are some examples of different scores being given to the same occupations in different ethnic groups, many of which are substantively plausible—for example, Pakistani men in 'associate professional' groups seem to be in worse relative positions than other men in the same occupations, which might make sense if we consider that many Pakistani men in the UK work in relatively insecure and low-paid technical professions such as the mobile phone industry. However, this example is somewhat unsatisfactory, because the difference in the major group profiles could well reflect underlying compositional differences in the occupations within the major group—in this case, we used highly aggregated occupational categories due to low volumes of data, but this compromise may well impede our ability to draw more useful conclusions.

Although hitherto there have been few other relevant implementations, there seem to us to be many rich analytical opportunities for disaggregating SID analyses between different social groups. As exploratory activities, these might reveal patterns of inequality that would otherwise be overlooked. Analytically, there is clearly potential for refined statistical results to emerge from studies that allow other socio-demographic contextualisations to be built into analytical tools. Practical constraints in terms of the number of cases available for particular sub-groups suggest an enduring tension between disaggregation of occupational units and disaggregation of socio-demographic groups, but similar constraints might diminish in future as larger and richer data resources accumulate.

## 11.4 Social Network Analysis and Demographic Groups

Just as it is plausible to estimate SID scales for distinctive social groups within a wider population, we could similarly apply exploratory network analysis approaches to different groups. When the difference is binary, for instance, the male-female division, the adjustment could potentially be summarised by assessing 'directed' rather than 'undirected' ties between occupations (Sect. 11.4.1). Alternatively, we could consider depicting the full network structure within a particular social group (making comparisons to others) (Sect. 11.4.2).

### 11.4.1 Directed Ties for Male-Female Relationships

Because we were primarily interested in the occupational connections, the network analyses shown in Chap. 8 were presented in an 'undirected' way. If a tie was drawn between two occupations, this meant that either the ego-alter, or the alter-ego, tie between the two occupations was unusually common (or that both combinations were unusually common). Remember however that many of the datasets on ties are naturally 'directed', in that there is a meaningful difference between ego-alter and alter-ego combinations. For example, when egos and alters are defined as males and females from married couples, it may be relevant to distinguish between male-female and female-male linkages in further analysis.

Information on the direction of ties can be readily incorporated into most network depictions, if the relationship is structured in a relevant way. Figure 11.3, for instance, uses arrows to show the nature of the relationship for networked occupations in the USA in 2000 (using the 'microclass' occupational taxonomy). Following the terms used in Chap. 8, the sociogram depicts those combinations of networked occupations that occurred at least twice as often as would be expected if social ties were distributed by chance, and the diagram only plots nodes and ties if at least 15 cases represented the relevant combination. The arrows in Fig. 11.3 point to the female occupation in the male-female combination.[4] The size of the cells is proportional to the CAMSIS scale score for the occupations,

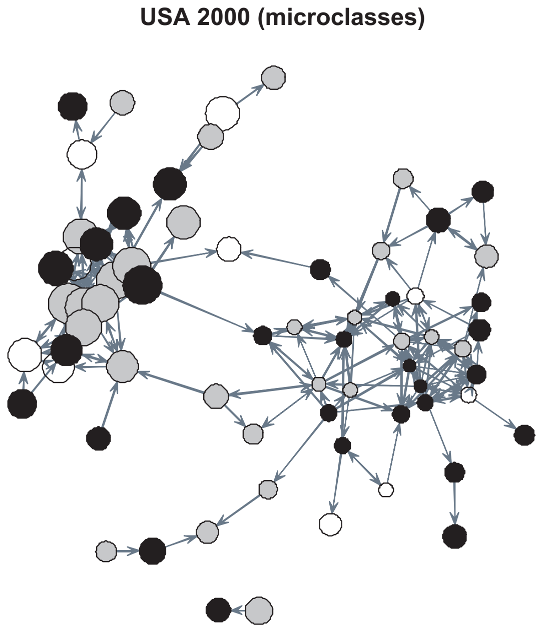



**Fig. 11.3** USA 2000 networked occupations as directed ties. Source: IPUMS-I data for USA 2000. Arrows indicate that the link occurs from males to females: the arrow points to the female occupation for the relevant over-represented combination. The nodes are shaded white for female-dominated occupations (at least 75% female), black for male-dominated occupations (at least 75% male), and grey for mixed occupations. Node size is proportional to the CAMSIS scale score (male scale) for the occupation

and because we anticipate that gender segregation will be related to the volume of male-female ties, the occupations in Fig. 11.3 are also shaded in a way which indicates whether the occupation is male dominated (black), female dominated (white), or mixed gender (grey) (cf. Hakim 1998). The figure reiterates the importance of social stratification position in influencing whether or not ties tend to be over-represented—in general, ties are much more likely to occur between occupations with similar CAMSIS scores. In this depiction, there might be some evidence of a boundary within the main network component, between segments of less advantaged and more advantaged occupations. Whilst some of the ties shown in the figure go from a male in a male-dominated job to a female in a female-dominated job, perhaps surprisingly there are also numerous exceptions, including ties from males in female-dominated occupations to females in male-dominated positions (to reiterate, these ties are relatively over-represented, but might not be common in the underlying population). Relatively few of the ties are reciprocal (indicated in the figure with a double-headed arrow)—that is, there are few scenarios where it is unusually common for the link to arise both from males to females and also from females to males. We should also reiterate that some ties could never occur (or never be reciprocated) as a function of the minimum threshold approach that we have used: for example, some male-dominated occupations contain so few women in our data that it would be impossible for a combination that involves women in that occupation to be defined as over-represented, irrespective of the size of the representation ratio.

The arrows in Fig. 11.3 can help us to see that gender segregation patterns alone are not the most important feature of the directed networked occupational structure. They can also help us understand particularly important ties within the network. Arguably, there are three key bridging routes evident within the network, that is, links that connect the relatively less advantaged nodes with the relatively more advantaged ones. One of these routes seems to involve a traditional gender segregation pattern: bridging connections that go from men in male-dominated jobs that are both more and less advantaged, to women in a female-dominated job that has an intermediate CAMSIS score (the specific connection involves women in the female-dominated 'non-medical technicians' microclass, who are unusually often married both to men who are natural scientists and also to men who are chemical process workers). However, the other two bridging routes do not involve a conventional gender segregation pattern—one group links several similarly advantaged mixed-gender occupations (the occupations involve several jobs that are linked to agricultural production and its administration); a third link hinges on unusually common connections that involve women in a male-dominated occupation (gardeners). If we were interested in promoting wider social connections or reducing social distances, this evidence might suggest that heterogeneity in social networks can emerge both within traditional gender-typed arrangements, and in alternative non-gender-typed roles.

Similar depictions of directed networks could help us to explore how other demographic differences relate to social connections and social inequalities. The use of directed ties is plausible if the demographic division can be summarised by a dichotomy in the network relationship—for instance, links from younger to older adults or involving two key ethnic categories. An explicit indication of the directed nature of ties in SNA exploratory sociograms is sometimes unimportant, insofar as it is generally easy to look up further details on any particular social connections (as in the examples shown in Chap. 8). However, the use of directed graph depictions could sometimes help draw attention to a key social difference within the network and foster interpretations and theories that might otherwise have been overlooked.

### 11.4.2 Network Comparative Analysis for Social Groups

Another way of extending the depiction of networks between occupations is to compare different network images for different social groups within a society. Section 8.4 showed one relevant example, when we explored networks of social connections within two specific US states, rather than for the USA as a whole. Two further examples are discussed below—the possibility of drawing different network diagrams for the social connections between men and women, as in Figs. 11.4 and 11.5; and using separate diagrams for male-male connections for specific age groups within the UK, as in Fig. 11.6. Indeed, although not discussed further below, we could also consider depicting and comparing network ties between a deliberately reduced range of nodes—for instance, by showing networks between occupations when we limit analysis of people in public sector occupations only.

Figures 11.4, 11.5, and 11.6 are based on data on ego-alter links that can be identified in the UK's BHPS study (University of Essex 2010) on the basis of friendship and/or co-residence—we linked people of the same gender who had shared a household at any point in the survey, and used other data from the survey where respondents reported the occupations of their friends. This meant it was possible to construct sizeable

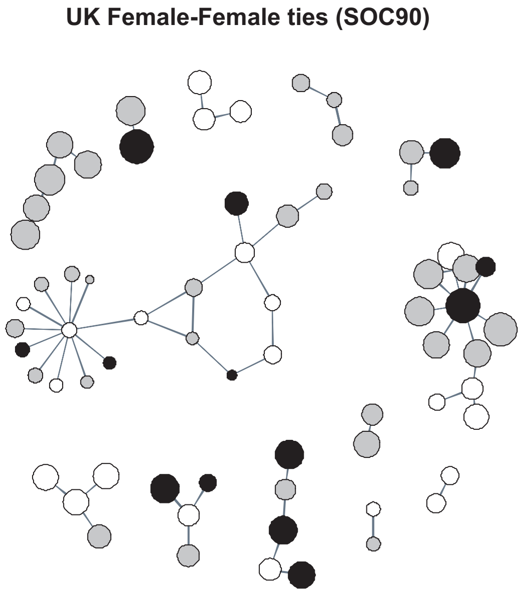



**Fig. 11.4** Female network (BHPS, all waves). Source: Female-female social interactions from BHPS dataset (1991–2008). Ties are shown if the combination has a representation ratio of 2 or more and occurs at least 15 times in the data

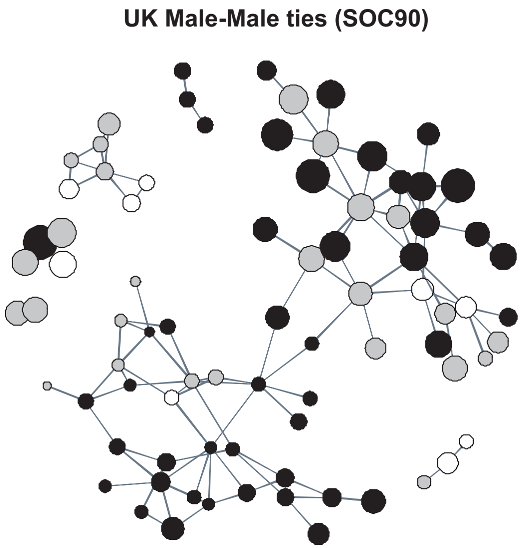



**Fig. 11.5** Male network (BHPS, all waves). Source: As Fig. 11.4

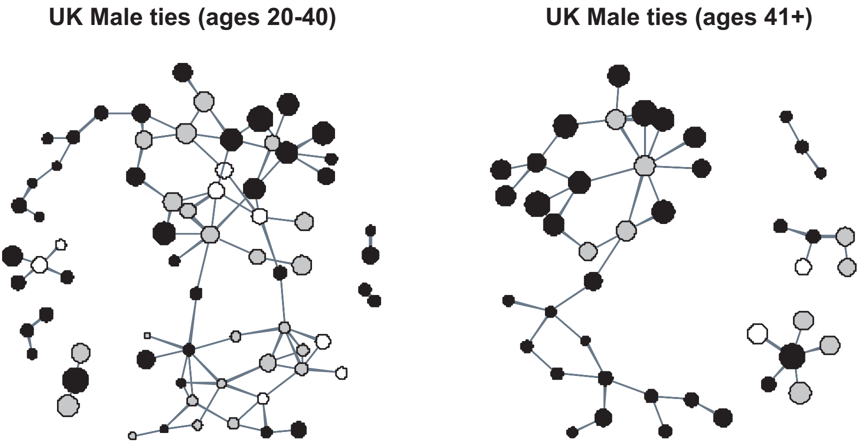



**Fig. 11.6** Networked occupations in the UK, by age groups. Source: Male-male ties from the BHPS. Ties are shown if the combination has a representation ratio of two or more and occurs at least seven times in the dataset

databases both of male-male social connections between occupations and, separately, of female-female connections. The sociograms also use the same coding scheme as in Fig. 11.3: the size of the node indicates the CAMSIS score for the occupation, and the colour of the node is related to the relative proportion of women in the job.

Without reviewing the images in extensive detail, the figures show that we could draw different insights into the structure of social connections between occupations by focussing upon different social groups. For men (Fig. 11.4) the most over-connected occupations mostly feature in one network component; the component is divided internally by stratification score, but there are some ties that connect between more and less advantaged occupations. For women (Fig. 11.5), there are many more separate components in social connection patterns—these are typically links between jobs that are of similar positions in the stratification structure and have similar male/female profiles, and (whilst the data is not shown in the diagram itself) the jobs that are within the same components typically fall within similar industries. At first sight, Figs. 11.4 and 11.5 might suggest that there is more social segregation between women than between men in social relationships linked to occupations, although we could consider a wider range of analyses, for instance, exploring different criteria for defining networked occupations and exploring similar patterns in other societies.

Figure 11.6 shows another potentially interesting pattern of social difference in network connections between occupations, although again the comparison could be taken much further than our illustrative example. The figure suggests some differences between younger and older men in the characteristics of their social connections between occupations—there are relatively fewer links involving men aged 41 and above, and most of the most over-represented links involve male-dominated jobs of similar positions in the stratification structure. Men aged 20–40 by contrast have a wider range of unusually common social connections between occupations, and these frequently involve jobs that are 'mixed gender' in composition, or even men who work in female-dominated occupations. It is tempting to conclude that these patterns suggest a generational shift—perhaps younger generations hold more diverse occupational positions and wider patterns of social connections. However, from this data alone, it is also plausible that the age difference represents an effect of aging (rather than of cohort change). An 'aging' process here arises if as men age they increasingly 'settle down' into more 'traditional' occupational roles (e.g. that are male dominated and that are more clearly defined in the stratification structure). Prospectively, in future work the evidence for either of these hypotheses could be resolved by using longitudinal career history data, and connecting that to data on social connections patterns.

## 11.5 Summary

In summary, the tools of analysis of social interactions between occupations provide us with several opportunities to identify and explore secondary structural patterns that are related to the social interactions between occupations. The dimension reduction tools of the SID approach can help us identify structures that might not have been anticipated (and provide us with nuanced representations of them, which are designed to be orthogonal to dimensions of social stratification). Strategies of analysis can also be used to foreground socio-demographic differences in social connections within larger populations: by splitting populations into different social groups, we can explore interaction patterns within different groups, or evaluate how often interactions cut across groups. The examples presented in this chapter only scratch the surface of possible comparisons. It is not always feasible to construct data on social interaction patterns involving occupations within specific social groups, not least due to low volumes of cases, but there is certainly evidence to suggest that it is an endeavour worth exploring.

## Notes

1. The separation between 'subsidiary dimensions' and 'pseudo-diagonals' can be fuzzy. Subsidiary dimensional structures might sometimes be driven by just a few occupational combinations (in which case it is common practice to define 'pseudo-diagonals' for those combinations and then rerun the model at which point the subsidiary dimension should no longer emerge—see Sect. 6.7).
2. It is also possible for the dimension to run from those occupations that are the most gender segregated (i.e. male dominated or female dominated), to those that are the least segregated (i.e. most even proportions of male and female incumbents).
3. Perhaps counterintuitively, the orthogonal construction of the different dimensions does not ensure that, when dimension scores are assigned to other data, they are entirely uncorrelated.
4. An arrow pointing from B to A suggests that the combination of males from B living with females from A is unusually common; a double-headed arrow indicates that both couples where women in B are married to males in A, and couples where women in A are married to males in B, are over-represented.

## Bibliography

Alderson, A. S., Junisbai, A., & Heacock, I. (2007). Social Status and Cultural Consumption in the United States. *Poetics, 35*(2/3), 191–212.

Blackwell, L. (2003). Gender and Ethnicity at Work: Occupational Segregation and Disadvantage in the 1991 British Census. *Sociology, 37*(4), 713–731.

Blau, P. M., & Duncan, O. D. (1967). *The American Occupational Structure*. New York: Wiley.

Chan, T. W. (Ed.). (2010). *Social Status and Cultural Consumption*. Cambridge: Cambridge University Press.

Chan, T. W., & Goldthorpe, J. H. (2004). Is There a Status Order in Contemporary British Society? *European Sociological Review, 20*(5), 383–401.

Ganzeboom, H. B. G., & Treiman, D. J. (2003). Three Internationally Standardised Measures for Comparative Research on Occupational Status. In J. H. P. Hoffmeyer-Zlotnik & C. Wolf (Eds.), *Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables* (pp. 159–193). New York: Kluwer Academic Press.

Hakim, C. (1998). *Social Change and Innovation in the Labour Market: Evidence from the Census SARs on Occupational Segregation and Labour Mobility, Part-Time Work and Student Jobs, Homework and Self-Employment*. Oxford: Oxford University Press.

Hodge, R. W., Treiman, D. J., & Rossi, P. H. (1967). A Comparative Study of Occupational Prestige. In R. Bendix & S. M. Lipset (Eds.), *Class, Status and Power: Social Stratification in Comparative Perspective* (2nd ed., pp. 309–321). London: Routlege and Kegan Paul.

Hout, M., & DiPrete, T. A. (2006). What We Have Learned: RC28s Contributions to Knowledge About Social Stratification. *Research Into Social Stratification and Mobility, 24*, 1–20.

Khattab, N. (2016). The Ethno-Religious Wage Gap Within the British Salariat Class: How Severe Is the Penalty. *Sociology, 50*(4), 813–824.

Lambert, P. S., Tan, K. L. L., Gayle, V., Prandy, K., & Bergman, M. M. (2008). The Importance of Specificity in Occupation-Based Social Classifications. *International Journal of Sociology and Social Policy, 28*(5/6), 179–192.

Lambert, P. S., Zijdeman, R. L., Maas, I., van Leeuwen, M. H. D., & Prandy, K. (2013). The Construction of HISCAM: A Stratification Scale Based on Social Interactions for Historical Research. *Historical Methods, 46*(2), 77–89.

Laumann, E. O., & Guttman, L. (1966). The Relative Associational Contiguity of Occupations in an Urban Setting. *American Sociological Review, 31*, 169–178.

Levine, J. H., & Spadaro, J. (1988). Occupational Mobility: A Structural Model. In B. Wellman & S. D. Berkowitz (Eds.), *Social Structures: A Network Approach* (pp. 452–475). Cambridge: Cambridge University Press.

Martin, J., & Roberts, C. (1984). *Women and Employment: A Lifetime Perspective*. London: HMSO.

Playford, C., & Gayle, V. (2016). The Concealed Middle? An Exploration of Ordinary Young People and School GCSE Subject Area Attainment. *Journal of Youth Studies, 19*(2), 149–168.

Prandy, K. (1986). Similarities of Life-Style and the Occupations of Women. In R. Crompton & M. Mann (Eds.), *Gender and Stratification* (pp. 137–153). Cambridge: Polity Press.

Prandy, K., & Jones, F. L. (2001). An International Comparative Analysis of Marriage Patterns and Social Stratification. *International Journal of Sociology and Social Policy, 21*, 165–183.

Prandy, K., & Lambert, P. S. (2003). Marriage, Social Distance and the Social Space: An Alternative Derivation and Validation of the Cambridge Scale. *Sociology, 37*(3), 397–411.

Prandy, K., Lambert, P. S., & Bergman, M. M. (2002, July 7–13). *National Contexts and Cross-National Comparisons of Structures of Social Stratification*. Paper Presented at the ISA IVth World Congress of Sociology, Brisbane, RC20 Panel 16.

Treiman, D. J. (1977). *Occupational Prestige in Comparative Perspective*. New York: Academic Press.

University of Essex, & Institute for Social and Economic Research. (2010, July). *British Household Panel Survey: Waves 1–18, 1991–2009 [Computer File]* (7th ed.). Colchester, Essex: UK Data Archive [distributor], SN: 5151.

van der Lippe, T., & van Dijk, L. (2002). Comparative Research on Women's Employment. *Annual Review of Sociology, 28*, 221–241.

# 12

# Conclusions

## 12.1 Overview: Social Connections and Occupational Inequality

Our study has explored how data on the social connections between people, insofar as they are linked to occupations, can shed light upon patterns of consequential social inequalities. There are many different ways of measuring and understanding long-term social inequalities, but we argued that the best understanding of all comes from a characterisation of social positions through empirical data on the social interaction patterns exhibited between detailed occupational categories. Specifically, in Chaps. 4, 5, and 6 we described in depth the 'CAMSIS' approach to constructing 'Social Interaction Distance' scales for occupations. We demonstrated how CAMSIS scales—that should be interpreted as measures of the average position in the stratification structure held by the incumbents of an occupation—provide useful, arguably optimal, measures of social stratification. The construction of CAMSIS scales raise complex issues, and results can be contingent upon low-level decisions made during the scale construction process, but these challenges are surmountable and they should not put us off from exploring and exploiting


P. Lambert, D. Griffiths, *Social Inequalities and Occupational Stratification*,
https://doi.org/10.1057/978-1-137-02253-0_12

a very promising tool for understanding the social structure. The CAMSIS approach has been around for many decades, but the use of SID scales for occupations has recently enjoyed an upsurge in popularity, and this text incorporates some extended reference materials for people interested in the approach, and advice on the interpretation and exploitation of CAMSIS scales.

Many previous studies in the tradition of social network analysis have focussed upon work and employment. Some look at the ways in which relations between organisations define structural boundaries or impact upon economic outcomes (e.g. Knoke 2012; Berkowitz 1988), and others at the way that an individual's employment outcomes might be influenced by their social connections (e.g. Lin 1999; Granovetter 1973). Some studies have argued that channels of network relationships represent important structural forms (e.g. Fuhse 2009; Boorman and White 1976), to the extent that a network analysis might provide a rival model to social class in portraying structures of socio-economic inequality (e.g. Mercklé 2012). Both the CAMSIS approach and the supplementary network-based descriptive methods that we introduced in Chaps. 7 and 8 take forward these traditions by demonstrating that a summary of network patterns in social connections between occupations can itself provide valuable insights.

The relevance of social connections between occupations to understanding social inequality scales up across countries and over time. In our illustrative applications mentioned earlier, we have presented data from many different countries and time periods. There are a small number of consistent and plausible comparative differences—for instance, we reported greater skew in the stratification distribution for countries with a relatively large agricultural sector. In general however the evidence points to similar arrangements of social stratification across a wide range of societies. To us, this implies not only that the methodology of analysing social interactions between the incumbents of occupations is of universal relevance, but, moreover, that a theorisation of the importance of social interactions and occupations—such as our metaphor of the 'social resin'—has relevance across different societies.

One important aspect of the arrangements of stratification that our analyses emphasised was the gradational representation of inequality that emerges from the analysis of social interaction patterns in the CAMSIS tradition. It is possible that the same methods might have suggested significant discontinuities within dimensions of the occupational structure, but we are not aware of any examples of social interaction distance that have led to this result—indeed, users of SID techniques have taken this result as evidence for the gradational nature of the stratification structure more widely (e.g. Griffiths and Lambert 2012). Although a gradational approach is sometimes portrayed as a simplifying, or even a normative, way of studying social inequality, we would argue that the gradational representation associated with CAMSIS can contribute a nuanced, cautious, and appropriately parsimonious characterisation of social inequality. Compared against a popular alternative approach, of defining social classes and comparing between categories, those analyses of social inequalities that use appropriately contextualised gradational measures have the capacity to contribute more parsimonious, and ultimately more appropriately complex and accurate, empirical research evidence about social stratification.

Analytical tools for summarising social connections between occupations are complicated by the variety of different social mechanisms that influence social connections, and the variety of units of analysis between which social distances could be explored. Chapters 9, 10, and 11 elaborated on selected further issues related to the analysis of social connections between occupations—concerned with the characterisation of specific occupational circumstances in the context of other factors; the use of data on educational experiences in combination with data on occupations; and the value of characterising additional dimensions and social structures that might be linked to subsidiary patterns of social connections.

Two issues persisted throughout our exploration of empirical patterns in social connections between occupations, and we close our discussion on these points. One concerns the considerable operational challenges of using empirical data on occupations to characterise social distances and social inequalities. The second concerns the most helpful ways of theorising the strong empirical link between social connections and occupational inequalities.

## 12.2 Operational Issues

### 12.2.1 Advocacy of Occupations

There are many ways to measure social inequality and interpret its patterns and processes, but during the formative years of sociology, the occupation-based approach reigned supreme: it seemed obvious that measures of occupations, and social classifications derived from them, provided the most powerful and interesting indicators of inequality (e.g. Coxon and Jones 1978). In the twenty-first century, however, more social science publications that examine inequality look to other means of measurement, particularly focussing upon income and wealth (e.g. Dorling 2010), and measures that take account of cultural participation, social capital, and/or lifestyle indicators (e.g. Savage et al. 2013). Some non-occupational analyses have experienced considerable impact: Dorling (2010), for instance, has become a very popular and widely cited paean to the need to challenge social inequality; Savage et al. (2013) became the most downloaded academic article in the history of the journal *Sociology* when it was published and sparked numerous follow-on projects and reviews.

Non-occupational measures of social inequality can be seen as rivals to those measures that draw upon occupations—and rivals who may well have been winning the battle of late. Indeed, it is possible that occupation-based measures are seen by some as old fashioned, or perhaps as tied problematically to over-rehearsed, esoteric, cantankerous academic disputes (cf. Heath 1981). It is also possible that occupations are seen as 'hard' to use, because the numerous alternative occupation-based measures introduce confusion and uncertainty. Additionally, we suspect a disciplinary reason has diminished the popularity of occupational measures—the so-called qualitative turn in sociology dating from the 1970s widened the divide between quantitative and qualitative approaches, as qualitative sociologists increasingly turned their attention to 'non-traditional' topics of research and non-empirical tools for theorising and understanding social class and stratification. Arguably, later generations of sociologists were increasingly unaware of or uninspired by tools based upon occupations; at the same time, empirical researchers in other social science disciplines were increasingly drawn to the analysis of social inequality, but followed the conventions of their own disciplines which did not prioritise the use of data about occupations (cf. Goldthorpe 2010).

In spite of such trends, how could an approach that is inspired by the analysis of occupations seek to reassert itself as central to social science research? One route could be by empirical validation. In the preceding chapters, we hope that we have persuaded readers that data on occupations is readily available and is easy to transform into convenient occupation-based indicators. Moreover, it has powerful empirical properties—it is strongly associated with important correlates and outcomes of inequality, and it is particularly suited to comparative analysis across countries and time periods. Many of the barriers to using occupational data might also be overcome. Steps forward might be taken through routine distribution of easy-to-use occupation-based measures on empirical datasets, and by making efforts to engage between researchers who focus on occupations and those who focus on other inequalities. At the least, it could be unsatisfactory that many social scientists under-exploit detailed occupational data, but we believe that the need for change is even greater, because the story about inequality that may be told by different measures is likely to be qualitatively different: different conclusions are likely to be available, and given the centrality of occupational social relations to social inequalities, the best accounts are likely to be those that take fair account of occupations.

### 12.2.2 The Problems of Details

Whilst we take a strong position in favour of using detailed occupational data to explore inequalities, our own results have highlighted that this is not always an easy path to follow. Analysis that uses detailed occupational data is labour intensive and sometimes delivers a rather underwhelming premium (when compared with using more broad-brush occupation-based, or alternative, measures). Moreover, attention to detailed occupations raises further complications in results and interpretations—consider, for instance, the treatment of 'pseudo-diagonals' in the CAMSIS approach, which raises challenges of data organisation, statistical power, consistency of method, and coherency of interpretation.

It is easy to retreat from such complexities and work with simpler indicators and measures, but our evidence is that this is not satisfactory—this will neglect highly specific empirical patterns related to social inequalities and could lead ultimately to spurious interpretations and theorisations. On the positive side, the steady evolution in standards for documentation and replication in software-based analysis of social science datasets suggest a more promising future for those willing to confront detailed, but powerful, data!

## 12.3 Theorising the 'Social Resin'

> *Nothing stamps a man as much as his occupation. Daily work determines the mode of life … it constrains our ideas, feelings and tastes … People of the same occupation know one another, seek each other's company and frequent one another by necessity and choice.* (Goblot 1961, as cited by Coxon and Jones 1978, p. 10)
>
> *Whom you know has much to do with what you do. Most job-seekers activate their social connections to find jobs. Employers use ties linking the workers whom they know to the new people they may like to hire.* (Waldinger and Lichter 2003, p. 83)

We called the social ties that connect occupations a 'social resin', because this metaphor suggests a powerful force that shapes structure, whilst nevertheless exhibiting some malleability. Empirical patterns in social connections are consistent with a circular and mutually reinforcing relationship between social connections and occupational inequalities—that is, occupational inequalities foster social connection patterns, and social interactions foster occupational outcomes, as suggested by the two quotes above. But in what way, if at all, is the 'social resin'—the powerful empirical relationship between social interaction patterns, and social stratification systems—of importance in theorising social stratification and social reproduction?

> *two contradictory principles for legitimizing power were struggling for mastery—the principle of kinship and the principle of merit—and nearly everyone, in his heart of hearts, believed in both.* (Young 1958, p. 103)

One contribution might be in better reconciling a long-standing puzzle about social reproduction, such as characterised in Young's famous quote. The 'principle of kinship' for Young referred to individuals' tendencies to offer 'ascriptive' support on consequential resources to their family (and friends); the 'principle of merit' referred to the belief that allocations to consequential positions such as occupations should largely reflect an individual's 'ability' and 'effort'. 'Ascription' and 'meritocracy' coexist, but accounts of this are often relatively crude, for instance, the simplified portrayal of ascription as bad and meritocracy as good that dominates public and political discourses (cf. Payne 2017; Saunders 2010; Swift 2004; Young 1958). Our position is that the recognition of the 'social resin' helps bridge the gap between ascriptive and meritocratic models in providing a more effective mechanism for describing the persistence of social inequalities. On the one hand, it helps to empirically explain the mechanisms behind patterns of reproduction that are consistent with moderate levels of ascription: the social interactions defined by the social resin account for a large proportion of those patterns of outcomes that are consistent with 'ascription'. For instance, Ermisch et al. (2006) argue that around 40–50% of the intergenerational economic correlation is sustained through partnership homogamy. On the other hand, the 'social resin' also helps to explain the widespread social acceptance of a certain level of non-meritocratic allocation—people understand and find it acceptable that social interactions contribute to inequalities in outcomes. Few people object, for example, if a plumber appoints his nephew as his apprentice; many people aspire to economic positions of reproduction and stability, surrounded by their social contacts, even if on the face of it reproduction might do them few favours.

> *'… all around him was this ethic of the parents watching the test scores and, "What college is your son going to?"'. That is to say, their local neighbours and friends were … doctors, engineers, teachers, managers and so forth. Like the interviewees' parents, they were mostly educated people who also wanted their children to do well in school. At home and school, therefore, the interviewees socialised with friends who were like them.* (Devine 2004, p. 124)

Another contribution from thinking in terms of a 'social resin' is in theorising the driving forces behind its existence and influence. We suggested much earlier (Chap. 2) that there were at least two plausible accounts. The first puts the emphasis on social and cultural capital, as, for instance, the mechanisms of support described above by Devine. Sharing an orientation with a number of recent Bourdieusian accounts of social inequality (e.g. Atkinson 2017; Savage et al. 2015), this perspective foregrounds the differentiating influence of social and cultural capital and often suggests a model of sustained reproduction: we could say that social and cultural capital often help to lock individuals into structural positions. Detailed investigations of contemporary inequalities with this emphasis do not in fact assert an unbending social structure, and are typically most compelling when they suggest that fuller attention to social and cultural capital as mechanisms might help us to more effectively chart the depth of empirical inequalities (e.g. Laurison and Friedman 2016). Empirically, however, our analysis has also shown many social connections between individuals that seem to cross-cut barriers of social or cultural capital—for instance, in disproportionate social connections between disparate occupations that are linked by 'situs'. In the light of such patterns, it might be compelling to think of social and cultural capital as important considerations in characterising the social resin, but it seems less convincing to suggest their primacy.

> *The women were proud of some aspects of their community despite the problems they faced, and they spoke about the neighbourhood fondly, and with some gratitude … The neighbourhood, regardless of its problems, represented home, community, and also their place of safety.* (McKenzie 2015, p. 72)

McKenzie's account above refers to people living in quite deprived circumstances. From such examples we see more support for a second theoretical model, albeit one with a controversial ontology. This is the model of a deep-rooted preference for stability. Our empirical analysis has demonstrated that social connections patterns are certainly consistent with this model—people behave as if they sought out similarity and familiarity, and as if they have a preference for stability. Further investigation is desirable but it does seem to us that part of the reason for social reproduction could plausibly lie in psychological or psycho-social orientations towards stability.

The idea that individuals 'seek social reproduction' makes a compelling empirical model across the life course. From the earliest stages of childrearing, for instance, we see marked social differences in parenting and educational practices which apparently reflect intergenerationally imbued values (e.g. Ermisch et al. 2012). Through the years of schooling, habit and reproduction shape the experiences of children at the most fine-grained level—not least when privileged alumni send their children to the same private schools that they attended (e.g. Packard 1959, c16). We see that individuals disproportionately hold jobs which are exactly the same as their family and friends and that they are also more likely to hold exactly the same jobs as their parents (e.g. Bingley et al. 2012). A tendency to seek reproduction is not the only relevant force—individuals also express aspirations to 'better themselves' such as by improving their own socio-economic circumstances or by supporting their children in 'doing better for themselves'. Such aspirations are common, yet in an empirical sense social outcomes are more consistent with the expectation that most people mostly choose stability in those areas of life where they have agency.

Many sociological accounts have emphasised problematic aspects of preferences for, or rationalisations of, stability—one example, highlighted previously, is in Bourdieu's (1984) negative portrayal of how the disadvantaged 'refuse what they are refused'. However, we could also consider representing social reproduction more positively, as something that is likely to satisfy stability preferences. Our analysis of social connection patterns does seem to suggest that many individuals behave through their social interactions in ways which are consistent with them actively seeking occupational social reproduction. Perhaps 'seeking social reproduction' should be treated as a more benign social preference, one that should not necessarily be confronted?

It is important to recognise that a more nuanced attitude to social reproduction need not imply resistance to other social change. Hypothetically, in one domain of social life, we might see stability of circumstances through social reproduction (such as occupational inheritance from parent to child), but this could co-exist with other forms of social change in other domains and circumstances. In particular, levels of internal economic inequalities within a society could in principle be reduced without any reduction in social reproduction in how individuals connect with occupations, such as through social policies which reduce occupational income disparities. As such, an evaluation that positions some aspects of social reproduction in a positive light—that is, as satisfying a preference for stability—should not be equated with a conservative or apologist perspective towards other social inequalities.

Nevertheless, difficult normative questions are raised by entertaining the idea that satisfying a preference for stability might be valued as a social outcome. If individuals are taking paths, which they seem to prefer, but which may have the consequence of disadvantaging them, should social policies seek to intervene and change this? For example, both gender and ethnic segregation in occupations are associated with the perpetuation of gender and ethnic inequalities (e.g. Brynin and Guveli 2012; Charles and Grusky 2004), yet both are often shown to arise due to aspirations, preferences, and social patterns in recruitment networks (e.g. Waldinger and Lichter 2003; Hakim 2000)—there is potentially a trade-off between the reduction of other social inequalities and denial of the preference for stability.

Further research might usefully assess the extent to which social ills arise when social reproduction is *not* supported. There are numerous demonstrations that parents and children alike become distressed when they are unable to sustain intergenerational occupational reproduction, even when it would be of a modest or arguably disadvantaged form, or when mobility has been, economically, for the better (e.g. Friedman 2014). The dissonance of expectation mismatch in occupations is a popular cultural genre—the UK film *Brassed off*, for instance, portrays the challenges faced by miners during industrial transformations which force them into different employment sectors. From such perspectives, supporting social reproduction as a strategy to reduce dissonance and discomfort seems plausible. This objective is ambiguous however because many social problems evident in circumstances of deprivation are probably themselves sustained by social adaptations to deprivation: perhaps new generations come to expect the sustained disadvantage and disengagement as were experienced by their parents and seek to reproduce rather than avoid them (cf. Jones 2011). Perhaps such reasons explain why relatively few previous sociological accounts have considered a model of preferences for social reproduction, or reflected upon positive rather than negative aspects of reproduction. On the contrary, many influential texts have focussed upon social change and strategies for promoting change and reducing social reproduction. Our analysis, by looking at the finer details of social reproduction and occupations, suggests that it is valuable to consider nuanced elements of social reproduction that are likely to include positive as well as negative features.

## Bibliography


Atkinson, W. (2017). *Class in the New Millennium: The Structure, Homologies and Experience of the British Social Space*. London: Routledge.

Berkowitz, S. D. (1988). Markets and Market-Areas: Some Preliminary Formulations. In B. Wellman & S. D. Berkowitz (Eds.), *Social Structures: A Network Approach* (pp. 261–303). Cambridge: Cambridge University Press.

Bingley, P., Corak, M., & Westergard-Nielsen, N. (2012). Equality of Opportunity and Intergenerational Transmission of Employers. In J. Ermisch, M. Jantti, & T. Smeeding (Eds.), *From Parents to Children: The Intergenerational Transmission of Advantage* (pp. 441–460). New York: Russell Sage Foundation.

Boorman, S. A., & White, H. C. (1976). Social Structure from Multiple Networks: Part II. Role Interlock. *American Journal of Sociology, 81*(6), 1384–1446.

Bourdieu, P. (1984). *Distinction*. London: Routledge and Kegan Paul.

Brynin, M., & Guveli, A. (2012). Understanding the Ethnic Pay Gap in Britain. *Work, Employment and Society, 26*(4), 574–587.

Charles, M., & Grusky, D. B. (2004). *Occupational Ghettos: The Worldwide Segregation of Women and Men*. Stanford, CA: Stanford University Press.

Coxon, A. P. M., & Jones, C. L. (1978). *The Images of Occupational Prestige: A Study in Social Cognition*. London: Macmillan Press.

Devine, F. (2004). *Class Practices: How Parents Help Their Children Get Good Jobs*. Cambridge: Cambridge University Press.

Dorling, D. (2010). *Injustice: Why Social Inequality Persists*. Bristol: The Policy Press.

Ermisch, J., Francesconi, M., & Siedler, T. (2006). Intergenerational Mobility and Marital Sorting. *The Economic Journal, 116*, 659–679.

Ermisch, J., Jantti, M., & Smeeding, T. (Eds.). (2012). *From Parents to Children: The Intergenerational Transmission of Advantage*. New York: Russell Sage Foundation.

Friedman, S. (2014). The Price of the Ticket: Rethinking the Experience of Social Mobility. *Sociology, 48*(2), 352–368.
Fuhse, J. (2009). The Meaning Structure of Social Networks. *Sociological Theory, 27*(1), 51–73.
Goblot, E. (1961). Class and Occupation. In T. Parsons (Ed.), *Theories of Society*. New York: Free Press.
Goldthorpe, J. H. (2010). Analysing Social Inequality: A Critique of Two Recent Contributions from Economics and Epidemiology. *European Sociological Review, 26*(6), 731–744.
Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology, 78*(6), 1360–1380.
Griffiths, D., & Lambert, P. S. (2012). Dimensions and Boundaries: Comparative Analysis of Occupational Structures Using Social Network and Social Interaction Distance Analysis. *Sociological Research Online, 17*(2), 5.
Hakim, C. (2000). *Work-Lifestyle Choices in the 21st Century—Preference Theory*. Oxford: Oxford University Press.
Heath, A. F. (1981). Book Review: Social Stratification and Occupations. *Sociology, 15*(4), 620–623.
Jones, O. (2011). *Chavs—The Demonisation of the Working Class*. London: Verso.
Knoke, D. (2012). *Economic Networks*. Cambridge: Polity Press.
Laurison, D., & Friedman, S. (2016). The Class Pay Gap in Higher Professional and Managerial Occupations. *American Sociological Reivew, 81*(4), 668–695.
Lin, N. (1999). Social Networks and Status Attainment. *Annual Review of Sociology, 25*, 467–488.
McKenzie, L. (2015). *Getting By: Estates, Class and Culture in Austerity Britain*. Bristol: Policy Press.
Mercklé, P. (2012). Who Is Responsible for the Disappearance of Social Classes? *Bulletin de Methodologie Socilogique, 116*(1), 67–75.
Packard, V. (1959). *The Status Seekers*. London: Penguin.
Payne, G. (2017). *The New Social Mobility: How the Politicians Got it Wrong*. Bristol: Policy Press.
Saunders, P. (2010). *Social Mobility Myths*. London: CIVITAS: Institute for the Study of Civil Society.
Savage, M., Cunningham, N., Devine, F., Friedman, S., Laurison, D., McKenzie, L., Miles, A., Snee, H., & Wakeling, P. (2015). *Social Class in the 21st Century*. London: Pelican Books.

Savage, M., Devine, F., Cunningham, N., Taylor, M., Li, Y., Hjellbrekke, J., Le Roux, B., Friedman, S., & Miles, A. (2013). A New Model of Social Class: Findings from the BBC's Great British Class Survey Experiment. *Sociology, 47*(2), 219–250.
Swift, A. (2004). Would Perfect Mobility be Perfect? *European Sociological Review, 20*(1), 1–11.
Waldinger, R., & Lichter, M. I. (2003). *How the Other Half Works: Immigration and the Social Organization of Labor*. Berkeley: University of California Press.
Young, M. (1958). *The Rise of the Meritocracy 1870–2033*. Harmondsworth: Penguin.

# Bibliography


Abbott, A. (2006). Mobility: What? When? How? In S. L. Morgan, D. B. Grusky, & G. S. Fields (Eds.), *Mobility and Inequality* (pp. 137–161). Stanford, CA: Stanford University Press.

Abel, T. (2008). Cultural Capital and Social Inequality in Health. *Journal of Epidemiology and Community Health, 62*(7), e13.

ADLS. (2012). *P-ADLS: Portal of the Administrative Data Liaison Service*. Retrieved May 1, 2012, from http://www.adls.ac.uk/padls/.

Alderson, A. S., Junisbai, A., & Heacock, I. (2007). Social Status and Cultural Consumption in the United States. *Poetics, 35*(2/3), 191–212.

Allison, P. D. (2009). *Fixed Effects Regression Models*. Thousand Oaks, CA: Sage.

Archer, M. S. (2007). *Making Our Way Through the World: Human Reflexivity and Social Mobility*. Cambridge: Cambridge University Press.

Arum, R., Gamoran, A., & Shavit, Y. (2007). More Inclusion Than Diversion: Expansion, Differentiation and Market Structure in Higher Education. In Y. Shavit, R. Arum, & A. Gamoran (Eds.), *Stratification in Higher Education: A Comparative Study* (pp. 1–35). Stanford, CA: Stanford University Press.

Atkinson, A. B. (2005). Social Indicators, Policy, and Measuring Progress. In S. Svallfors (Ed.), *Analyzing Inequality* (pp. 83–107). Stanford, CA: Stanford University Press.

Atkinson, A. B., & Brandolini, A. (2006). The Panel-of-Countries Approach to Explaining Income Inequality: An Interdisciplinary Research Agenda. In




P. Lambert, D. Griffiths, *Social Inequalities and Occupational Stratification*,
https://doi.org/10.1057/978-1-137-02253-0

S. L. Morgan, D. B. Grusky, & G. S. Fields (Eds.), *Mobility and Inequality* (pp. 400–448). Stanford, CA: Stanford University Press.

Atkinson, W. (2017). *Class in the New Millennium: The Structure, Homologies and Experience of the British Social Space*. London: Routledge.

Attewell, P., & Lavin, D. E. (2007). *Passing the Torch: Does Higher Education for the Disadvantaged Pay Off Across the Generations?* New York: Russell Sage Foundation.

Bakker, B. F. M. (1993). A New Measure of Social Status for Men and Women: The Social Distance Scale. *Netherlands Journal of Social Sciences, 29*, 113–129.

Bechhofer, F. (1969). Occupations. In M. Stacey (Ed.), *Comparability in Social Research* (pp. 94–122). London: Heinemann.

Beck, U. (2000). *The Brave New World of Work*. Cambridge: Polity Press.

Bellotti, E., & Mora, E. (2016). Networks of Practices in Critical Consumption. *Journal of Consumer Culture, 16*(3), 718–760.

Bennett, T., Savage, M., Silva, E. B., Warde, A., Gayo-Cal, M., & Wright, D. (2009). *Culture, Class, Distinction*. London: Routledge.

Bergman, M. M., Lambert, P. S., Prandy, K., & Joye, D. (2002). Theorisation, Construction and Validation of a Social Stratification Scale: Cambridge Social Interaction and Stratification Scale (CAMSIS) for Switzerland. *Swiss Journal of Sociology, 28*(1), 441–460.

Berkowitz, S. D. (1988). Markets and Market-Areas: Some Preliminary Formulations. In B. Wellman & S. D. Berkowitz (Eds.), *Social Structures: A Network Approach* (pp. 261–303). Cambridge: Cambridge University Press.

Bernardi, F., & Ballarino, G. (Eds.). (2016). *Education, Occupation and Social Origin: A Comparative Analysis of the Transmission of Socio-Economic Inequalities*. Cheltenham: Elgar.

Bessudnov, A. (2012). A Relational Occupational Scale for Russia. In P. S. Lambert, R. Connelly, R. M. Blackburn, & V. Gayle (Eds.), *Social Stratification: Trends and Processes* (pp. 53–65). Aldershot: Ashgate.

Bihagen, E., & Lambert, P. S. (2012, September, 24–26). *Class and Status: One or Two Forms of Stratification?* Paper Presented at the Equalsoc Conference, Stockholm.

Bingley, P., Corak, M., & Westergard-Nielsen, N. (2012). Equality of Opportunity and Intergenerational Transmission of Employers. In J. Ermisch, M. Jantti, & T. Smeeding (Eds.), *From Parents to Children: The Intergenerational Transmission of Advantage* (pp. 441–460). New York: Russell Sage Foundation.

Biressi, A., & Nunn, H. (2013). *Class and Contemporary British Culture*. Basingstoke: Palgrave Macmillan.

Blackburn, R. M. (2008). What Is Social Inequality? *International Journal of Sociology and Social Policy, 28*(7/8), 250–259.

Blackwell, L. (2003). Gender and Ethnicity at Work: Occupational Segregation and Disadvantage in the 1991 British Census. *Sociology, 37*(4), 713–731.

Blau, P. M., & Duncan, O. D. (1967). *The American Occupational Structure*. New York: Wiley.

Blossfeld, H. P., Mills, M., & Bernardi, F. (Eds.). (2006). *Globalization, Uncertainty and Men's Careers: An International Comparison*. Cheltenham: Edward Elgar.

Blossfeld, H. P., & Timm, A. (Eds.). (2003). *Who Marries Whom: Educational Systems as Marriage Markets in Modern Societies*. London: Kluwer Academic.

Blum, K., Oscar-Berman, M., Giordano, J., Downs, B. W., Simpatico, T., Han, D., & Femino, J. (2012). Neurogenetic Impairments of Brain Reward Circuitry Links to Reward Deficiency Syndrome (RDS): Potential Nutrigenomic Induced Dopaminergic Activation. *Journal of Genetic Syndromes and Gene Therapy, 3*(4), 1000e1115.

Bol, T., & Weeden, K. A. (2015). Occupational Closure and Wage Inequality in Germany and the United Kingdom. *European Sociological Review, 31*(3), 354–369.

Bolton, P. (2012). *Education: Historical Statistics*. London: House of Commons Library, Standard Note SN04252. Retrieved January 1, 2016, from http://researchbriefings.files.parliament.uk/documents/SN04252/SN04252.pdf.

Boorman, S. A., & White, H. C. (1976). Social Structure from Multiple Networks: Part II. Role Interlock. *American Journal of Sociology, 81*(6), 1384–1446.

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analysing Social Networks*. London: Sage.

Bott, E. (1957). *Family and Social Network*. London: Tavistock Publications.

Bottero, W. (2000). Gender and the Labour Market at the Turn of the Century: Complexity, Ambiguity and Change. *Work Employment and Society, 14*(4), 781–791.

Bottero, W. (2005a). *Stratification: Social Division and Inequality*. London: Routledge.

Bottero, W. (2005b). Interaction Distance and the Social Meaning of Occupations. *Sociological Review, 53*(2), 56–72.

Bottero, W. (2009). Relationality and Social Interaction. *British Journal of Sociology, 60*(2), 399–420.

Bottero, W., Lambert, P. S., Prandy, K., & McTaggart, S. (2009). Occupational Structures: The Stratification Space of Social Interaction. In K. Robson & C. Sanders (Eds.), *Quantifying Theory: Pierre Bourdieu* (pp. 141–150). Amsterdam: Springer.

Bottero, W., & Prandy, K. (2003). Social Interaction Distance and Stratification. *British Journal of Sociology, 54*(2), 177–197.

Bourdieu, P. (1984). *Distinction*. London: Routledge and Kegan Paul.

Bradley, H. (1989). *Men's Work, Women's Work: A Sociological History of the Sexual Division of Labour in Employment*. Cambridge: Polity Press.

Braverman, H. (1974). *Labour and Monopoly Capital*. London: Monthly Review Press.

Breen, R. (Ed.). (2004). *Social Mobility in Europe*. Oxford: Oxford University Press.

Breen, R., & Goldthorpe, J. H. (1997). Explaining Educational Differentials: Towards a Formal Rational Action Theory. *Rationality and Society, 9*(3), 275–305.

Breiger, R. L. (1981). The Social Class Structure of Occupational Mobility. *American Journal of Sociology, 87*(3), 578–611.

Brown, P., & Scase, R. (1995). *Higher Education and Corporate Realities: Class, Culture and the Decline of Graduate Careers*. London: UCL Press.

Brynin, M., & Ermisch, J. (Eds.). (2008). *Changing Relationships*. London: Routledge.

Brynin, M., & Guveli, A. (2012). Understanding the Ethnic Pay Gap in Britain. *Work, Employment and Society, 26*(4), 574–587.

Brynin, M., Martinez Perez, A., & Longhi, S. (2008). How Close are Couples? In M. Brynin & J. Ermisch (Eds.), *Changing Relationships* (pp. 93–108). London: Routledge.

Buis, M. L. (2010). *Inequality of Educational Outcome and Inequality of Educational Opportunity in the Netherlands During the 20th Century*. Amsterdam: VU University Amsterdam.

Bukodi, E., & Goldthorpe, J. H. (2013). Decomposing 'Social Origins': The Effects of Parents' Class, Status and Education on the Educational Attainment of Their Children. *European Sociological Review, 29*(5), 1024–1039.

Bulmer, M., Gibbs, J., & Hyman, L. (Eds.). (2010). *Social Measurement Through Social Surveys: An Applied Approach*. Aldershot: Ashgate.

Burchell, B. (1994). The Effects of Labour Market Position, Job Insecurity and Unemployment on Psychological Health. In D. Gallie, C. Marsh, & C. Vogler (Eds.), *Social Change and the Experience of Unemployment* (pp. 188–212). Oxford: Oxford University Press.

Burrows, R., & Crow, G. (2006). Geodemographics, Software and Class. *Sociology, 40*(5), 793–812.

Butler, P. (2010). *Visualizing Friendships*. Retrieved December 1, 2012, from http://www.facebook.com/notes/facebook-engineering/visualizing-friendships/469716398919.

Centers, R. (1949). Marital Selection and Occupational Strata. *American Journal of Sociology, 54*(6), 530–535.

Chan, T. W. (Ed.). (2010a). *Social Status and Cultural Consumption*. Cambridge: Cambridge University Press.

Chan, T. W. (2010b). The Social Status Scale: Its Construction and Properties. In T. W. Chan (Ed.), *Social Status and Cultural Consumption* (pp. 28–56). Cambridge: Cambridge University Press.

Chan, T. W., Birkelund, G. E., Aas, A. K., & Wiborg, O. (2011). Social Status in Norway. *European Sociological Review, 27*(4), 451–468.

Chan, T. W., & Goldthorpe, J. H. (2004). Is There a Status Order in Contemporary British Society? *European Sociological Review, 20*(5), 383–401.

Chan, T. W., & Goldthorpe, J. H. (2007). Class and Status: The Conceptual Distinction and Its Empirical Relevance. *American Sociological Review, 72*(4), 512–532.

Charles, M., & Grusky, D. B. (2004). *Occupational Ghettos: The Worldwide Segregation of Women and Men*. Stanford, CA: Stanford University Press.

Christakis, N., & Fowler, J. (2010). *Connected: The Amazing Power of Social Networks and How They Shape Our Lives*. London: Harper Press.

Clark, G. (2014). *The Son Also Rises: Surnames and the History of Social Mobility*. Princeton, NJ: Princeton University Press.

Clogg, C. C. (1982). Using Association Models in Sociological Research: Some Examples. *American Journal of Sociology, 88*(1), 114–134.

Clogg, C. C., & Shihadeh, E. S. (1994). *Statistical Models for Ordinal Variables*. Thousand Oaks, CA: Sage.

Connelly, R., Gayle, V., & Lambert, P. S. (2016a). Modelling Key Variables in Social Science Research: Introduction to the Special Section. *Methodological Innovations Online*. https://doi.org/10.1177/2059799116637782.

Connelly, R., Gayle, V., & Lambert, P. S. (2016b). A Review of Educational Attainment Measures for Social Survey Research. *Methodological Innovations Online*. https://doi.org/10.1177/2059799116638001.

Corak, M. (Ed.). (2004). *Generational Income Mobility in North America and Europe*. Cambridge: Cambridge University Press.

Coxon, A. P. M., & Jones, C. L. (1978). *The Images of Occupational Prestige: A Study in Social Cognition*. London: Macmillan Press.

Crompton, R. (1998). *Class and Stratification: An Introduction to Current Debates* (2nd ed.). Cambridge: Polity Press.

Crompton, R., & Mann, M. (Eds.). (1986). *Gender and Stratification*. Cambridge: Polity Press.

Crompton, R., & Sanderson, K. (1990). *Gendered Jobs and Social Change*. London: Unwin Hyman.

Croon, M. A. (2002). Ordering the Classes. In J. A. Hagenaars & A. McCutcheon (Eds.), *Applied Latent Class Analysis* (pp. 137–162). Cambridge: Cambridge University Press.

Crossley, N. (2015). *Networks of Sound, Style and Subversion: The Punk and Post-Punk Worlds of Manchester, London, Liverpool and Sheffield, 1975–1980*. Manchester: Manchester University Press.

Curington, C. V., Lin, K.-H., & Lunquist, J. H. (2015). Treatment of Multiracial Daters in an Online Dating Website. *American Journal of Sociology, 80*(4), 764–788.

Dale, A. (2006). Quality Issues with Survey Research. *International Journal of Social Research Methodology, 9*(2), 143–158.

Davies, R., & Elias, P. (2010). The Application of ESeC to Three Sources of Comparative European Data. In D. Rose & E. Harrison (Eds.), *Social Class in Europe: An Introduction to the European Socio-Economic Classification* (pp. 61–86). London: Routledge.

de Luca, D., Meraviglia, C., & Ganzeboom, H. B. G. (2010, April 14). *ICAM: The International CAMSIS Scale*. Hafia: First Presented at ISA RC28, Hafia. Software Retrieved from http://www.harryganzeboom.nl.

de Luca, D., Meraviglia, C., & Ganzeboom, H. B. G. (2012). Measures and Dimensions of Occupational Stratification: The Case of a Relational Scale for Italy. In P. S. Lambert, R. Connelly, R. M. Blackburn, & V. Gayle (Eds.), *Social Stratification: Trends and Processes* (pp. 29–51). Aldershot: Ashgate.

de Nooy, W., Mrvar, A., & Batagelj, V. (2011). *Exploratory Social Network Analysis with Pajek*. Cambridge: Cambridge University Press.

Devine, F. (2004). *Class Practices: How Parents Help Their Children Get Good Jobs*. Cambridge: Cambridge University Press.

Dolton, P. J., & Makepeace, G. H. (1993). Female Labour Force Participation and the Choice of Occupation: The Supply of Teachers. *European Economic Review, 37*(7), 1393–1411.

Doogan, K. (2009). *New Capitalism? The Transformation of Work*. Cambridge: Polity.

Dorling, D. (2010). *Injustice: Why Social Inequality Persists*. Bristol: The Policy Press.

Dorling, D. (2014). Thinking About Class. *Sociology, 48*(3), 452–462.

Duncan-Jones, P. (1972). Social Mobility, Canonical Scoring and Occupational Classification. In K. Hope (Ed.), *The Analysis of Social Mobility* (pp. 191–210). Oxford: Oxford University Press.

Elias, P. (2000). Status in Employment: A World Survey of Practices and Problems. *Bulletin of Labour Statistics, 1*, 11–19.

Elias, P., & Birch, M. (1994). *Establishment of Community-Wide Occupational Statistics: ISCO-88(COM), A Guide for Users*. Warwick: Institute for Employment Research, University of Warwick.

Elias, P., & McKnight, A. (2001). Skill Measurement in Official Statistics: Recent Developments in the UK and the Rest of Europe. *Oxford Economic Papers, New Series, 53*(3), 508–540.

Erikson, R., & Goldthorpe, J. H. (1992). *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford: Clarendon Press.

Erikson, R., Goldthorpe, J. H., & Portocarero, L. (1979). Intergenerational Class Mobility in Three Western European Societies: England, France and Sweden. *British Journal of Sociology, 30*, 415–440.

Ermisch, J., Francesconi, M., & Siedler, T. (2006). Intergenerational Mobility and Marital Sorting. *The Economic Journal, 116*, 659–679.

Ermisch, J., Jantti, M., & Smeeding, T. (Eds.). (2012). *From Parents to Children: The Intergenerational Transmission of Advantage*. New York: Russell Sage Foundation.

Evans, G., & Mills, C. (1998). Identifying Class Structure: A Latent Class Analysis of the Criterion-Related and Construct Validity of the Goldthorpe Class Schema. *European Sociological Review, 14*(1), 87–106.

Evans, G., & Mills, C. (2000). In Search of the Wage-Labour/Service Contract: New Evidence on the Validity of the Goldthorpe Class Schema. *British Journal of Sociology, 51*(4), 641–661.

Feld, S. (1981). The Focused Organization of Social Ties. *American Journal of Sociology, 86*(5), 1015–1035.

Field, J. (2015). Social Ties, Agency and Change: Education and Social Capital in Adult Life. In Y. Li (Ed.), *Handbook of Research Methods and Applications in Social Capital* (pp. 292–306). Cheltenham: Edward Elgar.

Flap, H., & Boxman, E. (1999). Getting a Job as a Manager. In R. T. A. J. Leenders & S. M. Gabbay (Eds.), *Corporate Social Capital and Liability* (pp. 197–216). Boston: Springer US.

Foner, N. (2000). *From Ellis Island to JFK: New York's Two Great Waves of Immigration*. New Haven: Yale University Press.

Friedman, S. (2014). The Price of the Ticket: Rethinking the Experience of Social Mobility. *Sociology, 48*(2), 352–368.

Fuhse, J. (2009). The Meaning Structure of Social Networks. *Sociological Theory, 27*(1), 51–73.

Ganesh, J. (2012). *George Osborne: The Austerity Chancellor*. Hull: Biteback Publishing.

Ganzeboom, H. B. G. (2005). On the Cost of Being Crude: A Comparison of Detailed and Coarse Occupational Coding. In J. H. P. Hoffmeyer-Zlotnik & J. Harkness (Eds.), *Methodological Aspects in Cross-National Research* (pp. 241–257). Mannheim: ZUMA, Nachrichten Spezial.

Ganzeboom, H. B. G. (2010). *Do's and Don'ts of Occupation Coding with an Extension to ISCO-08*. Working Paper of the Department of Social Research Methodology. Amsterdam: Free University of Amsterdam.

Ganzeboom, H. B. G. (2016). *Tools for Deriving Occupational Status Measures from ISCO-88*. Retrieved March 1, 2016, from http://www.harryganzeboom.nl/isco88/index.htm.

Ganzeboom, H. B. G., de Graaf, P. M., & Treiman, D. J. (1992). A Standard International Socio-Economic Index of Occupational Status. *Social Science Research, 21*, 1–56.

Ganzeboom, H. B. G., & Treiman, D. J. (2003). Three Internationally Standardised Measures for Comparative Research on Occupational Status. In J. H. P. Hoffmeyer-Zlotnik & C. Wolf (Eds.), *Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables* (pp. 159–193). New York: Kluwer Academic Press.

Gascoigne, P. (2005). *Gazza: My Story*. London: Headline Press.

Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Gilbert, G. N. (1993). *Analyzing Tabular Data: Loglinear and Logistic Models for Social Researchers*. London: UCL Press.

Giuffre, K. (2013). *Communities and Networks: Using Social Network Analysis to Rethink Urban and Community Studies*. Cambridge: Polity Press.

Glass, D. V. (1954). *Social Mobility in Britain*. London: Routledge & Kegan Paul.

Goblot, E. (1961). Class and Occupation. In T. Parsons (Ed.), *Theories of Society*. New York: Free Press.

Goldstein, H. (2003). *Multilevel Statistical Models* (3rd ed.). London: Arnold.

Goldthorpe, J. H. (1997). The “Goldthorpe” Class Schema: Some Observations on Conceptual and Operational Issues in Relation to the ESRC Review of Government Social Classifications. In D. Rose & K. O’Reilly (Eds.), *Constructing Classes* (pp. 40–48). Colchester: University of Essex.

Goldthorpe, J. H. (2007). *On Sociology, Second Edition: Volume One—Critique and Programme*. Stanford, CA: Stanford University Press.

Goldthorpe, J. H. (2010). Analysing Social Inequality: A Critique of Two Recent Contributions from Economics and Epidemiology. *European Sociological Review, 26*(6), 731–744.

Goodman, L. A. (1979). Simple Models for the Analysis of Association in Cross-Classifications Having Ordered Categories. *Journal of the American Statistical Association, 74*, 537–552.

Goodman, L. A. (2002). Latent Class Analysis: The Empirical Study of Latent Types, Latent Variables, and Latent Structures. In J. A. Hagenaars & A. McCutcheon (Eds.), *Applied Latent Class Analysis* (pp. 3–55). Cambridge: Cambridge University Press.

Gordon, D. (2006). The Concept and Measurement of Poverty. In C. Pantazis, D. Gordon, & R. Levitas (Eds.), *Poverty and Social Exclusion in Britain: The Millenium Survey* (pp. 29–69). Bristol: The Policy Press.

Gornick, J., & Jantti, M. (Eds.). (2013). *Income Inequality: Economic Disparities and the Middle Class in Affluent Countries*. Stanford, CA: Stanford University Press.

Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology, 78*(6), 1360–1380.

Greenacre, M., & Blasius, J. (1994). *Correspondence Analysis in the Social Sciences: Recent Developments and Applications*. London: Academic Press.

Griffiths, D., & Lambert, P. S. (2012). Dimensions and Boundaries: Comparative Analysis of Occupational Structures Using Social Network and Social Interaction Distance Analysis. *Sociological Research Online, 17*(2), 5.

Griffiths, D., & Lambert, P. S. (2013). The American Occupational Marriage Structure, 1980 and 2010. *Procedia—Social and Behavioural Sciences, 100*, 21–39.

Griffiths, D., & Lambert, P. S. (2015). Social Capital and the Social Relations of Occupational Structure. In Y. Li (Ed.), *Handbook of Research Methods and Applications in Social Capital* (pp. 204–224). London: Sage.

Grund, T. U. (2014). *nwcommands: Software Tools for the Statistical Modeling of Network Data in Stata*. Retrieved March 1, 2017, from http://nwcommands.org.

Grusky, D. B., & Weeden, K. A. (2006). Does the Sociological Approach to Studying Social Mobility Have a Future? In S. L. Morgan, D. B. Grusky, & G. S. Fields (Eds.), *Mobility and Inequality* (pp. 85–108). Stanford, CA: Stanford University Press.

Guinea-Martin, D., Blackwell, L., & Elliott, J. (2010). Occupational Segregation and Concentration: An Analysis by Sex, Employment Status and Ethnic Group in England and Wales. In J. Stillwell, P. Norman, C. Thomas, & P. Surridge (Eds.), *Spatial and Social Disparities: Understanding Population Trends and Processes* (Vol. 2, pp. 99–111). Dordrecht: Springer.

Guttman, L. (1944). A Basis for Scaling Quantitative Data. *American Sociological Reivew, 9*(2), 139–150.

Guveli, A. (2006). *New Social Classes Within the Service Class in the Netherlands and Britain: Adjusting the EGP Class Schema for the Technocrats and the Social and Cultural Specialists*. Nijmegen: Radboud University Nijmegen.

Hagenaars, J. A. (2008). *COMPSOC: Exploiting, Documenting and Enriching Comparative Data from Large-Scale Surveys in the Social Sciences*. Tilburg University, Investeringen NWO-Middelgroot.

Hakim, C. (1998). *Social Change and Innovation in the Labour Market: Evidence from the Census SARs on Occupational Segregation and Labour Mobility, Part-Time Work and Student Jobs, Homework and Self-Employment*. Oxford: Oxford University Press.

Hakim, C. (2000). *Work-Lifestyle Choices in the 21st Century—Preference Theory*. Oxford: Oxford University Press.

Hallsten, M. (2013). The Class-Origin Wage Gap: Heterogeneity in Education and Variations Across Market Segments. *British Journal of Sociology, 64*(4), 662–690.

Halsey, A. H. (1988). Higher Education. In A. H. Halsey (Ed.), *British Social Trends Since 1900* (pp. 268–296). London: Macmillan.

Hanneman, R. A., & Riddle, M. (2011). Concepts and Measures for Basic Network Analysis. In J. Scott & P. J. Carrington (Eds.), *The Sage Handbook of Social Network Analysis* (pp. 340–369). London: Sage.

Harkness, S. (2013). Women's Employment and Household Income Inequality. In J. C. Gornick & M. Jantti (Eds.), *Income Inequality: Economic Disparities and the Middle Class in Affluent Countries* (pp. 207–233). Stanford, CA: Stanford University Press.

Hauser, R. M., & Warren, J. R. (1997). Socioeconomic Indexes for Occupations: A Review, Update, and Critique. *Sociological Methodology, 27*, 177–298.

Haythornthwaite, C. (2002). Strong, Weak, and Latent Ties and the Impact of New Media. *The Information Society, 18*(5), 385–401.

Heath, A. F. (1981). Book Review: Social Stratification and Occupations. *Sociology, 15*(4), 620–623.

Hennig, C., & Liao, T. F. (2013). How to Find an Appropriate Clustering for Mixed Type Variables with Application to Socio-Economic Stratification. *Journal of the Royal Statistical Society Series C, 62*(3), 309–369.

Hinde, R. A. (1997). *Relationships: A Dialectical Perspective*. Hove: Psychology Press.

Hodge, R. W., Treiman, D. J., & Rossi, P. H. (1967). A Comparative Study of Occupational Prestige. In R. Bendix & S. M. Lipset (Eds.), *Class, Status and Power: Social Stratification in Comparative Perspective* (2nd ed., pp. 309–321). London: Routlege and Kegan Paul.

Hoffmeyer-Zlotnik, J. H. P., & Warner, U. (2014). *Harmonising Demographic and Socio-Economic Variables for Cross-National Comparative Survey Research*. Berlin: Springer.

Hout, M., & DiPrete, T. A. (2006). What We Have Learned: RC28s Contributions to Knowledge About Social Stratification. *Research Into Social Stratification and Mobility, 24*, 1–20.

Hox, J. (2010). *Multilevel Analysis* (2nd ed.). London: Routledge.

Hyman, H. H. (1967). The Value Systems of Different Classes: A Social Psychological Contribution to the Analysis of Stratification. In R. Bendix & S. M. Lipset (Eds.), *Class, Status and Power: Social Stratification in Comparative Perspective* (2nd ed., pp. 488–499). London: Routlege and Kegan Paul.

ILO. (1990). *ISCO-88: International Standard Classification of Occupations*. New York: International Labour Office.

Jarman, J., Blackburn, R. M., & Racko, G. (2012). The Dimensions of Occupational Gender Segregation in Industrial Countries. *Sociology, 46*(6), 1003–1019.

Jenkins, S. P. (2011). *Changing Fortunes: Income Mobility and Poverty Dynamics in Britain*. Oxford: Oxford University Press.

Jones, O. (2011). *Chavs—The Demonisation of the Working Class*. London: Verso.

Jonsson, J. O., Grusky, D. B., Di Carlo, M., Pollak, R., & Brinton, M. C. (2009). Microclass Mobility: Social Reproduction in Four Countries. *American Journal of Sociology, 114*(4), 977–1036.

Kadushin, C. (2012). *Understanding Social Networks: Theories, Concepts, and Findings*. Oxford: Oxford University Press.

Kalmijn, M. (1998). Intermarriage and Homogamy: Causes, Patterns, Trends. *Annual Review of Sociology, 24*, 395–421.

Kalmijn, M., & Flap, H. (2001). Assortative Meeting and Mating: Unintended Consequences of Organized Settings for Partner Choices. *Social Forces, 79*(4), 1289–1312.

Kerbo, H. R. (2003). *Social Stratification and Inequality: Class Conflict in Historical, Comparative and Global Perspective* (5th ed.). London: McGraw-Hill.

Khan, S. (2011). *Privilege: The Making of an Adolescent Elite at St Paul's School*. Princeton, NJ: Princeton University Press.

Khattab, N. (2016). The Ethno-Religious Wage Gap Within the British Salariat Class: How Severe Is the Penalty. *Sociology, 50*(4), 813–824.

Knoke, D. (2012). *Economic Networks*. Cambridge: Polity Press.

Knoke, D., & Yang, S. (2008). *Social Network Analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Kraaykamp, G., van Eijck, K., & Utlee, W. (2010). Status, Class and Culture in the Netherlands. In T. W. Chan (Ed.), *Social Status and Cultural Consumption*. Cambridge: Cambridge University Press.

Krebs, V. (2008). *A Network of Books About Recent US Politics Sold by the Online Bookseller Amazon.com*. Retrieved March 1, 2017, from http://www.orgnet.com/divided.html.

Lambert, P. S. (2012). *Comparative Scaling of Educational Categories by Homogamy—Analysis of UK Data from the BHPS*. Technical Paper 2012-1 of the DAMES Node, Data Management Through e-Social Science. Stirling: University of Stirling. Retrieved from www.dames.org.uk/publications.html.

Lambert, P. S. (2015). Advances in Data Management for Social Survey Research. In P. Halfpenny & R. Procter (Eds.), *Innovations in Digital Research Methods* (pp. 105–122). London: Sage.

Lambert, P. S. (2016). *CAMSIS Project: Files for Distribution Covering Occupational Unit Codes and Translations*. Retrieved December 1, 2016, from http://www.camsis.stir.ac.uk/occunits/distribution.html.

Lambert, P. S., & Bihagen, E. (2014). Using Occupation-Based Social Classifications. *Work, Employment and Society, 28*(3), 481–494.

Lambert, P. S., Griffiths, D., Bihagen, E., Zijdeman, R. L., Gayle, V., & Tranmer, M. (2014). *Is Britain Pulling Apart? Evidence from the Analysis of Social Distance.* Stirling: Technical Paper of the 'Is Britain Pulling Apart' Research Project, School of Applied Social Science, University of Stirling. Retrieved from www.camsis.stir.ac.uk/pullingapart.

Lambert, P. S., & Prandy, K. (2002). *Intra-Generational Mobility and Social Distance: Work History Analysis and Occupational Structure*. Paper Presented at the ESRC Seminar Series: Occupational Information in Social and Economic Research: Occupational Careers.

Lambert, P. S., Tan, K. L. L., Gayle, V., Prandy, K., & Bergman, M. M. (2008). The Importance of Specificity in Occupation-Based Social Classifications. *International Journal of Sociology and Social Policy, 28*(5/6), 179–192.

Lambert, P. S., Tan, K. L. L., Turner, K. J., Gayle, V., Prandy, K., & Sinnott, R. O. (2007). Data Curation Standards and Social Science Occupational Information Resources. *International Journal of Digital Curation, 2*(1), 73–91.

Lambert, P. S., Zijdeman, R. L., Maas, I., van Leeuwen, M. H. D., & Prandy, K. (2013). The Construction of HISCAM: A Stratification Scale Based on Social Interactions for Historical Research. *Historical Methods, 46*(2), 77–89.

Lampard, R. (1997). Party Political Homogamy in the United Kingdom. *European Sociological Review, 13*(*1*), 79–99.

Laumann, E. O. (1966). *Prestige and Association in an Urban Community*. Indianapolis, IN: Bobbs-Merrill.

Laumann, E. O. (1973). *Bonds of Pluralism: The Form and Substance of Urban Social Networks*. New York: Wiley.

Laumann, E. O., & Guttman, L. (1966). The Relative Associational Contiguity of Occupations in an Urban Setting. *American Sociological Review, 31*, 169–178.

Laumann, E. O., & Senter, R. (1976). Subjective Social Distance, Occupational Stratification, and Forms of Status and Class Consciousness: A Cross-National Replication and Extension. *American Journal of Sociology, 81*(6), 1304–1338.

Laurison, D., & Friedman, S. (2016). The Class Pay Gap in Higher Professional and Managerial Occupations. *American Sociological Reiwew, 81*(4), 668–695.

Leiulfsrud, H., Bison, I., & Jensberg, H. (2005). *Social Class in Europe: European Social Survey 2002/3*. Trondheim: NTNU Samfunnsforskning/NTNU Social Research Ltd.

Levine, J. H. (1972). A Two-Parameter Model of Interaction in Father-Son Status Mobility. *Behavioral Science, 18*, 455–465.

Levine, J. H. (1990). Measuring Occupational Stratification Using Loglinear Distance Models. In R. L. Breiger (Ed.), *Social Mobility and Social Structure*. Cambridge: Cambridge University Press.

Levine, J. H., & Spadaro, J. (1988). Occupational Mobility: A Structural Model. In B. Wellman & S. D. Berkowitz (Eds.), *Social Structures: A Network Approach* (pp. 452–475). Cambridge: Cambridge University Press.

Li, Y. (Ed.). (2015). *Handbook of Research Methods and Applications in Social Capital*. Cheltenham: Edward Elgar.

Li, Y., Savage, M., & Pickles, A. (2003). Social Capital and Social Exclusion in England and Wales (1972–1999). *British Journal of Sociology, 54*(4), 497–426.

Liao, T. F. (2006). Measuring and Analyzing Class Inequality with the Gini Index Informed by Model-Based Clustering. *Sociological Methodology, 36*(1), 201–224.

Lin, N. (1999). Social Networks and Status Attainment. *Annual Review of Sociology, 25*, 467–488.

Lin, N., & Dumin, M. (1986). Access to Occupations Through Social Ties. *Social Networks, 8*, 365–385.

Lin, N., Ensel, W. M., & Vaughn, J. C. (1981). Social Resources, Strength of Ties and Occupational Status Attainment. *American Sociological Review, 46*(4), 393–405.

Lipset, S. M., & Bendix, R. (1959). *Social Mobility in Industrial Society*. Berkeley, CA: University of California Press.

Lizardo, O. (2006). How Cultural Tastes Shape Personal Networks. *American Sociological Review, 71*(5), 778–807.

Long, J. S. (2009). *The Workflow of Data Analysis Using Stata*. Boca Raton: CRC Press.

Longhi, S., & Nandi, A. (2015). *A Practical Guide to Using Panel Data*. London: Sage.

Lui, Y., & Grusky, D. B. (2013). The Payoff to Skill in the Third Industrial Revolution. *American Journal of Sociology, 118*(5), 1330–1374.

Luijkx, R. (1994). *Comparative Loglinear Analyses of Social Mobility and Heterogamy*. Tilburg: Tilburg University Press.

Lupton, R., Heath, N., & Salter, E. (2009). Education: New Labour's Top Priority. In J. Hills, T. Sefton, & K. Stewart (Eds.), *Towards a More Equal Society? Poverty, Inequality and Policy Since 1997* (pp. 71–90). Bristol: Policy Press.

Mare, R. D. (1991). Five Decades of Educational Assortative Mating. *American Sociological Review, 56*, 15–32.

Marin, A., & Wellman, B. (2011). Social Network Analysis: An Introduction. In J. Scott & P. J. Carrington (Eds.), *The Sage Handbook of Social Network Analysis* (pp. 11–25). London: Sage.

Marsden, P. V. (2011). Survey Methods for Network Data. In J. Scott & P. J. Carrington (Eds.), *The Sage Handbook of Social Network Analysis* (pp. 370–388). London: Sage.

Marshall, G., Rose, D., Newby, H., & Vogler, C. (1988). *Social Class in Modern Britain*. London: Routledge.

Martin, J., & Roberts, C. (1984). *Women and Employment: A Lifetime Perspective*. London: HMSO.

Marvin, D. M. (1918). Occupational Propinquity as a Factor in Marriage Selection. *Journal of the American Statistical Association, 16*(123), 131–150.

McDonald, K. I. (1972). MDSCAL and Distances Between Socio-Economic Groups. In K. Hope (Ed.), *The Analysis of Social Mobility: Methods and Approaches* (pp. 211–234). London: Clarendon.

McKenzie, L. (2015). *Getting By: Estates, Class and Culture in Austerity Britain*. Bristol: Policy Press.

McKnight, A., & Elias, P. (1997). A Database of Information on Unit Groups of the Standard Occupational Classification. In D. Rose & K. O'Reilly (Eds.), *Constructing Classes* (pp. 116–145). Colchester: University of Essex.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology, 27*, 415–444.

Mercklé, P. (2012). Who Is Responsible for the Disappearance of Social Classes? *Bulletin de Methodologie Socilogique, 116*(1), 67–75.

Miller, R. L. (1998). Worklife Mobility Typologies as Background to Current Class Position—A Research Note. *Economic and Social Review, 29*(3), 259–284.

Mills, C. (2007). Unequal Jobs: Job Quality and job Satisfaction. In P. McGovern, S. Hill, C. Mills, & M. White (Eds.), *Market, Class and Employment* (pp. 233–282). Oxford: Oxford University Press.

Mills, C. (2014). The Great British Class Fiasco: A Comment on Savage et al. *Sociology, 48*(3), 437–444.

Mills, C., & Evans, G. (2003). Employment Relations, Employment Conditions and the NS-SEC. In D. Rose & D. J. Pevalin (Eds.), *A Researchers Guide to the National Statistics Socio-economic Classification* (pp. 77–106). London: Sage.

Minnesota Population Center. (2012). *IPUMS USA: Integrated Occupation and Industry Codes and Occupational Standing Variables in the IPUMS*. Retrieved March 3, 2012, from http://usa.ipums.org/usa/chapter4/chapter4.shtml.

Minnesota Population Center. (2015). *Integrated Public Use Microdata Series, International: Version 6.4 [Machine-Readable Database]*. Minneapolis, MN: University of Minnesota.

Mirowsky, J., & Ross, C. E. (2003). *Education, Social Status and Health*. Hawthorne, NY: Aldine de Gruyter.

Modood, T. (2005). The Educational Attainments of Ethnic Minorities in Britain. In G. C. Loury, T. Modood, & S. M. Teles (Eds.), *Ethnicity, Social Mobility and Public Policy: Comparing the US and UK* (pp. 288–308). Cambridge: Cambridge University Press.

Mohler, P. P., Pennell, B.-E., & Hubbard, F. (2008). Survey Documentation: Toward Professional Knowledge Management in Sample Surveys. In E. De Leeuw, J. Hox, & D. A. Dillman (Eds.), *International Handbook of Survey Methodology* (pp. 403–420). Hove: Psychology Press.

More, D. M., & Suchner, R. W. (1976). Occupational Situs, Prestige and Stereotypes. *Work and Occupations, 3*(2), 169–186.

Moreno, J. L. (1953). *Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Sociodrama*. New York: Beacon House.

Morris, R. T., & Murphy, R. J. (1959). The Situs Dimension in Occupational Structure. *American Sociological Review, 24*(2), 231–239.

Mouw, T. (2003). Social Capital and Finding a Job: Do Contacts Matter? *American Sociological Review, 68*(6), 868–898.

North Atlantic Population Project and Minnesota Population Center. (2008). *NAPP: Complete Count Microdata. NAPP Version 2.0 [Computer Files]*. Minneapolis, MN: Minnesota Population Center [distributor]. Retrieved from http://www.nappdata.org.

O*NET. (2008). *O*NET Online: Occupational Information Network*. Retrieved December 1, 2008, from http://online.onetcenter.org/.

OECD. (2004). *OECD Handbook for Internationally Comparative Educational Statistics*. Paris: Organisation for Economic Cooperation and Development.

OECD. (2008). *Growing Unequal*. Paris: Organisation for Economic Cooperation and Development.

Oesch, D. (2006). *Redrawing the Class Map: Stratification and Institutions in Britain, Germany, Sweden and Switzerland*. Basingstoke: Palgrave.

Oesch, D. (2013). *Occupational Change in Europe: How Technology and Education Transform the Job Structure*. Oxford: Oxford University Press.

ONS. (2010). *Standard Occupational Classification 2010. Volume 1: Structure and Descriptions of Unit Groups*. Basingstoke: Palgrave Macmillan (with the Office for National Statistics).

ONS. (2011). *National Population Projections, 2010-Based Projections*. London: Office for National Statistics. Retrieved from www.ons.gov.uk.

ONS. (2012, September). *Labour Market Statistics Data Tables (Excel Spreadsheets)*. London: Office for National Statistics. Retrieved from www.ons.gov.uk.

ONS. (2017). *Labour Force Survey—User Guidance*. Retrieved March 1, 2017, from https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyuserguidance.

OPCS. (1990). *Standard Occupational Classification, Volume 1: Structure and Definition of Major, Minor and Unit Groups*. London: Office for Population Censuses and Surveys.

Packard, V. (1959). *The Status Seekers*. London: Penguin.

Pahl, R. (2000). *On Friendship*. Cambridge: Polity Press.

Parkin, F. (1972). *Class Inequality and Political Order: Social Stratification in Capitalist and Communist Societies*. London: MacGibbon and Kee.

Payne, G. (2012). Labouring Under a Misapprehension? Politicians' Perceptions and the Realities of Structural Social Mobility in Britain, 1995–2010. In P. S. Lambert, R. Connelly, R. M. Blackburn, & V. Gayle (Eds.), *Social Stratification: Trends and Processes* (pp. 223–242). Aldershot: Ashgate.

Payne, G. (2017). *The New Social Mobility: How the Politicians Got it Wrong*. Bristol: Policy Press.

Pearce, A., & Gambrell, D. (2016). *Bloomberg: This Chart Shows Who Marries CEOS, Doctors, Chefs and Janitors*. Retrieved June 1, 2016, from http://www.bloomberg.com/graphics/2016-who-marries-whom/.

Penn, R. D. (1985). *Skilled Workers in the Class Structure*. Cambridge: Cambridge University Press.

Penn, R. D. (2006). *Social Change and Economic Life in Britain*. Bologna: Homeless Book.

Penn, R. D., & Lambert, P. S. (2009). *Children of International Migrants in Europe: Comparative Perspectives*. Basingstoke: Palgrave.

Pfeffer, F. T. (2008). Persistent Inequality in Educational Attainment and Its Institutional Context. *European Sociological Review, 24*(5), 534–565.

Platt, L. (2005). The Intergenerational Social Mobility of Minority Ethnic Groups. *Sociology, 39*(3), 445–462.

Platt, L. (2011). *Understanding Inequalities: Stratification and Difference*. Cambridge: Polity.

Playford, C., & Gayle, V. (2016). The Concealed Middle? An Exploration of Ordinary Young People and School GCSE Subject Area Attainment. *Journal of Youth Studies, 19*(2), 149–168.

Playford, C., Gayle, V., Connelly, R., & Gray, A. J. G. (2016). Administrative Social Science Data: The Challenge of Reproducible Research. *Big Data and Society, 3*(2), 1–13.

Pollock, G. (2007). Holistic Trajectories: A Study of Combined Employment, Housing and Family Careers by Using Multiple-Sequence Analysis. *Journal of the Royal Statistical Society Series A, 170*(1), 167–183.

Portes, A., & Rumbaut, R. G. (2001). *Legacies: The Story of the Immigrant Second Generation*. Berkeley and Los Angeles: University of California Press.

Prandy, K. (1979). Ethnic Discrimination in Employment and Housing. *Ethnic and Racial Studies, 2*(1), 66–79.

Prandy, K. (1986). Similarities of Life-Style and the Occupations of Women. In R. Crompton & M. Mann (Eds.), *Gender and Stratification* (pp. 137–153). Cambridge: Polity Press.

Prandy, K. (1990). The Revised Cambridge Scale of Occupations. *Sociology, 24*(4), 629–655.

Prandy, K. (1998). Class and Continuity in Social Reproduction: An Empirical Investigation. *Sociological Review, 46*(2), 340–364.

Prandy, K. (1999). Class, Stratification and Inequalities in Health: A Comparison of the Registrar-General's Social Classes and the Cambridge Scale. *Sociology of Health & Illness, 21*(4), 466–484.

Prandy, K. (2000). Class, the Stratification Order and Party Identification. *British Journal of Political Science, 30*, 237–258.

Prandy, K. (2002). Ideal Types, Stereotypes and Classes. *British Journal of Sociology, 53*(4), 583–601.

Prandy, K., & Blackburn, R. M. (1997). Putting Men and Women into Classes: But Is That Where They Belong? A Comment on Evans. *Sociology, 31*(1), 143–152.

Prandy, K., & Bottero, W. (1998). The Use of Marriage Data to Measure the Social Order in Nineteenth-Century Britain. *Sociological Research Online, 3*(1), U43–U54.

Prandy, K., & Bottero, W. (2000). Reproduction Within and Between Generations—The Example of Nineteenth-Century Britain. *Historical Methods, 33*(1), 4–15.

Prandy, K., & Jones, F. L. (2001). An International Comparative Analysis of Marriage Patterns and Social Stratification. *International Journal of Sociology and Social Policy, 21*, 165–183.

Prandy, K., & Lambert, P. S. (2003). Marriage, Social Distance and the Social Space: An Alternative Derivation and Validation of the Cambridge Scale. *Sociology, 37*(3), 397–411.

Prandy, K., Lambert, P. S., & Bergman, M. M. (2002, July 7–13). *National Contexts and Cross-National Comparisons of Structures of Social Stratification*. Paper Presented at the ISA IVth World Congress of Sociology, Brisbane, RC20 Panel 16.

Prandy, K., Unt, M., & Lambert, P. S. (2004). *Not by Degrees: Education and Social Reproduction in Twentieth-Century Britain*. Paper Presented at the ISA

RC28 Research Committee on Social Stratification and Mobility, Neuchatel, 7–9 May 2004.
Putnam, R. D. (2000). *Bowling Alone: The Collapse and Revival of American Community*. London: Simon and Schuster.
Putnam, R. D., Leonardi, R., & Nanetti, R. Y. (1993). *Making Democracies Work: Civic Traditions in Modern Italy*. Princeton, NY: Princeton University Press.
R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata, Third Edition (Two-Volume Set)*. College Station, TX: Stata Press.
Raffe, D., & Croxford, L. (2015). How Stable Is the Stratification of Higher Education in England and Scotland? *British Journal of Sociology of Education, 36*(2), 2015.
Roemer, J. (2012). What Is the Justification for Studying Intergenerational Mobility of Socioeconomic Status? In J. Ermisch, M. Jantti, & T. Smeeding (Eds.), *From Parents to Children: The Intergenerational Transmission of Advantage* (pp. 482–487). New York: Russell Sage Foundation.
Rose, D., & Harrison, E. (2007). The European Socio-Economic Classification: A New Social Class Scheme for Comparative European Research. *European Societies, 9*(3), 459–490.
Rose, D., & Harrison, E. (Eds.). (2010). *Social Class in Europe: An Introduction to the European Socio-economic Classification*. London: Routledge.
Rose, D., & Pevalin, D. J. (Eds.). (2003). *A Researcher's Guide to the National Statistics Socio-Economic Classification*. London: Sage.
Rost, J. (1988). Rating Scale Analysis with Latent Class Models. *Psychometrika, 53*, 327–348.
Routh, G. (1980). *Occupation and Pay in Great Britain, 1906–79*. London: Macmillan.
Ruggles, S., Sobek, M., Alexander, T., Fitch, C. A., Goeken, R., Hall, P. K., King, M., & Ronnander, C. (2009). *Integrated Public Use Microdata Series: Version 4.0 [Machine-Readable Database]*. Minneapolis, MN: Minnesota Population Center [producer and distributor]. Retrieved from http://usa.ipums.org/usa/.
Rytina, S. (1992). Scaling the Intergenerational Continuity of Occupation: Is Occupational Inheritance Ascriptive After All? *American Journal of Sociology, 97*(6), 1658–1688.
Rytina, S. (2000). *A Keynote of Hierarchy, an Echo of Class: The Full Texture of Mobility in England and Wales.* Paper Presented at the RC33—Fifth International Conference on Logic and Methodology, Cologne.

Rytina, S. (2010). *Sticky Struggles: Stratification as Emergent from Network Infrastructure*. Presentation to Centre on Skills, Knowledge and Organisational Performance (SKOPE), University of Oxford.

Rytina, S., & Bianchi, S. M. (1984). Occupational Reclassification and Changes in Distribution by Gender. *Monthly Labor Review, 107*(3), 11–17.

Sacker, A., Bartley, M., Firth, D., & Fitzpatrick, R. (2001). Dimensions of Social Inequality in the Health of Women in England: Occupational, Material and Behavioural Pathways. *Social Science and Medicine, 52*(5), 763–781.

Saunders, P. (2010). *Social Mobility Myths*. London: CIVITAS: Institute for the Study of Civil Society.

Savage, M., Cunningham, N., Devine, F., Friedman, S., Laurison, D., McKenzie, L., Miles, A., Snee, H., & Wakeling, P. (2015). *Social Class in the 21st Century*. London: Pelican Books.

Savage, M., Devine, F., Cunningham, N., Taylor, M., Li, Y., Hjellbrekke, J., Le Roux, B., Friedman, S., & Miles, A. (2013). A New Model of Social Class: Findings from the BBC's Great British Class Survey Experiment. *Sociology, 47*(2), 219–250.

Sayer, A. (2005). *The Moral Significance of Class*. Cambridge: Cambridge University Press.

Sayer, A. (2016). *Why We Can't Afford the Rich*. Bristol: Policy Press.

Scase, R. (1992). *Class*. Buckingham: Open University Press.

Schmidt-Catran, A. W., & Fairbrother, M. (2016). The Random Effects in Multilevel Models: Getting Them Wrong and Getting Them Right. *European Sociological Review, 32*(1), 23–38.

Schneider, S. L. (2008). *The International Standard Classification of Education (ISCED-97). An Evaluation of Content and Criterion Validity for 15 European Countries*. Mannheim: MZES.

Schuller, T., Preston, J., Hammond, C., Basset-Grundy, A., & Brynner, J. (2004). *The Benefits of Learning: The Impact of Education on Health, Family Life and Social Capital*. London: Routledge.

Scott, J. (2017). *Social Network Analysis: A Handbook* (4th ed.). London: Sage.

Scott, J., & Carrington, P. J. (Eds.). (2011). *The Sage Handbook of Social Network Analysis*. London: Sage.

Scott, J., Dex, S., & Joshi, H. (Eds.). (2008). *Women and Employment: Changing Lives and New Challenges*. Cheltenham: Edward Elgar.

Scottish Government. (2016). *The Scottish Index of Multiple Deprivation*. Retrieved March 1, 2016, from http://www.gov.scot/Topics/Statistics/SIMD.

Shavit, Y., Arum, R., & Gamoran, A. (Eds.). (2007). *Stratification in Higher Education: A Comparative Study*. Stanford, CA: Stanford University Press.

Shavit, Y., & Blossfeld, H. P. (Eds.). (1993). *Persistent Inequality: Changing Educational Attainment in Thirteen Countries*. Boulder, CO: Westview.

Shaw, A. (2000). *Kinship and Continuity: Pakistani Families in Britain*. Amsterdam: Harwood Academic.

Shaw, M., Galobardes, B., Lawlor, D. A., Lynch, J., Wheeler, B., & Davey Smith, G. (2007). *The Handbook of Inequality and Socioeconomic Position: Concepts and Measures*. Bristol: Policy Press.

Simon, J. L., & Boggs, R. (1995). Trends in the Quantities of Education: USA and Elsewhere. In J. L. Simon (Ed.), *The State of Humanity* (pp. 208–223). Oxford: Blackwell.

Skopek, J., Schulz, F., & Blossfeld, H. P. (2011). Who Contacts Whom? Educational Homophily in Online Mate Selection. *European Sociological Review, 27*(2), 180–195.

Skvoretz, J. (2013). Diversity, Integration, and Social Ties: Attraction Versus Repulsion as Drivers of Intra- and Intergroup Relations. *American Journal of Sociology, 119*(2), 486–517.

Smith, J. A., McPherson, M., & Smith-Lovin, L. (2014). Social Distance in the United States: Sex, Race, Religion, Age and Education Homophily Among Confidants, 1985 to 2004. *American Sociological Reiwew, 79*(3), 432–456.

Smits, J. (2003). Social Closure Among the Higher Educated: Trends in Educational Homogamy in 55 Countries. *Social Science Research, 32*(2), 251–277.

Smits, J., Ultee, W., & Lammers, J. (1999). Occupational Homogamy in Eight Countries of the European Union, 1975–89. *Acta Sociologica, 42*(1), 55–68.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (2nd ed.). London: Sage.

Sokoloff, N. J. (1992). *Black Women and White Women in the Professions: Occupational Segregation by Race and Gender 1960–1980*. New York: Routledge.

Sorokin, P. A. (1927). *Social Mobility*. New York: Harper and Brothers.

Stansfeld, S. A., Rasul, F., Head, J., & Singleton, N. (2011). Occupation and Mental Health in a National UK Survey. *Social Psychiatry and Psychiatric Epidemiology, 46*(2), 101–110.

StataCorp. (2015). *Stata Statistical Software: Release 14*. College Station, TX: StatCorp LLC.

Statistics Finland. (1996). *The Finnish Census Panel Data*. Helsinki: Statistics Finland.

Statistics Sweden. (2016). *MONA—Microdata Online Access*. Stockholm: Statistics Sweden. Retrieved from www.scb.se/mona-en.

Stewart, A., Prandy, K., & Blackburn, R. M. (1973). Measuring the Class Structure. *Nature, 245*, 415–417.

Stewart, A., Prandy, K., & Blackburn, R. M. (1980). *Social Stratification and Occupations*. London: Macmillan.

Steyvers, M., & Tenenbaum, J. B. (2005). The Large Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science, 29*(1), 41–78.

Strangleman, T. (2012). Work Identity in Crisis? Rethinking the Problem of Attachment and Loss at Work. *Sociology, 46*(3), 411–425.

Stutzer, A., & Frey, B. S. (2006). Does Marriage Make People Happy, or do Happy People Get Married? *The Journal of Socio-Economics, 35*, 326–347.

Szreter, S. R. S. (1984). The Genesis of the Registrar-General: Social Classification of Occupations. *British Journal of Sociology, 35*(4), 522–546.

Swift, A. (2004). Would Perfect Mobility be Perfect? *European Sociological Review, 20*(1), 1–11.

Tahlin, M. (2007). Class Clues. *European Sociological Review, 23*(5), 557–572.

Therborn, G. (2013). *The Killing Fields of Inequality*. Cambridge: Polity Press.

Tomlinson, M., Walker, R., & Williams, G. (2008). Measuring Poverty in Britain as a Multi-Dimensional Concept, 1991–2003. *Journal of Social Policy, 37*(4), 597–620.

Torssander, J., & Erikson, R. (2010). Stratification and Mortality—A Comparison of Education, Class, Status, and Income. *European Sociological Review, 26*(4), 465–474.

Toubol, J., & Larsen, A. G. (2017). Mapping the Social Class Structure: From Occupational Mobility to Social Class Categories Using Network Analysis. *Sociology, 51*(6), 1257–1276.

Treiman, D. J. (1977). *Occupational Prestige in Comparative Perspective*. New York: Academic Press.

Treiman, D. J. (2009). *Quantitative Data Analysis: Doing Social Research to Test Ideas*. New York: Jossey Bass.

Treiman, D. J., & Ganzeboom, H. B. G. (2000). The Fourth Generation of Comparative Stratification Research. In S. R. Quah & A. Sales (Eds.), *The International Handbook of Sociology* (pp. 123–150). London: Sage.

Turner, H., & Firth, D. (2007). *Generalized Nonlinear Models in R: An Overview of the GNM Package*. Southampton: ESRC National Centre for Research Methods, Paper 6/07 of the NCRM Working Paper Series.

Tyler, I. (2013). *Revolting Subjects: Social Abjection and Resistance in Neoliberal Britain*. London: Zed Books.

Ultee, W. C., & Luijkx, R. (1990). Educational Heterogamy and Father-to-Son Occupational Mobility in 23 Industrial Nations: General Societal Openness or Compensatory Strategies of Reproduction. *European Sociological Review, 6*, 125–150.

University of Essex, & Institute for Social and Economic Research. (2010, July). *British Household Panel Survey: Waves 1–18, 1991–2009 [Computer File]* (7th ed.). Colchester, Essex: UK Data Archive [distributor], SN: 5151.

Upright, C. B. (2004). Social Capital and Cultural Participation: Spousal Influences on Attendance at Arts Events. *Poetics, 32*(2), 129–143.

van de Werfhorst, H. G., de Graaf, N. D., & Kraaykamp, G. (2001). Intergenerational Resemblance in Field of Study in the Netherlands. *European Sociological Review, 17*(3), 275–293.

van de Werfhorst, H. G., & Luijkx, R. (2010). Educational Field of Study and Social Mobility: Disaggregating Social Origin and Education. *Sociology, 44*(4), 695–715.

van der Gaag, M. P. J., Snijders, T. A. B., & Flap, H. (2008). Position Generator Measures and Their Relationship to Other Social Capital Measures. In N. Lin & B. H. Erikson (Eds.), *Social Capital: An International Research Program* (pp. 27–48). Oxford: Oxford University Press.

van der Lippe, T., & van Dijk, L. (2002). Comparative Research on Women's Employment. *Annual Review of Sociology, 28*, 221–241.

Verhaeghe, P.-P., Pattyn, E., Bracke, P., Verhaeghe, M., & Van de Putte, B. (2012). The Association Between Network Social Capital and Self-Rated Health: Pouring Old Wine in New Bottles? *Health and Place, 18*(2), 358–365.

Vermunt, J. K. (1997). *lEM: A General Program for the Analysis of Categorical Data*. Tilburg: Tilburg University.

Wageindicator. (2013). *Wageindicator.org: World Database of Occupations (WISCO)*. Retrieved March 1, 2013, from http://www.wageindicator.org/main/researchlab/occupation-data-base.

Waldinger, R., & Lichter, M. I. (2003). *How the Other Half Works: Immigration and the Social Organization of Labor*. Berkeley: University of California Press.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

Weber, M. (1968[1922]). *Economy and Society*. Berkeley: University of California Press.

Weeden, K. A., & Grusky, D. B. (2012). The Three Worlds of Inequality. *American Journal of Sociology, 117*(6), 1723–1785.

Weeden, K. A., Kim, Y.-M., Di Carlo, M., & Grusky, D. B. (2007). Social Class and Earnings Inequality. *American Behavioral Scientist, 50*(5), 702–736.

Weller, S. C., & Romney, A. K. (1990). *Metric Scaling: Correspondence Analysis*. London: Sage.

Wellman, B. (1979). The Community Question. *American Journal of Sociology, 84*, 1201–1231.

Wellman, B., & Berkowitz, S. D. (Eds.). (1988). *Social Structures: A Network Approach*. Cambridge: Cambridge University Press.

White, H. C. (1992). *Identity and Control: A Structural Theory of Social Action*. Princeton, NJ: Princeton University Press.

Wilkinson, R. G., & Pickett, K. E. (2009). *The Spirit Level: Why More Equal Societies Almost Always do Better*. London: Allen Lane.

Williams, M. (2013). Occupations and British Wage Inequality, 1970s–2000s. *European Sociological Review, 29*(4), 841–857.

Willis, P. (1977). *Learning to Labour: How Working Class Kids Get Working Class Jobs*. Westmead, Hants: Saxon House.

Wong, R. S. K. (2010). *Association Models*. Los Angeles: Sage.

Wright, E. O. (1997). *Class Counts: Comparative Studies in Class Analysis*. Cambridge: Cambridge University Press.

Wright, E. O. (Ed.). (2005). *Approaches to Class Analysis*. Cambridge: Cambridge University Press.

Xu, X., Ji, J., & Tung, Y.-Y. (2000). Social and Political Assortative Mating in Urban China. *Journal of Family Issues, 21*(1), 47–77.

Yamaguchi, K. (1983). The Structure of Intergenerational Occupational-Mobility: Generality and Specificity in Resources, Channels, and Barriers. *American Journal of Sociology, 88*(4), 718–745.

Yamaguchi, K. (1990). Homophily and Social Distance in the Choice of Multiple Friends—An Analysis Based on Conditionally Symmetric Log-Bilinear Association Models. *Journal of the American Statistical Association, 85*(410), 356–366.

Young, M. (1958). *The Rise of the Meritocracy 1870–2033*. Harmondsworth: Penguin.

Zijdeman, R. L., & Lambert, P. S. (2010). Measuring Social Structure in the Past: A Comparison of Historical Class Schemes and Occupational Stratification Scales on Dutch 19th and Early 20th Century Data. *Belgian Review of Contemporary History, 40*(1), 1111–1141.

Zimdars, A., Sullivan, A., & Heath, A. (2009). Elite Higher Education Admissions in the Arts and Sciences: Is Cultural Capital the Key? *Sociology, 43*(4), 648–666.

# Index[1]



[1] Note: Page numbers followed by 'n' refer to notes.



P. Lambert, D. Griffiths, *Social Inequalities and Occupational Stratification*,
https://doi.org/10.1057/978-1-137-02253-0