DOCTORAL THESIS

Measurement of work group diversity

Jeremy Dawson

MEASUREMENT OF WORK GROUP DIVERSITY

JEREMY FRANCIS DAWSON
Doctor of Philosophy

ASTON UNIVERSITY
September 2011

ASTON UNIVERSITY

MEASUREMENT OF WORK GROUP DIVERSITY

JEREMY FRANCIS DAWSON
Doctor of Philosophy
September 2011

THESIS SUMMARY

Whilst research on work group diversity has proliferated in recent years, relatively little attention has been paid to the precise definition of diversity or its measurement. One of the few studies to do so is Harrison and Klein's (2007) typology, which defined three types of diversity – separation, variety and disparity – and suggested possible indices with which they should be measured.

However, their typology is limited by its association of diversity types with variable measurement, by a lack of clarity over the meaning of variety, and by the absence of a clear guidance about which diversity index should be employed.

In this thesis I develop an extended version of the typology, including four diversity types (separation, range, spread and disparity), and propose specific indices to be used for each type of diversity with each variable type (ratio, interval, ordinal and nominal). Indices are chosen or derived from first principles based on the precise definition of the diversity type. I then test the usefulness of these indices in predicting outcomes of diversity compared with other indices, using both an extensive simulated data set (to estimate the effects of mis-specification of diversity type or index) and eight real data sets (to examine whether the proposed indices produce the strongest relationships with hypothesised outcomes).

The analyses lead to the conclusion that the indices proposed in the typology are at least as good as, and usually better than, other indices in terms of both measuring effect sizes and power to find significant results, and thus provide evidence to support the typology. Implications for theory and methodology are discussed.

Key Words: Work group diversity, Diversity Measurement, Typology, Simulation

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank those who helped and supported me throughout the writing of this thesis. In particular, I would like to thank my supervisors: Professor Sam Aryee, for guiding me through the writing and submission of the thesis, Dr Michael Grojean for helping me develop my thoughts and plans in the earlier days, and Professor Felix Brodbeck for inspiring greater interest in work group diversity and for helping me access two data sets.

I would also like to thank my family, friends and other colleagues who have helped me through the process. In particular I would like to thank Professor Michael West for his constant support, inspiration and intellectual stimulation.

CONTENTS

LIST OF TABLES

Tables in Appendices

LIST OF FIGURES

CHAPTER 1

WHAT IS DIVERSITY AND WHY IS IT IMPORTANT?

1.0 Chapter introduction

In this chapter I introduce the topic of work group diversity, explaining why it has become increasingly important over recent years, and giving a brief overview of topics found in the diversity literature, including outcomes of diversity in work groups. I also demonstrate some of the confusion around definitions of diversity and related terms. In doing so, I highlight why there is a lack of clarity about how work group diversity should be measured, and introduce the objectives of this thesis, primarily being to provide a typology that should provide such clarity.

1.1 Diversity at work

Diversity is a subject of growing importance for researchers and employers alike. The rise of global mobility, the increasing range and availability of education and training, and the growing legal and ethical responsibility of employers regarding equal opportunities for employees of different backgrounds all contribute to an overall increase in the extent to which workforces, and work groups and teams, display greater levels of diversity across multiple attributes. In the UK, the increasing proportion of the population who are from ethnic minorities (4.6 million people in Great Britain in the 2001 census compared with 3.0 million in the 1991 census; Office for National Statistics, 1991, 2001), combined with recent legislation which has made age discrimination illegal (along with other existing legislation on discrimination on the grounds of sex or ethnicity), means that the overall workforce has become demographically more diverse over recent decades. Additionally, the increased emphasis in some sectors on multidisciplinary working (e.g. cancer care: Calman & Hine, 1995; product development: Sarin & O'Connor, 2009) has led to increased functional diversity in teams also.

Diversity of work teams, groups, and organisations has been a subject of management research for several decades (Williams & O'Reilly, 1998), but has become an increasingly prevalent topic, with increasing numbers of articles published each year and new theoretical directions being explored continually (Van Knippenberg & Schippers, 2007). In the period from 2000 to 2010, the *Academy of Management Journal* alone published over 50 articles concerned with work group diversity, signifying a research topic that is of great importance to researchers and practitioners alike.

Work group diversity is a fact of working life for most people, and employees in all sectors are encouraged to embrace diverse working groups. For managers, however, there are complications, as diversity can lead to both positive and negative outcomes. Consequently huge resources are put into the management of diversity in organisations (Cox & Blake, 1991), even though much about the way diversity affects work outcomes is unclear. Indeed, much of the literature on work group diversity reveals contrasting results about whether diversity is (or can be) a positive attribute for employers, employees and clients. As a result, research on diversity has proliferated over the last 20 years – to a large extent attempting to discover the conditions which enable diversity to affect outcomes positively (e.g. Stewart, 2006; Horwitz & Horwitz, 2007; Joshi & Roh, 2009; Joshi, Liao & Roh, 2011).

1.2 Outcomes of diversity

Diversity in work groups has been shown to be related to outcomes such as group performance, organisational performance, creativity and innovation, cooperation, conflict, communication, cohesiveness, quality of decision making, absenteeism, turnover, stress, depression, satisfaction, organisational citizenship behaviours, and organisational commitment (Williams & O'Reilly, 1998; van Knippenberg & Schippers, 2007). If the make-up of a work group can explain such varied outcomes as group and organisational performance, group processes, and group member health, turnover and attitudes, then there are clearly implications of which managers need to be aware: in particular, the fact that diversity appears to have positive effects on some outcomes, but negative effects on others, presents particular problems for which solutions need to be found. In some cases diversity has been shown to have different effects on the same outcomes; for example, some studies have shown positive effects of demographic diversity on work group performance (e.g. McLeod & Lobel, 1992; O'Reilly, Williams & Barsade, 1997), whereas other studies have shown negative effects of the same relationship (e.g. Watson, Johnson & Zgourides, 2002; Ely, 2004).

1.3 Definitions of diversity

Despite this wide interest, few researchers had attempted to give a precise definition of diversity until recently. There appears to have been an assumption that diversity as a concept is understood generally without the need for a definition; indeed, White (1986) wrote that "Diversity and segregation are characteristics of a population most individuals can sense intuitively. Diversity is variety" (p. 198).

This relatively simple explanation implies that the very word "diversity" means the same for organisational researchers as it does to any other user of the English language. The New Oxford Dictionary of English defines diversity as “the state of being diverse”; “diverse”, in turn is defined as “showing a great deal of variety; very different”. (Pearsall, 1998, pp. 539 and 538 respectively). Thus White’s (1986) comment that “Diversity is variety” is one that probably would be understood by most people. However, this definition is broad, and when looking at diversity of work groups (or any specific target), a more precise definition may be necessary. Indeed, the lack of a precise definition is not limited to the study of diversity in organisations: in examining the plethora of attempts to define diversity of species from an ecological perspective, Hurlbert (1971) declared diversity to be a “non-concept”.

The twenty-first century has seen a number of attempts to define work group diversity precisely. Jackson, Joshi and Erhardt (2003) define diversity as the “distribution of personal attributes among interdependent members of a work unit” (p. 802). This is simultaneously focused (referring only to interdependent groups) and slightly vague (in that the word “distribution” can take on various meanings). Harrison and Sin (2005) define diversity as “the collective amount of differences among members within a social unit” (p. 196). This is a more useful definition to researchers, as it is not focused on a single type of group (instead referring to any social unit), and rather than describing a “distribution” of differences, mentions the “collective amount” – this can still be conceptualised in a number of ways, but calculation could be relatively simple, e.g. the sum of all possible individual differences within the unit. More recently, Harrison and Klein (2007) defined diversity as “the distribution of differences among the members of a unit with respect to a common attribute” (p. 1200) – more similar to Jackson et al.’s definition than to Harrison and Sin’s. They do, however, go on to define three distinct types of diversity, *separation*, *variety* and *disparity*, which have more precise definitions - these will be discussed in more detail in Chapter 3.

So despite an increased awareness of the need to define diversity precisely, there is no absolute consensus about what this definition should be; the fact that different researchers mean different things by the term is perhaps indicative that “diversity” should not be thought of as a single construct, but as an overarching idea embracing a number of different definitions.

1.4 Similar terms and related concepts

Complicating the definition of diversity further is the fact that there are a substantial number of similar terms used in the literature, and sometimes these are used interchangeably with diversity. For example, “heterogeneity”, “dissimilarity”, “dispersion”, “inequality”, “agreement”, “consensus”, “concentration”, “deviation”, “variety”, and “variation” have all been used to refer to concepts that are close to, or even identical to, diversity (or its inverse). However, although many of these overlap considerably with diversity in meaning, very often they have precise definitions in their own right, and this means that the meanings (and therefore measurements) cannot necessarily be transplanted. For example, the terms “agreement” and “consensus” are often used to refer to the degree to which perceptions of something (e.g. leadership or team climate) are shared by group members (e.g. Kozlowski & Hattrup, 1992; Lindell, Brandt & Whitney, 1999). Whilst this is a form of diversity, it may not correspond exactly with a form which is represented by the range of views shared within the group. Thus whichever term is used, it is necessary to define it precisely in order to determine how it should be operationalised and measured.

It is also worth mentioning and distinguishing some related terms and concepts that are often used in the diversity literature but differ in focus from what is usually meant by "group diversity" in one or more ways. *Relational demography* refers to the extent to which individuals are different or similar from other group members, or a specific individual (e.g. supervisor or supervisee) on one or more demographic variables, e.g. age, sex, ethnicity (e.g. Tsui & O'Reilly, 1989; Tsui, Egan & O'Reilly, 1992; Chattopadhay, 1999; Chattopadhay & George, 2001; Chatman & Flynn, 2001; Chattopadhay et al., 2004; Chatman & O'Reilly, 2004). *Organisational demography*, on the other hand, usually refers to the makeup of an entire organisation (e.g. Pfeffer, 1984; Wagner, Pfeffer & O'Reilly, 1984; Wiersema & Bird, 1993). Work in this area may refer to diversity in a similar way, but the theoretical perspectives underlying the relationships between diversity and organisational outcomes are often quite different from those occurring at the group level. Finally, *group faultlines* refer to subdivisions of groups on two or more separate attributes simultaneously, as opposed to diversity measured on a single attribute at a time (e.g. Lau & Murnighan, 1998; Thatcher, Jehn & Zanutto, 2003; Shaw, 2004). In particular it examines whether diversity on multiple attributes coincides (creating a "faultline", or subdivisions, within in a group) or is cross-categorised (so that values of one attribute are not associated with values of another attribute). The recent development of group faultline research has started to play a significant role in the understanding of diversity, and therefore I will refer to faultlines at times; however, the detail of the nature and measurement of these faultlines is beyond the scope of this thesis so I will limit the main content to group diversity on one attribute at once.

1.5 Chapter summary and objectives of thesis

Given the wide range of outcomes of diversity, and the fact that it leads to both positive and negative outcomes, it is of little surprise that research in the area has increased over the decades, and that different theoretical perspectives to explain these relationships have been developed. What is perhaps more surprising is that few researchers have questioned the mechanisms by which diversity is measured within this research, and therefore this is the main focus of my research in this thesis. I seek to build on recent definitions and operationalisations of diversity in order to give researchers in the area clear guidelines about how they should be measuring diversity in their studies. As such, I develop an extended typology of diversity definition and measurement that is intended to cover most situations diversity researchers are likely to encounter.

The thesis proceeds as follows. In Chapter 2 I review the substantive literature on work group diversity, in order to gain an understanding of the theoretical frameworks that have guided such research and the results that have been found. Chapter 3 is a critique of Harrison and Klein's (2007) typology, which is the only one to date that makes any systematic attempt to align diversity measurement with definitions. In Chapter 4 I review diversity indices that have been used in the literature so far, while in Chapter 5 I use the findings of the previous chapters to develop the extended typology, and formulate hypotheses about its usefulness.

In Chapter 6 I introduce the methods used to test these hypotheses, which include tests on both simulated and real data. Chapter 7 includes the results from the simulated data, and Chapter 8 the results from the analysis of real data. In Chapter 9 I bring together the results of the two studies, identifying what the results can tell us about measurement of work group diversity, and the implications this has for both theory and methods of future research.

CHAPTER 2

WORK GROUP DIVERSITY IN THE ORGANISATIONAL LITERATURE

2.0 Chapter introduction

In this chapter I review the literature on work group diversity, in order to understand what organisational researchers mean by diversity, the reasons it is expected to be linked to outcomes, and conclusions that can be drawn from existing studies.

I begin by examining the two main theoretical perspectives of diversity (social categorisation and information/decision making) and models that integrate them; following that I summarise the main findings to date, beginning with meta-analyses and narrative reviews, and then outline results from a wide range of studies examining different sources of diversity and different outcomes, highlighting the substantial inconsistencies in the literature. I conclude by examining some of the potential reasons for these inconsistencies, including moderating variables, and varying conceptualisations and measurements of diversity itself.

2.1 Theoretical perspectives of diversity

Research in work group diversity has proposed and used a variety of theoretical perspectives to explain the effects of diversity on outcomes (Mannix & Neale, 2005). These are grouped broadly into two areas: the social categorisation perspective and the information/decision making perspective; rather than being clear theories in their own right, however, each of these is a collection of theoretical positions that broadly lead to similar predictions about diversity. These alternative frameworks are important to consider in parallel, because they often predict differential effects of diversity on team outcomes.

The social categorisation perspective is usually considered either from the perspective of social identity theory or self-categorisation theory, but may also use the similarity-attraction paradigm. Self-categorisation (e.g. Turner, 1985; Turner et al., 1987) is a cognitive process whereby individuals define themselves in terms of membership of social groups – e.g. sex, race, religion, social class. Within a group context, self-categorisations become more or less salient depending on the distribution of the characteristic under consideration. For example, ethnicity would not be a very salient factor in an ethnically homogeneous team, but would be more so in a mixed team; even more still for minority ethnic group members in a team dominated by one ethnic group. Social identity theory (e.g. Tajfel & Turner, 1986; Hogg & Abrams, 1988) builds on this by suggesting the formation of subgroups creates motivational factors for group members. When individuals in diverse groups categorise themselves and others in this way, subgroups are distinguished, with ingroup members being favoured and trusted more than outgroup members. As a result, the perspective suggests that work groups would function more smoothly, and individual members would feel more satisfied with and attracted to the group, when they are more homogeneous in nature (Van Knippenberg & Schippers, 2007). A large number of studies of work group diversity have based their approach partly or wholly on self-categorisation and social identity theory (e.g. Tsui, Egan & O'Reilly, 1992; Harrison, Price & Bell, 1998; Pelled, Eisenhardt & Xin, 1999; Chatman & Flynn, 2001).

Related to this is the similarity-attraction paradigm (Byrne, 1971; Pfeffer, 1983). Individuals who are similar in background, and who share common experiences and values, may find interaction easier, positively reinforcing, and more desirable. Thus communication and interdependence are improved in groups that are less diverse. Various studies have explicitly cited the similarity-attraction paradigm as the underlying perspective of their research (e.g. Tsui & O'Reilly, 1989; Barsade et al., 2000; Harrison et al., 2002). Another related perspective is the Attraction-Selection-Attrition paradigm (ASA; Schneider, 1987), which suggests that the similarity/attraction process leads to individuals joining a group or organisation, and can subsequently lead to individuals leaving the group or organisation when the levels of similarity or attraction are not so high – thus the natural tendency is for groups and organisations to become more homogeneous over time. Although the process of selection (and, to a lesser extent, attrition) is clearly not appropriate to consider in all work groups, some studies – in particular those of top management teams – have used this paradigm as part of their theoretical framework (e.g. Jackson et al., 1991, Harrison et al., 2002).

In contrast, the information/decision making perspective suggests that variance in group composition can have a direct positive impact on outcomes. This is derived from the higher level of skills, abilities, sources of information and knowledge that diversity brings (Tziner & Eden, 1985), arguing that often the benefits of increased knowledge and improved decision making can create a process gain that would overcome any decrease in coordination or integration due to dissimilarity between individuals (Phillips et al., 2004). As a result, this perspective has more often been applied to diversity of functional background, education or information (e.g. Ancona & Caldwell, 1992; Pelled, Eisenhardt & Xin, 1999; Shin & Zhou, 2007; Somech & Drach-Zahavy,, 2007), although some authors have argued that this applies also to demographic diversity, as demographically diverse individuals may be expected to have a broader range of knowledge and experience than homogeneous individuals (e.g. Kent & McGrath, 1969; Bantel & Jackson, 1989). Moreover, functional diversity is itself often considered in two different ways: dominant-functional diversity and intrapersonal-functional diversity; the former referring to the extent of diversity in the area that individuals have spent the majority of their careers, and the latter the extent to which individuals have a broad as opposed to narrow focus as functional specialists (Bunderson & Sutcliffe, 2002). Although dominant-functional diversity could be viewed through the lens of either the social categorisation or information/decision-making perspectives, intrapersonal-functional diversity would not make sense from the social categorisation standpoint, since breadth of individuals' expertise would contribute to the extent of information available within a group, and may help to cross-cut some divisions that would be created by focussing solely on members' principal functions.

Along with these differing theoretical perspectives, there have been attempts to categorise diversity into distinct types. One of the most cited of such categorisations was by Jackson, May and Whitney (1995). Jackson and her colleagues suggested there were two main dimensions of diversity: task-relatedness (task versus relations) and observability (surface versus deep). Task-relatedness draws a distinction between features that are highly relevant to the task (e.g., education, functional background, knowledge, skills & abilities) and features that are generally not relevant to the team task (such as age, sex, race). Observability draws a distinction between those characteristics that are readily observed or available (such as sex, ethnic background, functional background, education) and those that are not generally observable (e.g. social status, personality, experience). This distinction has been influential in many subsequent studies, most particularly those by Harrison and his colleagues (Harrison et al. 1998, 2002), and does not align itself with any one theoretical perspective.

Thus there is considerable confusion theoretically regarding the impact of work group diversity. In much of the literature, self-categorisation/social identity theory and the similarity-attraction paradigm have been used to explain the negative effects of non-task related diversity on outcomes, whereas the information/decision-making framework has been used to explain the positive results of task-related diversity on outcomes. However, there is also a significant crossover of theory. For example, Jehn, Northcraft and Neale (1999) posited (and found) that informational diversity, although positively linked with performance (under certain conditions at least) would be associated with increased task conflict: a variable that might normally suggest poorer outcomes for the group. Nevertheless, when task complexity was high, they found that increased informational diversity was associated with better group performance. This is one example of how the theoretical distinction between the two perspectives has become blurred.

Moreover, not all variables fall neatly into a task-related/non-task related dichotomy. Consider for example the case of organisational tenure: this is a task-related variable, but is often highly correlated with age, which is not; as a result, the impact of tenure diversity on outcomes is unlikely to be consistently positive or negative according to theory. Certainly meta-analyses linking work group diversity and performance (e.g. Bowers, Pharmer & Salas, 2000; Webber & Donahue, 2001; Stewart, 2006; Horwitz & Horwitz, 2007; Hülsheger, Anderson & Salgado, 2009; Joshi & Roh, 2009: these are discussed in more detail in the next section) have generally concluded that there is no overall pattern of results for diversity as a whole, and for specific types of diversity any overall relationships are, at best, very small. This suggests that the adoption of a specific theoretical position to explain the effects of work group diversity is unhelpful.

As an attempt to reconcile these competing models, Van Knippenberg, De Dreu & Homan (2004) have attempted to integrate the two perspectives into a single overarching model, the Categorisation-Elaboration model (CEM) which allows for both the positive and negative effects suggested by the different theoretical standpoints. The CEM, summarised in Figure 2.1, proposes that all types of diversity may (in principle) elicit social categorisation processes as well as information/decision-making processes, because all attributes provide a basis for possible differentiation as well as being associated with differences in task-relevant information and perspectives. The potential of a particular attribute to increase group information and inform decision-making may vary, but even demographic factors such as age, gender and race may have some contribution – if not directly, then certainly as proxies for underlying variables such as culture. The CEM also posits that certain types of diversity will be more or less likely to lead to social categorisation processes, due to the cognitive accessibility and comparative and normative fit of categorisation – again building on social identity theory. These social categorisation processes in turn lead to affective and evaluative reactions (such as relational conflict, cohesion, identification and commitment), particularly when the identity elicited by the categorisation is threatened or challenged. It is these reactions that moderate the main relationship between diversity and elaboration of task-relevant information and perspectives, which in turn leads to group performance. As also suggested by other authors (e.g. Jehn et al., 1999; Webber & Donahue, 2001), this main effect is also moderated by task features: informational & decision requirements, task motivation and task ability being three key examples. Thus the often-observed negative relationships between diversity and performance would appear to be due to social categorisation processes leading to poorer intragroup processes, which – particularly when the task at hand is relatively simple and low in required ability – creates poorer task-relevant information elaboration and, subsequently, poorer outcomes. Although this provides a general framework for understanding the mixed effects of diversity, usage of the model in the wider diversity literature has so far remained limited, and it is still unclear in exactly what situations the effects of social categorisation may moderate the diversity-outcomes relationships.

Aston University

Information on this page has been removed for data protection purposes

Figure 2.1: The Categorization-Elaboration Model (van Knippenberg et al., 2004)

2.2 Research findings: themes and inconsistencies

The history of diversity research is perhaps characterised best by its very diversity – few firm conclusions can be drawn about the effects of diversity in organisations, due to the sheer inconsistency of results. As Williams and O'Reilly (1998) described in their review of over 80 articles across 40 years of diversity research, there were few findings that could be described as consistent. Milliken and Martins' (1996) earlier review drew very similar conclusions, leading them to describe diversity as a "double-edged sword: increasing the opportunity for creativity as well as the likelihood that group members will be dissatisfied and fail to identify with the group" (p. 403). Indeed both reviews found that some diversity constructs could have either positive or negative effects on different outcomes; some could even have positive or negative characteristics on the same outcomes. To try to draw meaning from the sheer mass of varying results, therefore, it is probably wise to start by considering the findings of meta-analyses.

Bowers, Pharmer and Salas (2000) conducted the first meta-analysis on the effects of work group diversity on performance. This was a comparatively small analysis, using only 13 separate studies (albeit 57 effect sizes within these studies), and looked specifically at diversity in relation to gender, ability and personality. They found that there were no significant overall relationships between diversity and performance, although the direction of the relationships appeared to favour heterogeneity over homogeneity. They also found that there was a significant, positive relationship between the effect size and task difficulty (as coded by the authors), suggesting that diversity may only become beneficial when the task demands it. For low task difficulty, or for "performance tasks" (those with low cognitive demand, e.g. many physical work tasks), there was a slight negative effect of diversity on performance.

Webber and Donahue's (2001) meta-analysis was slightly more comprehensive, including 24 studies (although only 45 separate effect sizes) and examining more types of diversity. As well as effects on performance, they looked at group cohesion as an additional outcome, and they included studies of diversity on age, gender, race/ethnicity, educational and functional background, and industry and occupational background. The first three of these were categorised as "less job-related"; the remainder as highly job-related. They found that neither type of diversity was associated with either outcome, broadly supporting the findings by Bowers et al. (2000): the overall conclusion from both meta-analyses being that there are no consistent relationships between group diversity and either group performance or cohesion.

Stewart (2006) studied many team design features, and analysed diversity as just one of these, not distinguishing between types of diversity. His meta-analysis included 26 effect sizes involving group diversity, which suggested a significant but very small negative relationship between diversity and group performance (r = -0.04). When broken down by type of team, there were also small, significant and negative relationships within production teams (r = -0.07) and management teams (r = -0.03), but a small positive relationship within project teams (r = 0.04), possibly lending support to the arguments presented by Bowers et al. (2000) about task difficulty and type being moderators.

Horwitz and Horwitz (2007) were able to draw upon a greater amount of research, including 78 effect sizes from 35 separate correlational studies, even though they eschewed top management team studies to ensure a greater consistency in their sample. Categorising diversity type in the same way as Webber and Donahue (2001), they found a significant positive overall relationship between task-related diversity and both quality and quantity of performance (again, small overall effects: r = 0.13 and 0.07 respectively); however, they found no similar relationship for bio-demographic (i.e. less task-related) diversity. Overall, diversity was found to have a small but statistically significant negative relationship with social integration, but this result did not hold for either type of diversity in isolation.

A broader meta-analysis looking at predictors of team innovation by Hülsheger, Anderson and Salgado (2009) used job-relevant and background diversity as two of its principal predictors. Using a smaller number of studies (15 for job-relevant diversity and 8 for background diversity) they found a significant (although still small) positive relationship between job-relevant diversity and innovation (r = 0.16), and no significant relationship for background diversity.

Joshi and Roh (2009) concentrated on the context of teams as a moderator of the diversity-performance relationship. They meta-analysed 39 studies and found similar results to earlier meta-analyses: no significant overall relationships between diversity and performance, but very small significant relationships for different types of diversity: positive for task-related diversity ($r = 0.04$) and negative for relations-oriented diversity ($r = -0.03$). They went on to find, though, a number of moderating effects. The occupational context of an attribute could be important: gender diversity had a significant, negative effect in majority male occupational settings, but a positive effect in gender-balanced settings. Likewise, racial diversity had a negative effect in majority white occupations, but a positive effect in more balanced occupations. Industrial settings were also important: relations-oriented diversity had a significant positive association with performance in service industries, but a negative effect in manufacturing and high-technology industries. Task-oriented diversity was positively related to performance in high-technology settings but not significantly related in any other setting. Team interdependence was also a significant moderator: relations-oriented diversity was positively linked to performance when interdependence was low, but negatively when interdependence was high (contrary to the authors' predictions). Relations-oriented diversity was also more positively associated with performance when teams were short-term in nature, as opposed to long-term or stable teams.

The pattern of results from these meta-analyses appears relatively consistent: very small overall main effects, often not significant, but sometimes positive for task-related diversity and negative for demographic diversity. More recently, however, van Dijk, van Engen and van Knippenberg (under review) conducted a considerably more comprehensive meta-analysis, including 146 studies featuring the relationship between work-group diversity and group performance. Unlike the previous studies which had distinguished between job-related and demographic diversity as a system for predicting whether diversity would have a positive or a negative effect on performance, van Dijk and his colleagues also considered the nature of the performance variables – particularly whether they were objective or subjective. They found that the apparent distinction between job-related and demographic diversity was based on biases in subjective ratings, and there was no evidence for this distinction in the 59 studies featuring objective measurement. Within the subjective performance ratings, the distinction between job-related and demographic diversity only became apparent when the rater was external to the team (n = 63; r = -0.06 for demographic diversity and 0.09 for job-related diversity), and not when the rater was internal (n = 41; r = 0.05 for demographic diversity and 0.03 for job-related diversity). This suggests that earlier differences between types of diversity may be artefacts caused (partially) by the source of performance data – thus rather than clarifying the situation, the results serve to muddy the waters even further in terms of when diversity has a positive or negative influence on outcomes. The authors also found, however, that job-related diversity is more positively related to performance when task complexity is high, suggesting that the role of diversity is contingent on the role of the work group, and that it is more strongly related to innovative performance than in-role performance (supporting the information/decision-making perspective).

Given the lack of strong directional effects in these meta-analyses, and the tendency to group different types of diversity together, it is worth examining the findings of individual studies of work group in more detail, paying particular attention to both the specific diversity attributes being studied. Until the 1990s, most diversity research focused on demographic and surface-level work characteristics (Williams & O’Reilly, 1998). Age, sex, race/ethnicity, education, tenure and functional diversity were all the subject of dozens of articles that examined a variety of outcome measures. Even today these attributes dominate the work group diversity research agenda, albeit now complemented by more research on deep-level attributes such as attitudes, personality and beliefs.

Age diversity has been found to have both positive and negative effects on group conflict (O'Reilly, Williams & Barsade, 1997), and negative effects on innovation (Zajac, Golden & Shortell, 1991), retention (O'Reilly, Caldwell & Barnett, 1989; Jackson et al., 1991; Tsui, Egan & O'Reilly, 1992), attendance (Cummings, Zhou & Oldham, 1993), performance (Judge & Ferris, 1993; Ely, 2004), role clarity and communication (Zenger & Lawrence, 1989). Sex (or gender) diversity seemed to have mainly negative effects, leading to poorer group processes (e.g. Kramer, 1991; Pelled, 1996), higher conflict (Alagna, Reddy & Collins, 1982; Randel, 2002), identification with the work group (Van Knippenberg, Haslam & Platow, 2007), creativity (Kent & McGrath, 1969; Pearsall, Ellis & Evans, 2008) and higher absenteeism and turnover (Tsui et al., 1992). In fact, sex diversity has perhaps the most consistent results of all the constructs: however, the vast majority of studies have pointed towards negative outcomes. One exception is Curşeu, Schruijer and Boroş (2007) who found that gender diversity was positively linked to group cognitive complexity in a student sample.

Racial/ethnic diversity has had mixed effects, and Williams and O'Reilly (1998) pointed out that the results may be influenced by the year of study, as attitudes towards race have changed over the past 50 years. Race diversity has been found to have positive relationships with creativity, creation and implementation of ideas (McLeod & Lobel, 1992; O'Reilly, Williams & Barsade, 1997), information sharing (Sommers, 2006) and cooperation (Earley, 1989), but negative effects on commitment, individual performance (for both minority members - Greenhaus, Parasuraman & Wormley, 1990, and overall group performance - Watson, Johnson & Zgrourides, 2002), and group processes (Linville & Jones, 1980; Pelled et al., 1997). Some studies have also found curvilinear effects, including Dahlin, Weingart and Hinds (2005) who found that the depth and integration of information use was highest with moderate levels of national diversity, although range of information use was lowest with moderate levels of national diversity, and Earley and Mosakowski (2000), who found that moderate levels of national diversity were associated with poorer performance.

Functional background diversity has also had very mixed results, with positive effects on performance (Pelled et al., 1997; Van der Vegt & Bunderson, 2005), creativity (Choi, 2007; Shin & Zhou, 2007), innovation (Drach-Zahavy & Somech, 2001; Fay et al., 2006; Somech, 2006; Somech & Drach-Zahavy, 2007), external and internal communication (Ancona & Caldwell, 1992; Glick, Miller & Huber, 1993), organisational citizenship behaviour (Pelled, Cummings & Kizilos, 2000), and firm growth (for top management teams: Eisenhardt and Schoonhoven, 1990; Wiersema & Bantel, 1992; Hambrick, Cho & Chen, 1996), as well as various team processes (Somech, 2006; Somech & Drach-Zahavy, 2007, Zellmer-Bruhn et al., 2008). It has also been shown to have negative effects on performance (Ancona & Caldwell, 1992; Jehn et al., 1999; Bunderson & Sutcliffe, 2002), and in other samples no effects on performance or communication (e.g. Smith et al., 1994); it has also been shown to be related to team conflict (Jehn et al., 1997; Zellmer-Bruhn et al., 2008) and lower team social capital (Reagans, Zuckerman & McEvily, 2004).

Educational background diversity – either in terms of level of education or major specialism – has also been a regular topic of interest; this has been positively linked to performance (Triandis, Hall & Ewen, 1965; Thornburg, 1991; Smith et al., 1994; Hambrick et al., 1996; Kearney et al., 2009), job search efficacy (Choi, Price &

Vinokur, 2003), range and depth of information use (Dahlin et al., 2005), perceived team viability (Foo, Sin & Yiong, 2006) and diversification strategies (Wiersema & Bantel, 1993), but also to staff turnover (Alexander et al., 1995) and to poorer integration (Kirchmeyer, 1995), whilst being negatively related to information integration (Dahlin et al., 2005).

Similarly, tenure diversity has been shown to lead to poorer communication and group processes (in most, but not all studies – Ancona & Caldwell, 1992, find a positive relationship with group functioning), more group conflict (O'Reilly et al., 1997), lower team social capital (Reagans et al., 2004), higher turnover (O'Reilly et al., 1989) and adaptability to change (O'Reilly et al., 1993); it has also been shown to have positive relationships with innovation (Flatt, 1996), firm international diversification (for top management teams – Tihanyi et al., 2000) and both positive and negative relationships with performance (e.g. Murray, 1989; Eisenhardt & Schoonhoven, 1990; Ancona & Caldwell, 1992; O'Reilly et al., 1997; Ely, 2004).

In recent years, more researchers have moved away from simply examining diversity of demographic characteristics, and have increasingly been looking at diversity of knowledge, skills, abilities, values, beliefs, attitudes and personality. This is not entirely separate from the diversity of demographic characteristics: for example, some authors have seen diversity of functional background as a proxy for diversity of knowledge, skills and abilities (KSAs), and other diversity in demographic characteristics can also be seen as proxies for deeper, underlying differences (e.g. Priem, Lyon & Dess, 1999; Olson, Parayitam & Bao, 2007; Gevers & Peeters, 2009); in addition, much of the research is tied together by the same theoretical perspectives (which will be discussed in the next section).

However, there have been few clear results with these types of diversity either. For example, several studies have demonstrated a link between diversity in group member personality and various outcome measures, e.g. diversity in extraversion associated with greater social cohesion (Barrick et al., 1998) and better performance (Neuman, Wagner & Christiansen, 1999); diversity in conscientiousness associated with lower team member satisfaction and poorer performance (Gevers & Peeters, 2009), lower social integration (Harrison et al., 2002) and lower task cohesion (Van Vianen & De Dreu, 2001); and diversity in emotional stability associated with better team performance (Neuman et al., 1999) but lower task cohesion (Van Vianen & De Dreu, 2001).

Similarly, Harrison and his colleagues (Harrison, Price & Bell, 1998; Harrison et al., 2002) examined the relationship between diversity in attitudes and social integration, with negative effects (diversity in job satisfaction, and diversity in attitudes towards the importance of outcomes), positive effects (functional background diversity) and no effects (diversity in values, diversity about the task) all found within these two studies.

Another area of research that has become increasingly prominent in recent years is that of socially shared cognition and affect. Although this is not usually considered a type of diversity research, the sharedness aspect of it undoubtedly lends itself to the diversity field. One example of this is the idea of climate strength (Gonzalez-Roma et al., 2002; Schneider et al., 2002), in which the extent of group members' shared perceptions of the group climate is considered as a variable of interest. Other examples in the literature include shared mental models linked with fewer errors in flight crews (Weick & Roberts, 1993); shared task representations linked to group performance in a variety of settings (Tindale et al., 1996), cognitive disparity being linked to increased task conflict (Olson et al., 2007) and lower group cognitive complexity (Curşeu et al., 2007) and shared beliefs about failure being linked to improved group performance (Cannon & Edmondson, 2001).

One area of diversity research that has taken off since 1998, partly as a result of the Williams & O'Reilly (1998) review, is the inclusion of possible moderators in diversity research. Prior to this, most of the studies presented in the literature examined main effects of diversity only (Van Knippenberg & Schippers, 2007). The absence of moderating effects was suggested by Williams and O'Reilly as a possible cause of the inconsistency of results found. For example, Van der Vegt, Van de Vliert and Oosterhof (2003) found that the relationship between informational dissimilarity and both team identification and organisational citizenship behaviour was moderated by task and goal interdependence; a separate study by Van der Vegt and Janssen (2003) found interactions between task and goal interdependence and both cognitive and demographic diversity in their relationships with innovative behaviour. Shin and Zhou (2007) found that the effects of educational specialisation diversity on creativity were positive when transformational leadership was high, but negative when transformational leadership was low. One key paper is that by Harrison et al. (1998), who found that surface-level diversity (diversity of constructs such as age, sex, race, and functional background, that can be observed easily within a group) had negative effects on social integration, but that these effects decreased over time; conversely, the effects of deep-level diversity (diversity of constructs such as values, attitudes and beliefs, that cannot easily be observed) increase over time. Thus team tenure and type of diversity moderate the relationships between diversity and social integration.

Other variables that have been found to be moderators of relationships between diversity and outcomes include complexity of task (Jehn, 1999; Pelled, Eisenhardt &

Xin, 1999), team interdependence (Timmerman, 2000; Schippers et al., 2003), team processes (Ely, 2004; Mohammed & Angel, 2004; Fay et al., 2006), leadership style (Somech, 2006; Hmieliski & Ensley, 2007; Shin & Zhou, 2007; Kearney & Gebert, 2009), organisational culture and HR practices (Jehn & Bezrukova, 2004), trust (Olson et al., 2007), interpersonal congruence (Polzer, Milton & Swann, 2002), distribution of information (Kooij-de Bode, van Knippenberg & van Ginkel, 2008), zip code (Sacco & Schmitt, 2005), diversity beliefs (van Knippenberg, Haslam & Platow, 2007), and diversity of other attributes (Jehn, 1999; Jackson & Joshi, 2004). However, the literature in recent years is also replete with examples of predicted moderators for which empirical support was not found (e.g. subgroup status - Jackson et al., 1991; task interdependence - Jehn, 1999; strength of competition - Murray, 1989; diversity in sales districts – Jackson & Joshi, 2004; environmental uncertainty – Canella, Park & Lee, 2008). These are a few of the examples in published studies, and it is likely (given the publication bias towards studies with positive findings) that many more researchers have failed to find predicted moderators of work group diversity. There are several possible reasons for these non-significant findings: as with any moderator effects, sample size and range restriction may have a substantial effect on test power (sample size because, with most research conducted at the team or organisational level, it is more difficult to achieve large samples; range restriction because many studies are concerned with a particular sector and/or geographical region, leading to relatively homogeneous samples). However, it also remains possible that variables are hypothesised as moderators incorrectly.

2.3 Chapter summary

The literature on work group diversity is littered with inconsistencies, and very few relationships have been found reliably; there are even several examples of opposite effects being found between the same variables. Those effects that do appear to be more consistent include both positive effects of diversity (e.g. functional background diversity and innovation), and negative effects of diversity (e.g. sex diversity and most outcomes). Although different theoretical perspectives are well-developed to explain the differing results individually, relatively little research has focused on the inconsistencies themselves. Various authors have suggested potential moderators to explain some of this variation; one potential explanatory variable of the inconsistencies that has only recently been addressed in detail (Harrison & Sin, 2005; Harrison & Klein, 2007; Roberson, Sturman & Simons, 2007) is the contrasting range of definitions and operationalisations of diversity itself.

It is this array of types of diversity and associated measurements that I will seek to clarify in this study. To enable this, in the next chapter I will examine the one significant attempt to date to create a typology of work group diversity (Harrison & Klein, 2007), identifying its strengths and weaknesses and how this can be expanded into a fuller typology with clearer recommendations for researchers.

CHAPTER 3

TYPES OF DIVERSITY: HARRISON AND KLEIN'S TYPOLOGY

3.0 Chapter introduction

In this chapter I provide a critique of Harrison and Klein's (2007) typology of diversity, in which they defined three types: separation, variety and disparity. I appraise the strengths and weaknesses of their work, in particular highlighting areas where it does not completely fit the needs of diversity researchers, and suggest four ways in which the typology could be improved.

3.1 Harrison and Klein's definitions

Despite a plethora of research on work group diversity, and a variety of attempts at defining it (e.g. Jehn et al., 1999; Pelled et al., 1999; Polzer, Milton & Swann, 2002; Jackson et al., 2003; Harrison & Sin, 2005), it was not until Harrison and Klein's (2007) article that any authors explicitly acknowledged that not all diversity researchers work on the basis of the same principles, and actually proposed definitions for different constructs. Harrison and Klein reviewed the various theoretical approaches of researchers and derived definitions for three distinct forms of diversity: separation, variety and disparity.

Separation refers to the extent of differences of opinion amongst unit members. It is conceived as a measure of "disagreement or separation - horizontal distance along a single continuum". Examples of constructs that fall into this category are value diversity (Williams & O'Reilly, 1998), and climate strength (e.g. Gonzalez-Romá et al., 2002; Schneider et al., 2002). Maximum separation would occur when unit members are equally split at opposite endpoints of the continuum; for example, in a team of six, three members "strongly agreeing" and three members "strongly disagreeing" that their team functions effectively. Minimum (zero) separation occurs when all members of the group had the same value of the attribute in question.

Separation is particularly useful when considering theoretical perspectives such as social categorisation (Tajfel & Turner, 1979; Hogg & Terry, 2000), similarity-attraction (Newcomb, 1961; Byrne, 1971; Clore & Byrne, 1974), and attraction-selection-attrition (Schneider, 1987; Schneider et al., 1995). This is because these theories all suggest reduced separation has positive effects on outcomes such as co-operation, trust, social integration, cohesion, lack of conflict, and performance.

Variety refers to differences across the unit on a qualitative (categorical) variable, such as functional background. It encapsulates the extent to which members are evenly spread across different possible values of this variable. Minimum variety is achieved when all members of the unit have the same value; maximum variety when members are equally spread across all possible values.

Harrison and Klein (2007) discuss variety as a positive attribute for unit outcomes such as effectiveness of decision making and innovation. This is based on the information and decision-making theoretical perspective of diversity (e.g. Ancona & Caldwell, 1992; Jackson, 1992; Tsui, Egan, & O'Reilly, 1992; Jehn, Northcraft & Neale, 1999), which posits that a greater level of diversity (or variety) among a unit is reflective of a greater amount of information held within the unit, allowing better decisions to be made.

Disparity refers to the concept of inequality, often seen in the sociological literature (e.g. Blau, 1977; Allison, 1978). It measures the extent to which members of a unit differ on some asymmetrical variable, where "more is always better". A typical example of such a variable is pay - maximum disparity occurs when one member of the unit has maximum pay, but all other members have minimum pay. Minimum disparity would occur when all members have the same level of pay. As well as pay dispersion, disparity is often seen in reference to attributes such as power, status or social capital, with theoretical perspectives such as the IUS conjecture, which posits a curvilinear relationship between rank and conformity (Phillips & Zuckerman, 2001).

3.2 Critique of Harrison and Klein

Harrison and Klein's typology is a welcome first step in clarifying the different meanings of diversity for different research perspectives. They clearly elucidate the rationale for considering diversity not as a single construct, but as a suite of related constructs with similar, although subtly different, definitions. Nevertheless, as is often the case with initial typologies, it is not comprehensive and leaves some issues unclear for researchers. There are four issues in particular that I argue they have failed to address adequately.

First, and most striking, is the alignment of separation and variety with explicit theoretical perspectives and types of measure. Specifically, separation (which is aligned with the social categorisation perspective) is only defined in terms of interval variables, and variety (aligned with the information and decision making perspective) is only defined in terms of categorical variables.

However, social categorisation by group members would not always be conducted in terms of a variable which is measured on such a continuum. In particular, Harrison and Klein (2007) do not mention one of the key categories of variable in the diversity literature - diversity of ethnicity, nationality, or culture. These variables are typically measured using a categorical indicator; under Harrison and Klein's (2007) typology, this would make their diversity most suitable as a measure of variety, and useful for the information and decision making theoretical perspective. Whilst it is certainly possible that diversity of culture, nationality and ethnicity may lead to an increased amount of information in a unit, leading to better decisions as a result, most diversity researchers studying these variables have treated them as a basis for social categorisation, which does not fit in with Harrison and Klein's typology. The same can be said for some other demographic variables, notably gender.

Conversely, some researchers have argued that variables such as value diversity (Jehn & Mannix, 2001) can lead to task conflict, which in turn engenders more careful consideration of the task in hand and thus leads to better group decisions. This clearly fits within the information and decision making perspective of diversity, yet the variable on which diversity is to be measured is more likely to be an interval variable than a categorical variable. Harrison and Klein's (2007) typology is insufficient to cover this scenario.

The one area where the Harrison and Klein (2007) typology does provide a consistent conceptual distinction between measurement types is that of disparity. Disparity is conceptually distinct from separation or variety, due to the inherent asymmetry of the underlying variable. Although it is conceivable that researchers may wish to measure the disparity of a non-ordinal (nominal) categorical variable, this would point to an error in the initial measurement decisions, as the asymmetry necessary for a measure of disparity implies some degree of relative magnitude in the underlying categories.

The second issue, related to the first, is that the separation of diversity research into the two distinct theoretical perspectives (social categorisation and information/decision making) is not consistent with models such as Van Knippenberg et al.'s (2004) Categorisation-Elaboration Model, which (as reviewed earlier) provides a theoretical basis for the combination of the two perspectives. Broadly speaking, diversity is expected to lead to elaboration of task-relevant information perspectives (and then subsequently to performance) as in the information and decision making perspective. However, it is also acknowledged that under certain conditions (cognitive accessibility, or normative fit, or categorisation) diversity also leads to social categorisation, which may lead (where identity implied by the categorisation is threatened) to affective or evaluative reactions, such as conflict, cohesion, identification and commitment, and that these may in turn moderate the relationship between group diversity and elaboration. The important consideration for Harrison and Klein's typology, though, is that it is the *same* diversity that is expected to lead to both outcomes - the social categorisation, and the elaboration of ideas due to increased information. Therefore the Harrison and Klein typology does not fit with this model.

Third, the three constructs provided are not comprehensive. This is most striking when considering their construct of variety: the definition, the "extent to which members are evenly spread across different possible values of this variable", could mean multiple things. It is not immediately clear whether the spread of values needs to be fairly even, or whether simply having any members with particular values is sufficient: for example, would a team comprising two engineers, two accountants and two people from a sales background be considered more varied than a team comprising four engineers, one accountant and one person with a sales background? To extend this further, if the latter team had three engineers, an accountant, a salesperson and a lawyer, it would have more categories represented than the former group, and yet according to Blau's (1977) index (one of the indices Harrison and Klein recommend for measuring variety), it is less diverse than the first group, with a value of 0.64 compared with 0.67 (higher values representing greater diversity). Thus the definition – or, at least, the operationalisation – of variety appears not to capture the total amount of information available to a group. Moreover, other definitions of "diversity" – beyond separation, variety and disparity – may well be desirable.

Fourth, and exemplified by the previous point, Harrison and Klein go on to suggest how researchers should measure each of the different types of diversity. Such a classification is certainly welcome, but the detail is not necessarily consistent with the earlier definitions. To demonstrate this, consider the indices proposed for each type of diversity: for each they propose two possible indices (described and discussed below), all of which had been used in prior research on work group diversity. This immediately presents two potential problems. First, as the two operationalisations for each diversity type are different from each other, they must therefore represent slightly differing versions of diversity. Can both be consistent with the original definition? Second, are the indices used in prior research adequate for representing these "new" definitions of diversity?

In the case of separation, the two indices presented are the standard deviation and mean Euclidean distance. These differ fundamentally because the standard deviation is based around differences from a group mean, whereas mean Euclidean distance is based on differences between pairs of individuals. The latter would appear to be more consistent with the definition of separation ("the extent of differences or opinion amongst unit members"). In practice they might be highly correlated, but considering the specific nature of the definition of separation, clearer guidance about which is more appropriate would be helpful.

In the case of variety, both Blau's (1977) index and Teachman's (1980) entropy index are presented, without much to suggest what the differences between the two are. Later I will explore these differences in some detail. For disparity, the coefficient of variation (CV) and Gini index are offered by Harrison and Klein (2007) as alternatives, but again without a clear rationale to choose one or the other. I will also explore the difference between these indices more fully later.

In summary, although the distinctions provided by Harrison and Klein are helpful and may assist in clearer theoretical thinking about how diversity is measured, they could also lead to confusion and inappropriate choices. Indeed, there are already examples in the literature of poor decisions about diversity indices: Kearney and Gebert (2009), and Kearney, Gebert and Voelpel (2009) are two papers that use Blau's index to measure age diversity. This was chosen because of Harrison and Klein's suggestion that Blau's index should be used to measure variety (which represented the theoretical perspective Kearney and his colleagues were taking in both papers); however, they also recommended variety should be used with nominal (non-ordered categorical) variables, and Blau's index is only appropriate for such variables. Age is clearly a continuous variable (or at worst ordinal if measured in categories), and therefore unless a nominal categorisation were justified would not fit with this measurement method, even though the definition of variety was appropriate for the authors' theory. This is a clear demonstration of a gap in Harrison and Klein's typology and how its use may result in adoption of incorrect indices.

3.3 Proposed extension of the typology

To address the issues raised in section 3.2, I propose an extension to the Harrison and Klein typology. First I will refine the definitions of the terms *separation*, *variety*, and *disparity*, so that they are not inextricably linked with a particular type of variable, and divide *variety* into two different constructs to represent the alternative perspectives described above. Later, following a review of diversity measures in the literature, I will go on to discuss measurement of each of these types of diversity for different types of variable.

I define *separation* as "the extent to which unit members are different from one another with respect to a variable *X*". This is different from the Harrison and Klein definition in that it does not require *X* to be measured along a single continuum, but allows categorical variables also. This is important because the concept of separation is particularly useful for the social categorisation theoretical perspective, and social categorisation may occur on the basis of categorical variables such as nationality. This definition of separation is of a group-level construct, but one which relies on the differences between individual pairs of unit members. This reflects the idea that social categorisation involves comparing oneself with all other members of a unit, rather than with the unit itself. This is an important consideration for measurement.

I define *variety* as "the extent to which members of a unit have different levels of a variable *X*". This differs from Harrison and Klein's definition in that it does not require *X* to be a purely categorical variable, but also allows use of interval (and ratio) variables for this purpose. This means that variety can be considered for variables such as values, attitudes, beliefs and perceptions, which are more likely to be measured on a scale, but may be useful for measuring diversity to capture the range of views or abilities for the information/decision-making perspective.

However, as mentioned earlier, this definition is still ambiguous, and represents a super-ordinate concept of diversity as the total extent of an attribute in a group. Within this group-level construct what is most important is the extent to which different possible values of the variable *X* are covered by the group. However, what does "coverage" mean here: does this refer to the total range of categories of *X* included in the group, or the extent to which any member of the group has a different level of *X* from other members?

Consider the example proposed earlier, in which two groups of six people are spread across different values of a categorical variable. The situations discussed previously are represented by groups 1 and 2 in Table 3.1. Group 2 has more different categories (levels) represented than group 1, so in one sense has a greater degree of variety. However, the probability that any two members selected at random have different levels of variable *X* is the same in both groups (0.8) – and therefore the probability (all else being equal) of coming into contact with a different viewpoint is the same, despite more categories being represented in one of the groups. I refer to the total number of levels represented as *range*, and the extent to which they are evenly spread out as *spread*.

Table 3.1: Examples of contrasting range and spread in groups

Group	Group members' values of *X*	Range	Spread
1	A, A, B, B, C, C	Lower	Similar
2	A, A, A, B, C, D	Higher	Similar
Group	Group members' values of *Y*	Range	Spread
3	1, 1, 4, 7, 7	Higher	Lower
4	2, 3, 4, 5, 6	Lower	Higher

Likewise, as I have now included continuous variables in the definition of variety, consider two groups of five people, each of whom responded to a 7-point questionnaire item on a variable *Y*. The first group may have had values (1, 1, 4, 7, 7); the second group had values (2, 3, 4, 5, 6); these are represented by groups 3 and 4 respectively in Table 3.1. In some senses, the first group has greater coverage, as the extremes of the scale at both ends are represented within the group (so this has the greatest possible range). In another sense, however, the second group has more coverage, because a larger number of distinct responses are included. So although the first group has more extreme values, in some senses there is more variety in the second. This is similar to the distinction between "species abundance" and "evenness" made by Pielou (1976) when describing the variation in species from an ecological perspective. As a diversity index, though, spread encompasses both range and evenness: a distribution that was perfectly even over a very small range could not be described as being diverse.

Whether the range or spread is more important will depend on two factors. First, the construct being measured. Is it predicted (theoretically) that more extreme values within a group will be associated (either positively or negatively) with outcomes, or is it having a range of different values that is more important? Second, the measurement scale being used. Can different values of the variable be reliably interpreted as representing different values within a group? This is a question for individual researchers to answer based on the theory being tested, just as the distinction between separation, variety and disparity is a decision to be made on a theoretical basis.

My definition of *disparity* is almost identical to that of Harrison and Klein, but rephrased such that it is consistent with my definitions of separation and variety. I define disparity as "the extent to which there is inequality between unit members with respect to an asymmetrical variable *X*". An asymmetrical variable is typically one which has an uneven distribution of a principally limited resource, e.g. wealth or power. Like the definition of separation, this implies disparity is a group level construct, but one which depends on the differences between individual pairs of unit members. However, the fact that *X* is necessarily asymmetrical implies that it cannot be a purely nominal variable, but must include at least some ordering of values.

3.4 Chapter summary

In this chapter I have provided a critique of Harrison and Klein's (2007) typology of diversity, highlighting where some gaps and weaknesses exist. I have provided revised definitions of four diversity types – two very similar to those of Harrison and Klein, and two which are different interpretations of the third, variety – which should help to address some of these limitations.

Having defined these constructs, I will move on in Chapter 5 to consider how they should be measured for each type of variable. However, beforehand in Chapter 4, I review the indices currently used in the literature to help determine which of these will be sufficient for measuring the constructs, and where new indices may have to be constructed.

CHAPTER 4

REVIEW OF DIVERSITY MEASURES FOUND IN THE LITERATURE

4.0 Chapter introduction

In the management and applied psychology literatures a wide range of indices have been used to measure group diversity. Some of these have achieved widespread use; others have been used only once or twice; others still have been suggested but have not (yet) appeared in empirical journal articles. In this chapter I describe each of these indices, commenting on their properties and appropriateness of use for measuring diversity. I will first consider measures applied to continuous variables (including both interval and ratio variables), then measures for ordinal variables, and finally, measures for categorical (nominal) variables, including measures specific to binary variables.

Note on terminology

Throughout this chapter, I will assume that we are interested in the diversity of a group of n individuals on variable X, with values $x_1, x_2, \ldots, x_n$. Where X is a categorical variable, I will assume there are k possible categories.

4.1 Continuous measures

4.1.1 Standard deviation

The standard deviation (SD) is one of the most well known and widely used indices of dispersion. It has been used in (at least) dozens of different studies of work group diversity (for example, Choi, 2007; Harrison et al., 2002, Marcel, 2009; Pegels, Song & Yang, 2000, Wegge et al., 2008). However, its use goes far beyond the realm of group diversity, and is one of the most common indices for denoting the amount of variation in any type of sample. It is defined as the positive square root of the variance, and as such has a wide range of useful statistical properties – one of the most used being that the ratio between a sample statistic, e.g. a sample mean, and its standard deviation (known as the standard error), follows a *t*-distribution, which allows significance testing.

The formula for the standard deviation is:

$$SD_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

Thus the standard deviation ranges from a minimum of 0 (when all members of the group have the same value, i.e. $x_1 = x_2 = \ldots = x_n$), to a theoretically unlimited maximum (constrained only by the range of the variable being measured), with the maximum value within a sample being achieved when members of the group fall into two equal-sized subgroups, with minimum and maximum values respectively; if these values are $x_{\min}$ and $x_{\max}$, then the standard deviation would have the value $\frac{1}{2}(x_{\max} - x_{\min})$.

Due to its evident numerical treatment of the data (in particular, the squaring and square rooting of numbers), this is obviously only appropriate where the numbers representing the data are meaningful relative to each other, i.e. continuous data (ratio or interval).

Often the standard deviation is calculated as the estimated population SD, rather than a sample SD: this is achieved by multiplying the sample SD by $\sqrt{n/(n-1)}$. This is an important consideration in group research, as it is often the case that all members of the group are measured; even if not, it is not correct to assume an infinite population, which is what the correction does. If group sizes are equal, then this would make little difference; however, when group sizes are small but different, this can be an important difference. For the purposes of measuring work-group diversity, it would be most sensible to use the sample standard deviation (if all members of a group are present) or a finite population-corrected version (if they are not); however, it is seldom mentioned in the literature whether this is done or not.

The standard deviation effectively compares each member of the group to the mean. As such, this does not necessarily represent the calculation needed for separation, which would imply group members comparing themselves against all other group members. This would be more accurately achieved by other indices which compare all possible pairs of individuals within groups, and which will be discussed later in the chapter.

4.1.2 Coefficient of Variation

The coefficient of variation (CV) is one of the more common indices for measuring work group diversity. It is defined as the standard deviation divided by the mean, i.e.

$$CV_x = \frac{SD_x}{\bar{x}}$$

Its use in the management and psychology literature (e.g. Jackson et al., 1991; Harrison, Price & Bell, 1998; Pelled, Eisenhart & Xin, 1999) can be traced back to a reference in the sociological literature. Allison (1978) described the CV as a straightforward scale-invariant measure of inequality, and those papers which give a rationale for its use tend to cite either Allison (1978), or quote the property that the index is independent of the mean, or cite a subsequent paper that can be traced back to this origin. Very often, however, they neglect the fact that the CV should only be used with ratio data, i.e. scales where the value zero has a meaning of zero (absence of the quality being measured, e.g. salary or age); because of the division by the mean it is entirely inappropriate that it should be used with non-ratio data. For example, a 1-5

Likert scale would produce very different values (and, importantly, relative values) of the CV compared with a 0-10 Likert scale. These limitations were pointed out at length by Bedeian and Mossholder (2000), and again by Harrison and Klein (2007).
Even when data are measured on a ratio scale, this might not be an appropriate property: for example, the CV of three people with ages 20, 25 and 30 would be the same as the CV of three people with ages 40, 50 and 60, even though the standard deviation of the latter group would be twice that of the former group. Whether the equivalence of the CV is a desirable property in this case depends on the theoretical perspective. Harrison and Klein (2007) suggest this is an appropriate measure for diversity considered as disparity, but not otherwise.

Despite these warnings, there are various examples of its inappropriate use in the literature. For example, Hmieleski and Ensley (2007) used the coefficient of variation to capture skill diversity in top management teams, despite the fact that skill was measured on a Likert (interval) scale ranging from 1-5, not a ratio scale. It can be seen that an arbitrary shift in the values used (e.g. changing to 0-4 or 3-7) would cause different values of diversity to be calculated (not related in a linear way); as the values have no intrinsic meaning, this is clear evidence of why such a measurement is inappropriate. Likewise, Jehn and Bezrukova (2004) used the CV to calculate diversity of educational background, when this was measured as an ordinal variable (ranging from 1 = some school to 8 = doctorate degree). This commits two errors. First, the value zero is not meaningful (and hence ratios do not make any sense), and second, the standard deviation itself cannot be considered a robust statistic because the gaps between categories are not necessarily equal.

Like the standard deviation, the coefficient of variation has a minimum value of zero (when all members of a group have the same value), but has no theoretical maximum. Indeed, assuming the value of zero is real and attainable, it has no practical maximum either. If all members have the value zero, then mathematically it would be undefined, but the value of zero would normally be attributed to identify the lack of variation present.

4.1.3 Variance

The variance of a variable is one of its most fundamental statistical properties. It is important because, mathematically, it is the second central moment of a variable (the first being the mean), and therefore it has important mathematical properties that most other measures of dispersion do not. It is defined as the average squared deviation from the mean, i.e.

$$Var_x = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

Like the standard deviation, there is a correction to estimate the variance of an (infinite) population from which the sample is taken; this is done by multiplying the sample variance by $n/(n - 1)$. Many software packages apply this correction automatically, which again has implications in small group research, as it will lead to bias in groups of different sizes (see e.g. Biemann & Kearney, 2009).

Because it is based on a sum of squares, the scale of the variance is not immediately meaningful. This is why its square root, the standard deviation, is often used as an interpretable index of variation instead. Indeed, the variance is seldom used as a group diversity index in its raw form (a rare exception being Van Knippenberg et al., 2007), but is included here partly for completeness, and partly because other indices, such as the standard deviation and r_{wg}, stem directly from it.

4.1.4 Absolute Deviation from the Mean

The AD_M index, which measures the average absolute deviation from the mean, was introduced into the organisational literature by Burke, Finkelstein and Dusig (1999), although it has existed in the statistical field for decades, known as the mean absolute deviation (MAD) or mean deviation (MD) (e.g. Edington, 1914; Fisher, 1920). It is defined as:

$$AD_m = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

The formula bears evident similarities to that of the standard deviation, the difference being that the standard deviation takes a root mean square (RMS) form for averaging differences, whereas AD_M is a straightforward mean of absolute differences. Like SD, it has a minimum value of 0, when all members of the group have the same value, and a maximum only constrained by the limits of the scale, with the maximum value within a sample being achieved when members of the group fall into two equal-sized subgroups, with minimum and maximum values respectively; if these values are x_{min} and x_{max}, then AD_M would have the value $\frac{1}{2}(x_{max} - x_{min})$.

Although Burke et al. (1999) did not introduce it as a work group diversity measure *per se* (but rather as an alternative index of agreement to the r_{wg} statistic), it has been compared directly with the standard deviation by Gorard (2005) and this demonstrates some relative advantages and disadvantages of AD_M.

The principal disadvantages are as follows. It is certainly less frequently used, and therefore less well known relative to the SD. It also lacks several mathematical properties that are brought about by the SD being the square root of the second central moment of a sample, and in addition is computationally more complex to manipulate because of the inclusion of absolute values.

Whilst these disadvantages are clear regarding wider analysis of the index, they do not preclude its use as a measure of group diversity (or, indeed, agreement). Gorard (2005) showed that AD_M has certain advantages, too. Because it squares the differences, SD gives greater prominence to outliers. There are two immediate corollaries of this. One is that AD_M is more immediately interpretable, as it gives a figure that is directly related to the scale underlying the variable in question. The other is that, when data are imperfect (i.e. there is some measurement error), SD will exaggerate the deviation due to this error to a greater extent than AD_M. A further advantage, claims Gorard, is the simplicity of AD_M compared with SD. Certainly, the computation is more intuitive for people who are not familiar with the standard deviation, or other RMS indices. However, it has had little if any use in the measurement of demographic diversity in work groups; it has seem a little more use as a measure of climate strength, however, which although is often thought of as its own concept, is in fact a form of deep-level diversity (see e.g. González-Romá, Peiró & Tordera, 2002; Dawson et al., 2008).

*4.1.5 r_{wg} and r_{wg}**

James, Demaree and Wolf (1984) first introduced the r_{wg} statistic as a measure of within-group rater reliability. Following responses to this by Schmidt and Hunter (1989) and Kozlowski and Hattrup (1992), they abandoned the idea of using it as a reliability statistic, but instead recast the same index as a measure of within-group agreement. It has principally been used in the organisational literature to justify aggregation of a variable to a higher (e.g. work group) level, with high values demonstrating an agreed perception of a higher-level construct. It is defined as:

$$r_{wg} = 1 - \frac{{s_x}^2}{{\sigma_{EU}}^2}$$

where s_x^2 is the sample variance and σ_{EU}^2 is the variance of a hypothetical distribution in which values are caused by random error only – this is often taken as a uniform distribution across the possible values of the variable. Thus, it is defined as the proportional reduction in error variance due to similarity in ratings. James et al. (1984) also defined a version for multiple-item scales, in which the agreement on *J* parallel items is taken into account; the formula for this is:

$$r_{wg(J)} = \frac{J\left[1 - \left(\bar{s}_{xj}{}^2/\sigma_{EU}{}^2\right)\right]}{J\left[1 - \left(\bar{s}_{xj}{}^2/\sigma_{EU}{}^2\right)\right] + \left(\bar{s}_{xj}{}^2/\sigma_{EU}{}^2\right)}$$

where $\bar{s}_{xj}^2$ represents the mean variance of the *J* items. This form of the index is based on the Spearman-Brown prophecy formula (Spearman, 1910; Brown, 1910), which predicts the reliability of a scale (or test) after changing the scale length. Although a value of zero for both r_{wg} and $r_{wg(J)}$ represents no agreement compared with the underlying "null" distribution, negative values are possible, representing even less agreement than would be expected by chance. (The lower bound of the indices is determined by the relationship between the null distribution variance and the maximum possible range of the variable in question: if this is unbounded, then technically so are r_{wg} and $r_{wg(J)}$.) Both indices would have the value 1 in the case of perfect agreement (no diversity of responses).

$r_{wg(J)}$ has been criticised as a measure of agreement for several reasons (LeBreton, James & Lindell, 2005; Lance, Butts & Michels, 2006). First, it relies on the specification of a "null" distribution – one which would be expected by chance alone. Most researchers favour a uniform distribution for this; indeed, this was suggested by James et al. (1984). However, this fails to take into account that random response to questionnaire items may not be completely uniform – in fact, there is plenty of evidence to suggest that respondents have a central tendency when responding to some types of

Likert scales (e.g. Bardo, Yeager & Klingsporn, 1982; Greenleaf, 1992). Therefore the default use of a uniform null distribution may not be appropriate. Second, its use as an agreement index is usually to justify sufficient levels of agreement for aggregation of individual data to a unit level. In order to provide justification, a cut-off is usually used above which respondents are said to agree. Following George (1990) the value of 0.70 has been taken by many researchers as the cut-off to use. However, the use of this cut-off has been substantially debunked by Lance, Michels and Butts (2006), and no alternative single value has been successfully proposed (indeed, whilst the index relies on specifying an arbitrary null distribution, such a cut-off value cannot exist). Third, it has been argued that the use of the Spearman-Brown prophecy formula to create $r_{wg(J)}$ is inappropriate as r_{wg} is not itself a reliability index (although LeBreton et al., 2005, have since shown that $r_{wg(J)}$ can be derived independently of the Spearman-Brown formula), and that corresponding values of $r_{wg(J)}$ can be insensitive or overly sensitive to small changes in $\bar{s}_{xj}^{\ 2}$ when J is large (Lindell, Brandt & Whitney, 1999). Fourth, the lower bound of $r_{wg(J)}$ is usually negative, and its precise value depends on the rating scale used and the null distribution chosen. Despite all of these flaws, $r_{wg(J)}$ remains popular as a measure of agreement for justifying aggregation of data (Newman & Sin, 2009).

Recognising some of the limitations of the r_{wg} index, Lindell, Brandt and Whitney (1999) proposed an adjustment to it, r_{wg}*. This alternative index addresses some, but not all, of the limitations. It eliminates the Spearman-Brown prophecy formula problem by using the overall scale sample variance, rather than the variance for each item:

$$r_{wg}^{\ *} = 1 - \frac{\bar{s}_x^{\ 2}}{\sigma_{EU}^{\ 2}}$$

Moreover, Lindell et al. (1999) suggest that, to prevent negative values occurring, σ_{EU}^2 is set to the *maximum* possible variance (rather than that for a uniform distribution). However, this does not address the question of whether this maximum variance is an appropriate choice for the construct, nor whether an appropriate cut-off may exist for determining whether agreement is sufficient to justify aggregation. Furthermore, it can easily be seen that r_{wg}* is a linear transformation of the scale sample variance, so beyond its maximum and minimum values (and orientation), its properties would be identical to those of the variance, described earlier. Neither r_{wg} nor r_{wg}* has gathered much use as an indicator of group-level diversity, although they have been considered alongside other indices, in particular by Harrison and Sin (2005), and by Roberson, Sturman and Simons (2007), who focused particularly on measures of climate strength.

4.1.6 Euclidean distance

The Euclidean distance is a mathematical concept that, in its simplest form, measures the shortest distance between two points in any given number of dimensions. It is based on the Pythagorean result that, if two points **x** and **y** in n-dimensional space have coordinates $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$, then the distance between them is given by the formula

$$|\mathbf{x} - \mathbf{y}| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

The Euclidean distance, in this original form, is often used in relational demography research (see e.g. Tonidandel et al., 2008), where it can be used to measure differences between pairs of individuals across multiple attributes at once (particularly if the attributes are measured on the same scale). However, it has also gained wider use in group diversity research, being mentioned as one of the suggested indicators of separation by Harrison and Klein (2007). Harrison and Klein give a different formula, for a single variable across multiple members of a group:

$$Mean\ Euclidean\ Distance = \frac{1}{n}\sum_{i=1}^{n}\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(x_i - x_j\right)^2}$$

The basic principle behind this is that each individual's mean distance from all other individuals is calculated, and then these are averaged across all members of the group. The division by *n* at each stage is included to ensure that larger groups do not automatically have larger mean distances. This is therefore a two-stage approach, clearly stemming from the relational demography perspective, whereas a more direct, one-stage approach could be seen as appropriate; this was the method used by Gevers & Peeters (2009), whose formula was:

$$Distance = \sqrt{\frac{1}{n-1}\sum_{i \neq j}\left(x_i - x_j\right)^2}$$

It could be argued, however, that the division here should be by $n(n-1)$, as the use of $i \neq j$ means that there are this number of squared differences included in the summation, and therefore this index would be biased by team size.

These formulae appear to stem from an inappropriate interpretation of the different dimensions within the mathematical formula for the Euclidean distance. If considered for a single dimension, the Euclidean distance between two individuals would simply be the absolute value of the difference in scores between those individuals, and the mean Euclidean distance would be the average of these absolute values. This would then be identical to the coefficient of mean difference (introduced in the next section).

The presence of different forms of the mean Euclidean distance in the diversity literature is potentially dangerous, as researchers may use the same term to use different things without realising that this is what they are doing. To maintain consistency of terminology I will use the term "mean Euclidean distance" to refer to the form as presented by Harrison and Klein (2007) unless otherwise stated.

The mean Euclidean distance has some desirable properties: it has value 0 when all members of the group have the same value of x, while its maximum value is bounded only by the possible range of the variable being measured: if the range were r, the maximum value would be $r/\sqrt{2}$, and this would be achieved when half of the group members had the maximum possible value of x, and the other half had the minimum possible value.

4.1.7 Coefficient of mean difference

The coefficient of mean difference (CMD) has existed in various forms for well over a hundred years. It has frequently been credited to Kendall and Stuart (1977) or Gini (1912), but has in fact been traced back to at least Helmert (1876) (and was thought to be around before even that). It is equivalent to the pure form of the mean Euclidean distance described in section 4.1.6 (as opposed to Harrison and Klein's version): a version based on average absolute difference, rather than a root mean square difference. Its formula is given by:

$$\mathrm{CMD} = \frac{1}{n(n-1)} \sum_{i,j=1}^{n} \left| x_i - x_j \right|$$

The principle behind this is that the difference between each pair of individuals within a group is averaged (there being $n(n-1)$ potentially non-zero differences between group members).

Despite its long history, CMD has relatively little use within the work group diversity literature, notable exceptions being Reagans and Zuckerman (2001), and Reagans, Zuckerman and McEvily (2004), who used it to measure tenure diversity. Like the Euclidean distance, it has value 0 when all members of the group have the same value of x, while its maximum value is achieved when half of the group members had the highest possible value of x, and the other half had the lowest possible value. The maximum value in this case, for a possible range of r, would simply be $r/2$.

4.1.8 Gini Index

Many different indices have been attributed to Gini (1912) over the years, causing a certain degree of confusion when one is simply referred to as the "Gini Index" or "Gini Coefficient". Certainly the Italian statistician's seminal 1912 paper has proved influential in many spheres: not just organisational studies, but sociology, economics, ecology and engineering have also found use for the different forms of dispersion index proposed. However, in keeping with the literature on work group diversity, I refer here to the index cited in Harrison and Klein (2007) as a potential measure of disparity. This has the formula:

$$\mathrm{G} = \frac{1}{2n^2\bar{x}} \sum_{i,j=1}^{n} |x_i - x_j|$$

As such, it is similar to the coefficient of variation (in that it is a measure of dispersion divided by the mean); the measure of dispersion in question is similar to the CMD, but instead divides by n^2 rather than $n(n - 1)$ (effectively looking at the mean difference of all pairs including pairs comprising the same individual). It also divides by 2, due to the interpretation of the index in terms of the Lorenz curve (Allison, 1978).

It has been widely used in the sociological literature (e.g. Allison, 1978; Coulter, 1989; Kimura, 1994; Lai et al., 2008; Mills, 2009), but less so in the work group diversity literature. At first glance, it may appear that this is because of the division by the group mean being an undesirable property, but this is countered by the widespread use of the coefficient of variation in the same literature. Two articles that did use it were Pfeffer and O'Reilly (1987), and Allen et al. (2007) – each of these also compared it with other indices. Like most other measures, it has its minimum value of zero when all group members have the same value for x; however, its maximum value occurs not when the group is polarised in equal terms, but when all but one group member have the minimum possible value and the other member has the maximum possible value – in this case G would equal $1-(1/n)$ as long as the minimum possible value of the variable is zero. Like the coefficient of variation, it should only be used for ratio data, due to the division by the mean being otherwise meaningless.

4.2 Ordinal measures

Indices for ordinal variables are seldom found within the work group diversity literature, despite the frequency of ordinal measurement. Often researchers have inappropriately used indices that were designed for either continuous (e.g. Jehn & Bezrukova, 2004) or nominal (e.g. Kearney & Gebert, 2009) variables; the articles by Harrison and Sin (2005) and Harrison and Klein (2007) are notable for their absence of recommendations about diversity of ordinal variables (Harrison and Sin do mention ordinal variation, but only recommend measures for categorical data that would fail to take into account the ordered nature of the variables). Therefore this section is somewhat shorter than the equivalent sections for continuous and nominal data, and relies mainly on indices from the sociological literature, with one index from the organisational literature only (a_{wg}). Even though the indices from sociology may have not been used in the work group diversity literature to date, their study is useful when considering the overall range of potential measures in the wider literature, not least so that work group diversity researchers are aware of their existence, and to give some frame of reference for considering what the most appropriate indices may be.

4.2.1 Measures from the sociological literature

Blair and Lacy (2000) described the d^2 index, which compares the distribution of an ordinal variable with that of a maximally dispersed one – that is, one with half of the group members in the highest category and half in the lowest. The formula is:

$$d^2 = \sum_{i=1}^{k-1} \left(F_i - \frac{1}{2}\right)^2$$

where F_i is the cumulative proportion for the ith category; that is, $F_i = \sum_{j=1}^{i} p_j$, with p_i representing the sample proportion for the jth of the k categories.

d^2 has its minimum value, 0, when the maximum dispersion described above occurs – that is, when half of the group members are in the highest and the other half in the lowest category. Therefore this is not an index of diversity as such, but an index of concentration (though a reversal of the score will of course solve that if diversity is the property being studied). The maximum possible value, representing no diversity but minimal concentration, is $(k – 1)/4$, and therefore is dependent on the number of possible categories of x. It is therefore a non-normed measure.

In contrast, the index l^2 proposed by Blair and Lacy (1996) gives a normed measure of concentration. It is defined as:

$$l^2 = \frac{d^2}{d^2_{max}}$$

where d^2_{max} is $(k-1)/4$ as described above. This then ranges from 0 (maximum diversity) to 1 (minimum diversity). A corresponding diversity index, therefore, would be given by $1 - l^2$. Blair and Lacy also describe the measure l, the square root of l^2, which is analogous to the Euclidean distance.

Other measures of ordinal dispersion that have appeared in the sociological literature are related to these. Blair and Lacy (2000) showed that Berry and Mielke's (1992) *IOV* is equivalent to $1 - l^2$, and Kvålseth's (1995) *COV* is equivalent to $1 - l$. Leik's (1966) index, *LOV*, is an unsquared ("city block") version of l^2, where the calculation is based on the raw frequencies rather than squared adjusted frequencies, as d^2 and l^2 are. Its formula, $\frac{2}{k-1}\sum_{i=1}^{k-1} B_i$, where $B_i = F_i$ if $F_i < ½$, else $B_i = 1 - F_i$, lacks the more subtle distance measurement of the other indices, and therefore may be computationally more straightforward but does not reflect smaller discrepancies from the extreme situations.

4.2.2 a_{wg}

An alternative to the r_{wg} agreement index for agreement on Likert scale data was a_{wg}, proposed by Brown and Hauenstein (2005). Despite its similar name, it takes a very different perspective from r_{wg}, instead being designed as analogous to Cohen's (1960) kappa – an agreement index for categorical variables. Like James et al. (1993), Brown and Hauenstein (2005) based their index on a ratio of observed variance to a theoretically calculated variance score; however, a_{wg} uses the maximum possible variance for a given mean, thus avoiding the artefactual problem of there existing a correlation between means and variances on a scale of limited range (e.g. a Likert scale). It is defined in terms of responses to Likert scale data, rather than more general continuous data, and hence I consider its use to be for ordinal data only. Its formula is as follows:

$$a_{wg} = 1 - \frac{2{s_x}^2}{[(H+L)M - (M^2) - (H \times L)] \times [k/(k-1)]}$$

where *H* and *L* are the highest and lowest possible values of the scale, *M* is the observed mean score, and *k* is the number of raters. It is scaled so that the maximum value (representing complete agreement, i.e. no diversity) would yield a value of 1, and maximum disagreement a value of -1. This has the consequence that the value of zero has no intrinsic interpretation (which it does with most indices), but merely represents half of the maximum variance possible.

As a relatively recent index, a_{wg} has yet to become widely used, with its implementation so far only as an agreement index, or as an index of climate strength (Roberson et al., 2007). Like r_{wg} and $r_{wg(J)}$*, it is best seen as a transformation of the sample variance (LeBreton & Senter, 2008), but in this case the transformation takes account of the sample mean. Its use of the highest and lowest possible values of the scale implies it should only be used with bounded variables.

4.3 Categorical (nominal) measures

4.3.1 Blau's (Simpson's) Index

The most commonly found measure of categorical diversity in the organizational literature usually goes by the name of Blau's (1977) index, after its appearance in the popular book "Equality and Heterogeneity" by the sociologist Peter Blau. Although Blau (1977) presented this without reference, he was not the first to suggest its use: it was originally proposed by Simpson (1949), and has been known by other names since then, notably including the "Herfindahl-Hirschman" index, and occasionally (and confusingly) the "Gini index" (e.g. Tsui, Egan & Xin, 1995). Nevertheless, such is the near-ubiquity of the name “Blau’s index” that I shall refer to it as such for the remainder of this article (a rare exception in the work group diversity literature is Leonard, Levine & Joshi, 2004, who do refer to it as Simpson’s index). Blau’s index has been used in the organisational literature to measure diversity on a wide variety of attributes, including race (e.g. Harrison et al., 1998; Mayo, Pastor & Meindl, 1996), gender (e.g. Harrison et al., 2002; Ely, 2004), functional background (e.g. Knight et al., 1999; Pegels, Song & Yang, 2000), education (e.g. Murray, 1989; Ferrier, 2001), and marital status (Harrison et al., 1998).

One reason for the popularity of Blau’s index is its simplicity, both in terms of calculation and interpretation. It is given by the formula

$$D_B = 1 - \sum_{j=1}^{k} {p_j}^2$$

where p_i is the proportion of members of a group belonging to category i for each of k categories present in the group. It is defined as the probability of any two group members, sampled at random, being from the same category. It has a theoretical range of 0 to 1, where 0 represents no diversity – i.e. all group members belong to the same category.

This interpretation is intuitively appealing for measurement of work group diversity – it is effectively the inverse of probability that a group member would meet someone of a similar nature within the group. For less diverse groups, this probability would be larger; for more diverse groups, this would be smaller. However, this definition is based on *sampling with replacement* – that is, the same member could be sampled twice for the purposes of the definition. Although this represents a reasonable approximation in large populations, for work groups and teams (which are often small in number) this is not the case. In particular, if all members of a group belong to different categories (say, different functional backgrounds), the probability of two individuals being from different backgrounds should be 1, whereas Blau's index would give this as $1 - 1/n$. Some authors (e.g. Harrison & Klein, 2007) suggest an adjustment to correct for this:

$$D_{B-adj} = 1 - \sum_{j=1}^{k} \frac{n_j(n_j - 1)}{n(n-1)}$$

where n_j is the number of members of the group belonging to the jth category. This truly does vary between 0 and 1, with 0 representing no diversity and 1 representing each member belonging to a separate category. An alternative formula for this (and one which may be computationally easier) is:

$$D_{B-adj} = \frac{n}{n-1}\left(1 - \sum_{j=1}^{k} {p_j}^2\right)$$

Thus it is equivalent to multiplying Blau's index by $n/(n-1)$: a similar adjustment for bias due to group size as is commonly applied to the standard deviation, for example (for a proof of this result, see Appendix 1). Whether this adjustment is desirable or not is discussed later.

Related to this is the Index of Qualitative Variation (IQV) which has been used in other fields (particularly sociology), with the formula:

$$IQV = \frac{k}{k-1}\left(1 - \sum_{j=1}^{k} {p_j}^2\right)$$

i.e. the same as Blau's index but multiplied by $k/(k-1)$ (Wilcox, 1973). Here, k is the number of categories represented in the group – so as with the previous adjustment, there is a maximum value of 1 for all groups (attained when all group members belong to a separate category, or when all possible categories are equally represented).

4.3.2 Teachman's (Shannon's) Index

Teachman's (1980) index is, like Blau's (1977) index, a reincarnation of an earlier index. Unlike Blau (1977), however, Teachman (1980) cited the original source (Shannon, 1948), but the term "Teachman's index" still retains greater currency in the organizational literature, possibly because Teachman was the first to use it to measure diversity of work groups. Teachman's index has been used to measure diversity on various attributes in work groups, including race (e.g. Choi, Price & Vinokur, 2003; Pelled, Eisenhardt & Xin, 1999), gender (e.g. Randel, 2002; Van der Vegt & Janssen, 2003), functional background (e.g. Jehn & Bezrukova, 2004; Murray, 1989), education (e.g. Jehn, 1999; Foo, Sin & Yiong, 2006), and hierarchical status (Choi, 2007).

Teachman's index is almost as simple as Blau's index in terms of calculation, but its interpretation appears far less understood. It is given by the formula

$$D_T = -\sum_{j=1}^{k} p_j . \ln\left(p_j\right)$$

where ln is the natural logarithm function. It also has a minimum possible value of 0, when all group members are of the same category; unlike Blau's index, however, it is not bounded above but the maximum depends on the size of the group, the theoretical maximum for a group of size n being –ln(1/n).

Whereas Blau's index can be interpreted in terms of simple probability theory, the rationale for Teachman's index comes from information theory. It is an "entropy-based" measure – that is, the amount of uncertainty associated with the group in terms of the variable being measured (this is distinct from the definition of entropy in thermodynamics, but is based on a related concept). This formula is equivalent to the expected amount of "information" contained by any member of the group, taking into account communication between group members (Shannon, 1948). This was derived for the purpose of telecommunications, but its use as a diversity measure stems from this: the more "information" available to group members, the more "diverse" the group is in some sense of the word. Unlike Blau's index, which can be transformed to the IQV so it remains unrelated to group size, there is no unbiased correction for the fact that Teachman's index will be more greatly underestimated in small groups (Biemann & Kearney, 2010), although Roulston (1999) did provide a formula that will estimate the necessary correction, albeit with bias.

The similarities and differences between Blau's and Teachman's indices are a matter of important concern, as there is very little in the literature that compares the two. I discuss this further in section 4.4.3.

4.3.3 Count

One very simple measure of diversity across a categorical variable is simply to count the number of categories represented in a group. This was used, for example, by Fay et al. (2006) who looked at the effects of occupational diversity in breast cancer teams. The computation and face validity of this measure are therefore obvious.

The count has been criticised as a measurement of diversity in some fields (e.g. ecology) because it is strongly affected by the presence of outliers (Hill, 1973). However, in group diversity research where group sizes are typically small, this is likely to be less of a concern because individuals who are the only representative of their category are still likely to have some influence on the group as a whole – certainly more so than if far larger populations are considered (e.g. the entirety of large organisations).

4.3.4 Indices for Binary Variables

Binary variables are a unique case of nominal variables, because knowledge of the proportion of one of the categories alone is sufficient for knowledge about the whole sample. This is seen particularly when measuring sex diversity, when some researchers (e.g. Randel, 2002; Rentsch & Klimoski, 2001) have used the proportion of a group that is female (or male) to represent group diversity.

The measures described earlier in this chapter for nominal variables – Blau's index and Teachman's index – are both widely used to measure diversity on binary variables also. However, Williams and Meân (2004) argue that the proportion of female members (or, alternatively, the proportion of male members) is a more appropriate index as it is measured on an interval scale: that is, the difference between a ten-person group with no female members and a group with one female member is the same as the difference between a group with two female members and a group with three female members; something that Williams and Meân (2004) demonstrate is not the case for Blau's (1977) and Teachman's (1980) indices. A counter-argument to this could be, however, that the introduction of a single female member to a group which is otherwise all male is more salient than moving from two to three female members, and therefore Blau's or Teachman's index may be more appropriate if this fits better with the theory in question.

An alternative form of this index is described by Wegge et al. (2008) as the "Heterogeneity Index" (HI). This has the form:

$$\mathrm{HI} = 1 - \frac{n\ (majority)}{n\ (total)}$$

This potentially ranges from 0 to 0.5, and is a non-directional form of the proportion of women recommended by Williams and Meân (2004), but retains the property of interval measurement. It could alternatively be defined as the proportion of group members in the minority, or P(minority). As a pure index of group diversity this is more relevant therefore, as it retains the symmetrical assumption ascribed to nominal variables, but individual research studies may focus on a particular gender imbalance, in which case the proportion of women (or men) may be more appropriate.

4.4 Comparison of indices

4.4.1 Comparison of continuous indices

The continuous indices of diversity described earlier in this chapter can be classified in one of four ways: (i) whether they are based on a raw score, variance or ratio metric, (ii) if on a raw score metric (i.e. can be interpreted on the same scale as the original variable), whether they based on an absolute value or a root mean square (RMS) calculation, (iii) if on a variance metric, whether they are absolute or comparative indices, and (iv) whether they are based on differences from individuals to a central point or on differences between individuals.

The indices, with some of the properties of each, are described in Table 4.1. This helps to distinguish between the indices on the basis of their properties. Those based on a raw score metric (standard deviation, AD_M, mean Euclidean distance, and CMD) are interpretable on the same scale as the original variable; those based on a ratio metric (coefficient of variation and Gini's coefficient) are useful for asymmetrical ratio variables only; while those based on a variance metric (variance, r_{wg}, r_{wg}*, a_{wg}) are mainly used to assess agreement (particularly the latter three, which are all comparative indices). There are also clear distinctions between those indices based on an absolute value calculation (AD_M, CMD and Gini's coefficient) and those based on an RMS calculation (standard deviation, coefficient of variation, mean Euclidean distance); and between indices which compare pairs of group members directly (mean Euclidean distance, CMD, Gini's coefficient), and those that compare members with the group mean (all other indices).

Thus, understanding the differences between the assorted indices is not particularly difficult. The question then becomes: which (if any) is the most appropriate index for each type of diversity? This will be addressed in Chapter 5.

Table 4.1: Properties of diversity indices for continuous variables

Index	Formula	Properties
Standard deviation (SD)	$\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$	Raw score metric with RMS calculation. Based on differences from group mean. Lower bound 0 (no diversity); no upper bound.
Coefficient of variation (CV)	$\frac{SD_x}{\bar{x}}$	Ratio metric derived from RMS calculation. Based on differences from group mean. Lower bound 0 (no diversity); no upper bound.
Variance	$\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$	Variance metric, absolute (not comparative) index. Based on differences from group mean. Lower bound 0 (no diversity); no upper bound.
Absolute deviation from the mean (AD_M)	$\frac{\sum_{i=1}^{n}\lvert x_i - \bar{x}\rvert}{n}$	Raw score metric with absolute value calculation. Based on differences from group mean. Lower bound 0 (no diversity); no upper bound.
r_{wg}	$1 - \frac{{s_x}^2}{{\sigma_{EU}}^2}$	Variance metric, comparative index. Based on differences from group mean. Upper bound 1 (no diversity); lower bound depends on comparator distribution.
r_{wg}*	$1 - \frac{{\bar{s}_x}^2}{{\sigma_{EU}}^2}$	Variance metric, comparative index. Based on differences from group mean. Upper bound 1 (no diversity); lower bound depends on comparator distribution.
a_{wg}	$1 - \frac{2{s_x}^2}{[(H+L)M - (M^2) - (H \times L)] \times [k/(k-1)]}$	Variance metric, comparative index. Based on differences from group mean. Upper bound 1 (no diversity); lower bound -1.
Mean Euclidean distance	$\frac{1}{n}\sum_{i=1}^{n}\sqrt{\frac{1}{n}\sum_{j=1}^{n}(x_i - x_j)^2}$	Raw score metric with RMS calculation. Based on differences between individuals. Lower bound 0 (no diversity); no upper bound.
Coefficient of mean difference (CMD)	$\frac{1}{n(n-1)}\sum_{i,j=1}^{n}\lvert x_i - x_j\rvert$	Raw score metric with absolute value calculation. Based on differences between individuals. Lower bound 0 (no diversity); no upper bound.
Gini's coefficient	$\frac{1}{2n^2\bar{x}}\sum_{i,j=1}^{n}\lvert x_i - x_j\rvert$	Ratio metric derived from absolute value calculation. Based on differences between individuals. Lower bound 0 (no diversity); no upper bound.

4.4.2 Comparison of ordinal indices

The six indices of ordinal variation described earlier in this chapter are shown in Table 4.2, together with some basic properties of the indices.

Table 4.2: Properties of diversity indices for ordinal variables

Index	Formula	Properties
d^2	$\sum_{i=1}^{k-1}\left(F_i - \frac{1}{2}\right)^2$, where $F_i = \sum_{j=1}^{i} p_j$	Compares distribution with that of maximum separation. Lower bound 0 (maximum diversity); upper bound $(k-1)/4$.
l^2	$\frac{4}{k-1}\sum_{i=1}^{k-1}\left(F_i - \frac{1}{2}\right)^2$	Normed version of d^2. Lower bound 0 (maximum diversity); upper bound 1.
l	$\left[\frac{4}{k-1}\sum_{i=1}^{k-1}\left(F_i - \frac{1}{2}\right)^2\right]^{1/2}$	Raw score normed version of d^2. Lower bound 0 (maximum diversity); upper bound 1.
LOV	$\frac{2}{k-1}\sum_{i=1}^{k-1} B_i$ where $B_i = F_i$ if $F_i < ½$, else $B_i = 1 - F_i$	"City block" distance measure. Lower bound 0 (no diversity); upper bound 1.
IOV	$1 - \frac{4}{k-1}\sum_{i=1}^{k-1}\left(F_i - \frac{1}{2}\right)^2$	Reversed version of l^2. Lower bound 0 (no diversity); upper bound 1.
COV	$1 - \left[\frac{4}{k-1}\sum_{i=1}^{k-1}\left(F_i - \frac{1}{2}\right)^2\right]^{1/2}$	Reversed version of l. Lower bound 0 (no diversity); upper bound 1.

It is worth noting that all indices are based on the cumulative relative frequency distribution, as this contains all of the distributional information of any ordinal variable (Blair and Lacy, 2000). As such, it can be informative to interpret the indices by considering the $(k-1)$-tuple $(F_1, F_2, \dots, F_{k-1})$, where $F_i = \sum_{j=1}^{i} p_j$, and comparing this to known distributions of high and low dispersion. For example, in the situation where all members fell into the first category, the associated $(k-1)$-tuple would be $(1, 1, \dots, 1)$. In a situation where all fell into the last category, it would be $(0, 0, \dots, 0)$. If half members fell into the first and half into the last (a classic situation of high dispersion), the associated ($k-1$)-tuple would be (½,½, …, ½). If there were completely even distribution between the k categories the ($k-1$)-tuple would be $\left(\frac{1}{k},\frac{2}{k},\ldots,\frac{k-1}{k}\right)$. Consideration of these distributions will be important when deciding which (if any) of these indices is appropriate to measure which type of diversity.

It is also worth noting that, of the six indices presented, five are closely related, varying only by norming, reversing and square rooting. The exception is Leik's (1966) *LOV* index, which is to Blair and Lacy's *l* index what the AD_M is to the standard deviation: an index based on absolute differences rather than a root mean square measure.

4.4.3 Comparison of nominal indices

A comparison of nominal indices is a little different, owing to the very different types of indices considered. I have commented on the merits of indices for binary variables in section 4.3.4. The total number of categories represented (count) is a straightforward index that requires no detailed explanation, and the Index of Qualitative Variation (IQV) is just a standardised version of Blau's index. This leaves, then, a comparison of Blau's and Teachman's indices (including the adjusted version of Blau's index), as the main task. Formulae for all of these indices, with some properties, are shown in Table 4.3.

Table 4.3: Properties of diversity indices for nominal variables

Index	Formula	Properties
Blau's index	$1-\sum_{j=1}^{k} {p_j}^2$	Probability of two group members being from the same category, sampled with replacement. Lower bound 0 (no diversity); upper bound 1 – 1/n.
Adjusted Blau's index ($D_{B\text{-}adj}$)	$\frac{n}{n-1}\left(1-\sum_{j=1}^{k} {p_j}^2\right)$	Probability of two group members being from the same category, sampled without replacement. Lower bound 0 (no diversity); upper bound 1.
IQV	$\frac{k}{k-1}\left(1-\sum_{j=1}^{k} {p_j}^2\right)$	Probability of two group members being from the same category, adjusted for categories represented. Lower bound 0 (no diversity); upper bound $k(n-1)/n(k-1)$.
Teachman's index	$-\sum_{j=1}^{k} p_j . \ln(p_j)$	Entropy-based measure. Lower bound 0 (no diversity); no upper bound.
Count	k (such that $n_i \neq 0$)	Simple number of categories represented. Lower bound 1 (no diversity); no upper bound.
% in one category	p_1	Simple percentage, relates to one category only. Lower bound 0 (no representation); upper bound 1 (no diversity).
Heterogeneity index (HI)	$1-\frac{n\ (majority)}{n\ (total)}$	Symmetric form of percentage in one category for binary variables. Lower bound 0 (no diversity); upper bound 0.5.

First, I consider Blau's index and its adjusted form, $D_{B\text{-}adj}$. The adjustment is straightforward to compute (multiplying the value of Blau's index by $n/(n-1)$, where n is the group size), but it does beg the question of whether it is in fact desirable. At first glance, it does appear that the interpretation brought about by the adjustment – a raw probability of any two different group members being from different categories – would be beneficial, and in some cases this might be what is required by the researcher. If the aim is to describe how diverse a group is relative to the maximum possible diversity for a group of its size, then the adjustment is appropriate – because the maximum diversity of 1 would describe that no matter the group size. However, if comparing groups of different sizes, then for example a group of 6 members, say would have an (unadjusted) maximum diversity of 0.833, whereas a group of 8 members would have an (unadjusted) maximum diversity of 0.875 according to Blau's index. Thus the unadjusted index would reflect the reality that, when all members of the group belong to different categories, the larger group is more diverse because more categories are represented (which would be consistent with the theoretical position that more information was potentially available to the group, for example) – whereas the adjusted index would have both groups with the same diversity, i.e. 1.

Biemann and Kearney (2010) expand on this argument, arguing that values of Blau's index are systematically biased by group size. Using simulated data, they demonstrate that using Blau's index to represent diversity when it is "correctly" represented by Harrison and Klein's (2007) corrected form of the index leads to underestimation by as much as 26.5% in groups of three people. However, this relies on the assumption that the corrected form is, indeed, the correct representation of diversity – which, as I have explained above, may not be the case.

A more complex comparison is that between Blau's and Teachman's indices. These are worth comparing for a number of reasons. First, in the work group diversity literature, both measures are highly prevalent, and both claim some prominence of use (Blau's index is used in substantially more of the studies cited in Chapter 2, and yet according to Tsui, Egan and Xin, 1995, Teachman's index is the 'most widely accepted' demography measure for categorical variables).

Second, despite their prevalence, very few authors have directly compared the two. Certainly the major articles about diversity measurement by Harrison and Sin (2005) and Harrison and Klein (2007) recommend both indices, with no clear criteria for choosing one or the other suggested. Likewise, Biemann and Kearney suggest corrections for the two indices without advising the use of one or the other in particular. Pelled, Eisenhardt and Xin (1999) were the first authors of an empirical study to mention the choice between indices, choosing Teachman's index for their main analysis, but repeatedly using Blau's index for a sensitivity analysis (and found the same pattern of results). They did not explain the reasons for their initial choice of Teachman's index besides citing previous literature, however.

Foo, Sin and Yiong (2006) provide the only empirical study of work group diversity to date that discusses the relative merits of work group diversity. They cite Tsui, Egan and Xin (1995)'s book chapter to argue that they should "not use the Blau Index because it is sensitive to the underlying frequency distribution that results in left-skewed distributions… [making it] unsuitable for this study where in some teams all members fell into the same category" (p. 393). Tsui et al. actually refer to this index both as Blau's index and the Gini index, and cite Allison (1978) as saying that a sensitivity to the underlying frequency distribution results in an overweighing of left-field distributions. Careful study of Allison's article, however, reveals that he does not make reference to Blau's index at all, and his observations were based on Gini's index of concentration (described in section 4.1.9) – a measure for continuous variables. This means that Tsui et al.'s (1995) rationale, and therefore Foo et al.'s (2006) too, are unfounded.

In the absence therefore of any clear message from the work group diversity literature to help choose between the indices, more can be learned from other disciplines. In particular, comparisons of these (or related) indices in sociology by Taagapera and Ray (1977), and in ecology by Hill (1973), provide some insight.

At first sight, there is an obvious link between the formulae for the two indices – one is based on multiplying probabilities by themselves, the other on multiplying probabilities by their logarithms. However, the computational link between them is not obvious. To understand it, the related concept of "concentration" in the sociological literature can be helpful. Taagepera and Ray (1977) presented a generalized index of concentration given by the following formula:

$$C_n = \left[\frac{\sum_{i=1}^{N} P_i^n - N^{1-n}}{1 - N^{1-n}}\right]$$

Here, P_i are the proportions of each of the N groups in the overall population; n is a parameter defining the type of concentration. It is relatively straightforward to show that Blau's index, $D_B = C_2^2\ (1 - 1/N) + 1/N$; it is less obvious and requiring of a mathematical proof (supplied by Taagepera and Ray) that Teachman's index, D_T, is equivalent to ln $N(1\text{-}\ C_1)$. So the relationship between the two indices is given some light by comparing C_2 and C_1. Although Taagepera and Ray did not explicitly define n, they did show how C_n varies for different balances of a binary variable: this is reproduced in Figure 4.1. It can be seen that whereas C_2 (represented as "CON" in the diagram) increases almost proportionally to the proportions in each category, C_1 (represented as "RR") is far more sensitive to large imbalances in the proportions.

This actually contradicts the advice given by Tsui et al. (1995); it appears that Teachman's index would be far more sensitive to small changes in a minority than Blau's index (because the lines representing the less evenly spread groups, e.g. 95-5 and 85-15, are further apart for C_1 than for C_2). It is worth repeating this figure for the raw values of Blau's and Teachman's indices, rather than the transformed values of C_n. A bar chart showing equivalent data for these two indices is shown in Figure 4.2.

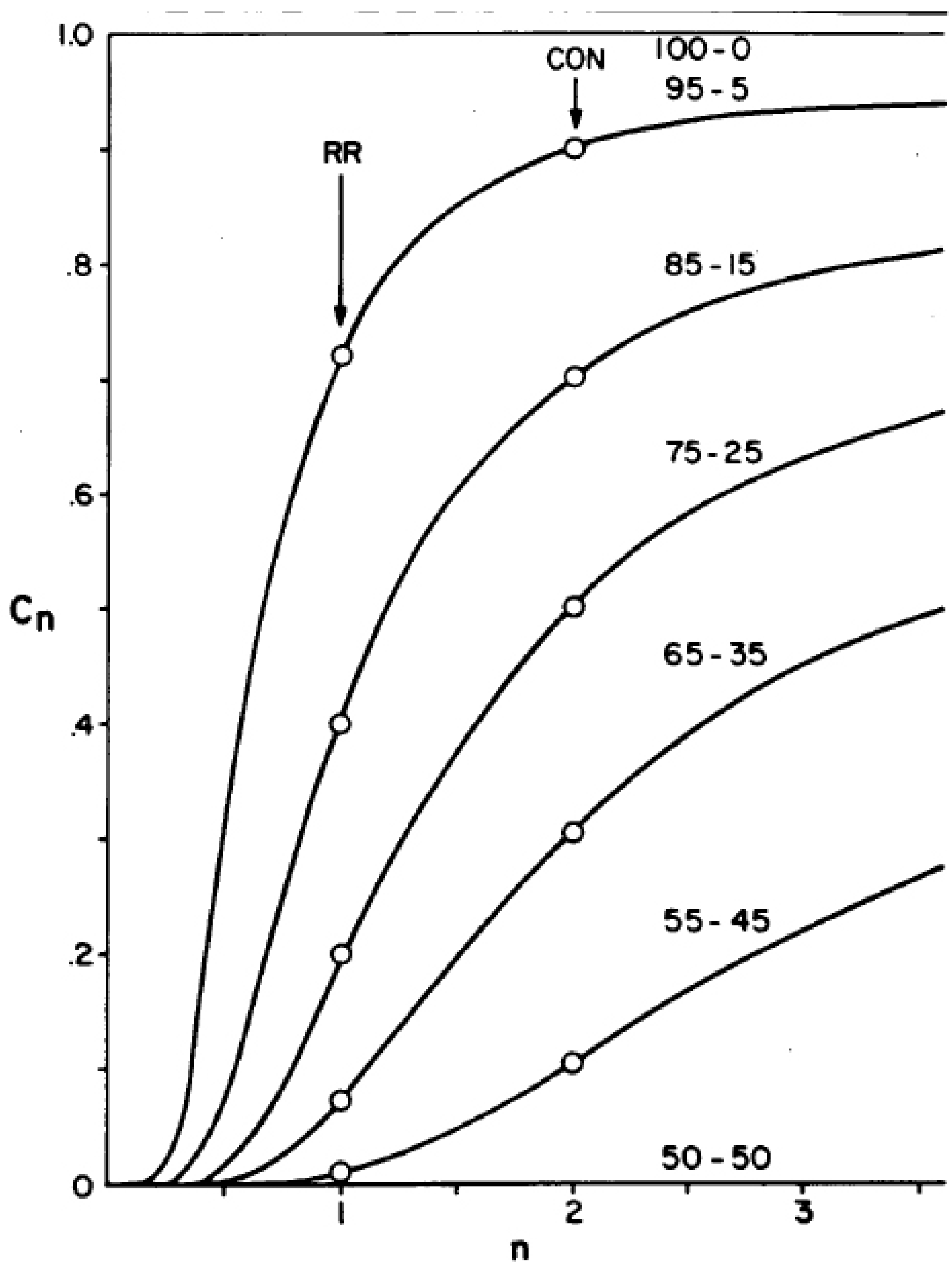


Figure 4.1: Taagepera & Ray's generalised index of concentration for different breakdowns of a binary variable

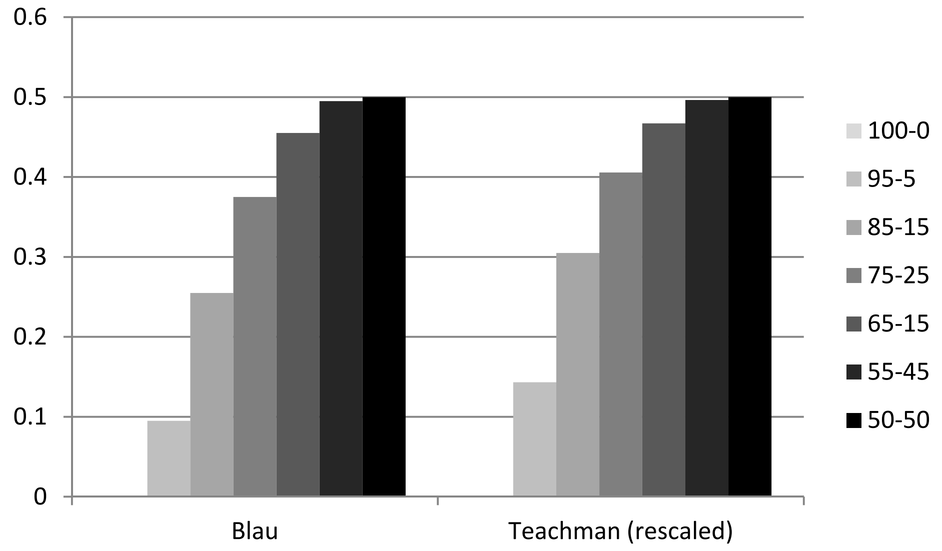


Figure 4.2: Comparison of Blau's and Teachman's indices for different breakdowns of a binary variable

Note that the data for Teachman's index in Figure 4.2 has been rescaled so that the maximum value (where the group is divided 50-50) is the same for both indices, namely 0.5, and also that the value of both indices when all members fall into one category is zero. However, it can clearly be seen that the pattern is consistent with that indicated by Taagapera and Ray (1977): Teachman's index is slightly more sensitive to changes in the composition when there is a clear majority/minority split, and less sensitive to slight changes from the equilibrium position. Thus, it appears that Tsui et al. (1995) inadvertently came to precisely the wrong conclusion about the two indices. Note that either sensitivity or lack of sensitivity might be desirable for a particular situation; for example, the literature on minority influence and decision making (e.g. Nemeth, 1986) would suggest that the presence of only one group member from a particular category could have an unrepresentative influence on a work group.

However, it is also clear that the indices bear a large level of similarity: they both monotonically increase with the breakdown of the variable in question; one index is consistently higher than the other; and it is likely that there would be a high correlation between the two on any given data set. Table 4.4 shows the values of the two indices for six fictional groups: the correspondence, if not proportionality, between them is clear to see.

Table 4.4: Blau's and Teachman's indices for different distributions within units

Team number	Distribution of categories	Blau's index	Teachman's index
(1)	A, A, A, A, A, A	0	0
(2)	A, A, A, B, B, B	0.5	0.69
(3)	A, A, B, B, C, C	0.67	1.10
(4)	A, A, A, B, B, C	0.61	1.01
(5)	A, A, A, A, B, C	0.5	0.87
(6)	A, A, B, B, C, D	0.72	1.33
(7)	A, B, C, D, E, F	0.83	1.79

In comparing methods for measuring the diversity of species in geographical regions, Hill (1973) more explicitly compared the two indices, to which he referred as Simpson's index and Shannon's index after their original authors. Hill also provides a formula for a generalised index of diversity, N_a:

$$N_a = \left[\frac{\sum_{i=1}^{n} w_i {p_i}^{a-1}}{\sum_{i=1}^{n} w_i}\right]^{1/(1-a)}$$

where w_i are weights that add to 1. Like Taagapera and Ray's (1977) C_n index, two forms of this are closely related to the two work group diversity indices under review. Specifically, N_1 is equal to the exponential of Teachman's (Shannon's) index, while N_2 is the reciprocal of [1 – Blau's (Simpson's) index]. Hill describes the continuum as a ranges from $-\infty$ to ∞; as a increases the nature of the diversity index "comes to depend more and more on the common species and less and less on the rare". In other words, higher values of a better represent "evenness" – the extent to which a distribution is evenly spread over all possible categories of a variable. Thus, Blau's index is more of a representation of evenness than Teachman's index. This tallies with the earlier suggestion from the Taagapera and Ray (1977) generalised index, which suggests that Blau's index is more sensitive to slight deviation from a position of equilibrium (where categories are equally represented, or "even"), but Teachman's index is more sensitive to changes in a strong majority/minority situation.

In summary, the choice between indices for measuring diversity of a nominal variable will involve some careful decisions. If the variable is binary, special options apply, but if not, the choice is between a simple count, and two fairly similar indices – Blau's (and its various adjusted forms) and Teachman's. The comparison of these presented here suggests the choice may come down to what type of sensitivity is most important for the theoretical framework being applied. How this relates to the different types of diversity will be discussed in the Chapter 5.

4.5 Chapter summary

In this chapter I have reviewed many indices used for measurement of work group diversity in the organisational and related literatures. Some have been used widely, others only occasionally. I have compared the suitability of each for measuring diversity, highlighting properties of each index, and identifying differences between them where they are not obvious (e.g., comparing Blau's and Teachman's indices).

The next step is to consider what the most appropriate indices are for measuring each type of diversity as defined in Chapter 3. Therefore in the next chapter I will consider the desirable properties for each diversity type, and match these up where possible with indices described in this chapter; where none exist I will adapt existing indices or derive new ones.

CHAPTER 5

A NEW DIVERSITY MEASUREMENT TYPOLOGY

5.0 Chapter introduction

In this chapter I develop a new typology for the measurement of work group diversity; this is based on Harrison and Klein's (2007) typology, but extends it in three ways. First, I provide definitions for four types of diversity (rather than Harrison and Klein's three types), three of which differ slightly from the Harrison and Klein (2007) definitions. Second, I clarify the types of variables to which these diversity types can be applied, and the principles of measurement underlying diversity indices. Third, I develop an index for each of the 14 possible combinations of diversity type and data type, using the definition as the basis for the form of the index. Some of these indices are already used in the literature; others are new. Finally, I conclude the chapter by presenting hypotheses about the comparative merits of using the proposed indices rather than other possible diversity indices.

5.1 Revised definitions of diversity: Separation, Range, Spread and Disparity

In Chapter 3, I extended Harrison and Klein's (2007) definitions so that they were not linked to specific types of variables, and so that variety was not a single construct but encompassed two different constructs, *range* and *spread*. The task now is to determine what the most appropriate methods of measurement are for each type of diversity and each type of variable. In order to do this, it is first helpful to review the new definitions, and clarify conditions when they would be at their highest and lowest.

S*eparation* is defined as "the extent to which unit members are different from one another with respect to a variable *X*". Crucially, although it is a group-level construct, it refers to the differences between individual pairs of unit members. It should be at its maximum when the total (or average) of differences between all pairs of group members is as high as possible, and at its minimum when all group members have the same value of *X*.

Variety (range) is defined as "the range of levels of a variable *X* represented within a group". If *X* is a continuous or ordinal variable, this would relate to the difference between the minimum and maximum values represented; if *X* is a nominal variable, then it would relate to the number of different categories represented. For a continuous or ordinal variable, it would be at its highest when the minimum and maximum possible values of *X* are represented within the group; for a nominal variable it would be at its highest when all different possible levels of *X* are represented within the group. In either case, its minimum value would occur when all group members have the same value of *X*.

Variety (spread) is defined as "the extent to which all possible levels of a variable *X* are equally represented within a group". This then encompasses both the concepts of range and evenness. For a continuous variable, it would be at its highest when group members are evenly spread across the whole possible range of a variable; for a categorical variable (ordinal or nominal), it would be at its highest when all possible categories are equally represented within the group. Again, its lowest value would occur when all group members have the same value of *X*.

Disparity is defined as "the extent to which there is inequality between unit members with respect to an asymmetrical variable *X*". The notion of asymmetricality implies that this cannot be defined for a nominal variable. It should be at its highest when one member of the group had the maximum possible value of *X* and all others had the minimum possible value. Once more, its lowest value would occur when all group members have the same value of X – regardless of what this value is.

5.2 Principles of measurement

Researchers have studied the diversity of many attributes within groups: a brief glance at the literature cited earlier reveals measurement of diversity of age, sex, ethnicity, nationality, culture, educational level, functional background, tenure, perceptions of climate, perceptions of leadership, work attitudes, personality, values and beliefs, amongst others. Given this wide range of foci of diversity, it is expedient to produce indices that are applicable to different variables, rather than specific attributes. As discussed in Chapter 4, indices used in the literature could be applied to any attribute with the same measurement type.

The theoretical alignment of an attribute with a type of diversity (separation, range, spread or disparity, as described in section 5.1) is the first step. Once the theoretical perspective for measurement is determined, it should not matter whether the attribute is in nature demographic, value-orientated, knowledge-orientated, or of any other form: the process for deriving a measurement of diversity from the raw variable depends only on the original form of measurement of the variable[1]. For example, ethnicity and functional background are very different attributes; however, if both are measured on a categorical (nominal) basis, and the definition of diversity is the same for each, then the process of forming the variables "ethnic diversity" and "functional background diversity" from the raw variables is identical.

---

[1] This assumes that the original form of measurement the variable represents the construct of interest. Sometimes this might not be the case: for example, age diversity may seek to examine the differences between those above and below retirement age. In such cases a categorisation of the raw variable into a new variable is entirely appropriate before diversity is calculated, and it is the categorised form – whether ordinal or nominal – that should determine the derivation of the diversity index.

It is therefore necessary to define only the different types of measurement of the original variables. Consistent with commonly accepted statistical/psychometric theory (e.g. Nunnally & Bernstein, 1994), and for the most part consistent with their use in the work group diversity literature, I use four different types of variables:

- Ratio: continuous variables with naturally meaningful values; i.e. scaled such that the value 0 represents an absence of the attribute in question (e.g. age, tenure, salary)
- Interval: scaled with meaningful differences but arbitrary values; i.e. if $x_a - x_b = x_c - x_d$, then the difference between *a* and *b* is equivalent to the difference between *c* and *d* (e.g. team climate, IQ), but without any assumptions about the value 0
- Ordinal: discrete ordered categories, no assumptions about nature of scale (e.g. ranks, individual Likert scale items)
- Nominal (categorical): discrete categories without natural order (e.g. nationality, functional background)

Principles for how indices for each of these types of variable should be derived are given in the following section.

In order to measure separation, variety (range and spread) and disparity in the most appropriate manner, I will present for each what (according to theory) would be the desirable properties of such an index, including the situations that would generate the highest and lowest values, as indicated in section 5.1. I will then use these properties to derive mathematically an index for each of the types of variable described above: ratio, interval, ordinal and nominal. These indices will often be the same as indices already used in the literature; however, where no such existing indices fits the definition accurately, a formula is derived from first principles.

It is recognised that the distinction between variables is not always as clear-cut as this, however. In particular, organizational researchers often use ordinal variables (e.g. Likert scales) as if they were interval variables (Beal & Dawson, 2007; Kotz, Balakrishnan & Johnson, 2000; Russell & Bobko, 1992). Whether an interval or ordinal measurement is used is left to the researcher, and should be consistent with the type of analysis usually performed with that variable. Additionally, nominal variables are not necessarily devoid of quantitative value, in the sense that some pairs of categories may be less different than other pairs. I do not include this possibility within the main typology, but instead discuss it further in Chapter 9.

For the sake of simplicity, I adopt the following notation: each diversity index is presented as D_{AB}, where A represents the type of diversity (S = separation, R = variety (range), V = variety (spread), and D = disparity), and B represents the type of variable (R = ratio, I = interval, O = ordinal, and N = nominal). For example, D_{SR} is the index to calculate the separation of a ratio variable.

5.3 Development of indices

5.3.1 Measurement of separation

Separation is defined as "the extent to which unit members are different from one another with respect to a variable *X*". The most important consideration for measurement here is that differences between individual pairs of unit members must be considered.

If the variable on which separation is to be calculated is a ratio or interval variable, then the difference between any pair of members can be calculated on the same scale on which the variable is measured (so, for example, if the variable in question is age, and three team members have ages 35, 40, and 50 years, then each pair of differences can be expressed in terms of years: 5, 10 and 15 being the three differences in this case). Therefore the group level variable should be expressed in terms of these differences; the obvious way to do this is via the average absolute difference between pairs. Expressed algebraically, this gives us a formula

$$D_{SR} = D_{SI} = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j<i} |x_i - x_j|$$

where the group has *N* members. The reason for this formula is that it is the total sum of all possible differences between individual members of the group, divided by the total number of pairs, ½ $N(N-1)$. This is the same as the Coefficient of Mean Difference (CMD) as used by many statisticians over the years (notably Gini, 1912 and Kendall & Stuart, 1977), but by relatively few work group diversity researchers (exceptions being Reagans & Klimoski, 2001, and Reagans, Zuckerman & McEvily, 2004); it is also equivalent to the pure form of the mean Euclidean distance with a single dimension (as opposed to Harrison and Klein's (2007) version). A slightly easier computational version of this formula is

$$D_{SR} = \frac{4}{N(N-1)} \sum_{i=1}^{N} i x_i - \frac{2(N+1)}{(N-1)} \bar{x}$$

where the individual values of *X* are ordered such that $x_1 \leq x_2 \leq \ldots \leq x_N$. (For a derivation of this formula, see Appendix 2).

Note that this index is different from those suggested by Harrison and Klein. One of their two suggestions for measuring separation was the standard deviation. However, this effectively compares each individual value with the group mean – not with other group members. This appears to be at odds with the definition of separation (whether Harrison and Klein's or mine), and the reason why this would be less appropriate as a measure of separation can be seen by looking at the social categorisation process, in which group members compare themselves with other group members, not some overall group average (Tajfel, 1978). Whilst comparing an individual value with the group mean is the same as the mean comparison with group members (including themselves), because the standard deviation uses a root mean square formula this does not produce the same result as the D_{SR} index.

Measurement of separation on ordinal variables cannot be achieved in exactly the same way. This is because the "distance" between any two points cannot be assumed to be a numerical quantity (as the differences between adjacent categories are not necessarily equal, by definition), and therefore taking an average of such a distance would be relatively meaningless. Instead, the measurement needs to take account of the relative distribution across the different categories. As noted previously, maximum separation should occur when half the group members have the highest possible value and the other half have the lowest possible value. Minimum separation should occur when all group members have the same value. If members are equally spread across all possible values, this should represent moderate separation. These are precisely the properties offered by Blair and Lacy's (2000) indices d^2, l^2 and l. (Note, however, that Blair and Lacy's indices need to be reversed to become measures of diversity rather than of concentration: that is, $1 - l^2$ and $1 - l$ are the appropriate indices rather than l^2 and l.) As described in Chapter 4, d^2 is a non-normed version of the index, and therefore depends on the number of possible categories, whereas l^2 scales this to between 0 and 1. Assuming the focus of diversity is a variable that has a consistent number of categories, it should not matter which of these indices is used. The question then becomes, however, whether the l^2-type measure, or the unsquared version, l, is more appropriate as a measure of separation.

To help determine this, it is useful to consider how the values would compare with the most appropriate index for an interval variable, i.e. assuming that adjacent categories were equally spaced. Although this is not necessarily realistic for an ordinal variable, it would help ensure that the indices were as consistent as possible. Table 5.1 shows five fictional ten-person teams with values of a variable that could be either interval (ranging from 1 to 5) or ordinal (with five categories). Team (1) is the condition of maximum separation, team (5) the condition of minimum separation, with the others falling in between. The adjusted CMD, or D_{SR} in my terminology – which represents separation for an interval variable – is divided by its maximum value so that it has the same potential scale as $1 - l^2$ and $1 - l$, making a direct comparison easier.

Table 5.1: Values of three diversity indices for an interval/ordinal variable

Team number	Values represented	$D_{SR}/D_{SR(max)}$	$1 - l^2$	$1 - l$
(1)	1, 1, 1, 1, 1, 5, 5, 5, 5, 5	1	1	1
(2)	1, 1, 1, 3, 3, 3, 3, 5, 5, 5	0.84	0.84	0.60
(3)	1, 1, 2, 2, 3, 3, 4, 4, 5, 5	0.80	0.80	0.55
(4)	2, 2, 2, 3, 3, 3, 3, 4, 4, 4	0.42	0.42	0.24
(5)	1, 1, 1, 1, 1, 1, 1, 1, 1, 1	0	0	0

It can clearly be seen that values of $1 - l^2$ are directly proportional to those of D_{SR}. Therefore I propose that the optimum method of measuring separation of ordinal variables is:

$$D_{SO} = 1 - l^2 = 1 - \frac{4}{k-1}\sum_{i=1}^{k-1}\left(F_i - \frac{1}{2}\right)^2$$

where F_i is the cumulative proportion for the ith category; that is, $F_i = \sum_{j=1}^{i} p_j$, with p_i representing the sample proportion for the jth of the k categories. Note that this does not imply that D_{SR} can be used in place of D_{SO} – an arbitrary shift in the values used to represent the ordinal variable would not change the meaning of the variable, or the calculation of D_{SO}, but would change the calculation of D_{SR}.

If X is a nominal variable, however, the notion of pairwise differences is not usually considered. Either two members belong to the same category, or they do not. If the distance between two members from separate categories is considered to be 1, and between two members from the same category is 0, then this gives rise to Blau's (1977) index, which is defined as the probability that any two members of the group, selected at random (with replacement), are from different categories. This is almost entirely in line with my definition of separation, the only difference being that sampling with replacement would not be appropriate. When the population is very large, the probability that the same individual is selected twice ($1/n$) is negligible, and therefore Blau's index is a very good approximation of the true probability when sampling without replacement. More appropriate, though, is the version of the formula that is based on sampling without replacement – i.e. the only differences between pairs of group members considered are where the pairs are of different individuals (rather than the same individual twice). For groups of the same size this would be directly proportional, but for groups of different sizes the uncorrected version of Blau's index would be biased downwards in smaller groups (Biemann & Kearney, 2009); this being particularly the case when the group size is less than or equal to the number of possible categories, so that each member could (potentially) belong to a different category. Therefore, I propose the corrected version of the index, which I presented in Chapter 4:

$$D_{SN} = 1 - \sum_{j=1}^{k} \frac{n_j(n_j - 1)}{n(n-1)} = \frac{n}{n-1}\left(1 - \sum_{j=1}^{k} {p_j}^2\right)$$

Note that in the case that the group size is the same as the number of potential categories ($n = k$), this is the same as the Index of Qualitative Variation. All of these proposed indices are summarised in Table 5.5.

5.3.2 Measurement of range

The range of a ratio or interval variable is easily captured. The normal method of measuring range, $x_{\max} - x_{\min}$, should suffice. It is worth noting, though, that in some cases, simply taking the maximum or minimum value may be sufficient. This is particularly likely to be the case in highly skewed variables, where many groups will have very similar mimina or maxima. Note that the range is not directly related to the standard deviation, the most commonly used measure of diversity in these types of variables - it is true that groups with a larger range will tend to have larger standard deviations, but this is not necessarily the case. Range of ordinal variables can be measured in the same way. Therefore I propose:

$$D_{RR} = D_{RI} = D_{RO} = x_{\max} - x_{\min}$$

The range of a nominal variable is also easy to calculate. It is simply the number of categories of the variable that are represented within the group. This is the measure used by Fay et al. (2006), who examined the relationship between the number of occupational groups represented within primary health care teams, and team innovation. Although Fay et al. did not refer to it as such, this is clearly a measure of variety – and range in particular – within the teams. Formally, this can be expressed by the equation

$$D_{RN} = \sum_{i=1}^{k} \delta_i$$

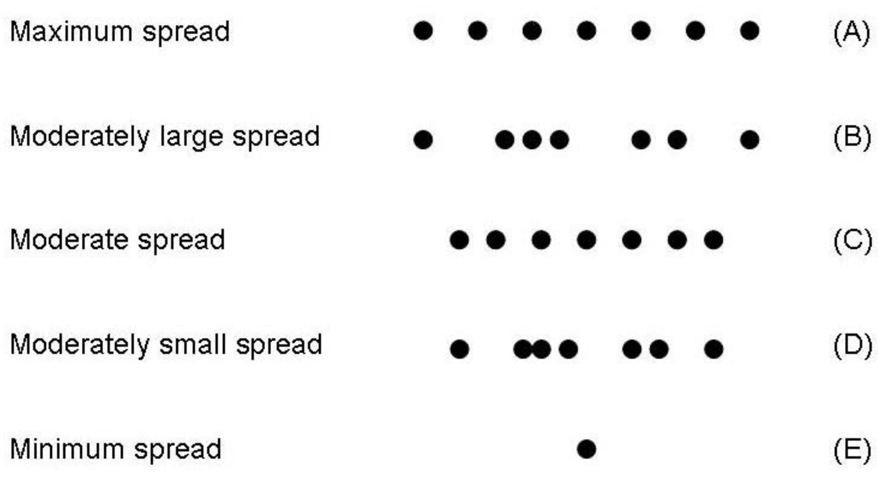
where $\delta_i = 1$ if $p_i > 0$, and $\delta_i = 0$ otherwise. However, it is far more easily understood as a simple count.

5.3.3 Measurement of spread

Spread is a more difficult concept to measure, as it covers the concepts of both range and evenness. It has a qualitatively different meaning for interval and categorical variables, although they are closely linked. For continuous (ratio and interval) variables, it effectively means "the extent to which all parts of the scale are represented by group members"; for categorical (ordinal and nominal) variables this becomes "the extent to which all possible categories are equally represented". The reason for this difference is the slightly different notion of "evenness" – for a continuous variable the most even scenario would be one where all sections of the scale are represented, but with equal distances between adjacent group members; for categorical variables, however, only a finite (and often small) number of possible values exist, so the most even scenario would be one where they are each represented by the same number of group members. However, in keeping with the concept of diversity as represented in the literature, spread should represent more than evenness alone – for example, a team of six members, each of a separate occupational background, would normally be considered more diverse than a team comprising two people from each of three occupational backgrounds, although they would both have maximum evenness.

For ratio and interval variables, it is worth first considering the desirable properties that this definition entails. First, a group with zero spread would mean all members have the same value (as is the case with all measures of diversity). Second, a group with maximum spread would include members at both extremes of the scale, with other members distributed equally between these. (Note that for some variables, there may not be absolute extremes – for example, there is no upper limit of age. In these situations it would be possible to use the extreme observed values across all samples as the extremes.) Third, a group whose values are equally distributed should have greater spread than a group with the same extreme values but whose other members are more bunched. Figure 5.1 shows five groups with different levels of spread. In group (A), the maximum spread is achieved - the seven group members are evenly distributed across the entire continuum. In group (B), the range is the same, but the spread is lower because the distribution of points is less even. In group (C), the distribution of members is even within the range represented by the group, but this is less than the possible range of the scale – the level of spread is probably fairly similar to that of group (B). In group (D), the range is the same as that of (C), but the distribution of scores is less even. Finally, in group (E), all members have the same value, representing minimum spread.

Figure 5.1: Representations of different levels of spread for ratio or interval variables



In the case of maximum spread, the distance between adjacent group members' scores is $(x_{\max} - x_{\min})/(N - 1)$, where $x_{\min}$ and $x_{\max}$ represent the minimum and maximum observed values of X in the group respectively. Thus, the maximum distance between any two adjacent group members is given by this formula. Where the extremes are the same, but the values less evenly distributed, this maximum distance will be larger, up to a maximum possible value of $x_{\max} - x_{\min}$ (which occurs when all group members occupy one of the extreme points). Thus, the evenness of the spread can be represented by

$$Evenness \propto \frac{x_{\max} - x_{\min}}{\max\,(x_i - x_{i-1})}$$

where the x_i are placed in order: i.e. the evenness is inversely proportional to the maximum possible distance between adjacent points. The reason it is inversely proportional is because a higher maximum distance implies a less even spread – in Figure 5.1, for example, group (B) has a larger maximum difference than group (A) and is therefore less even, whilst the same is true of groups (D) and (C) respectively.

The other component of spread, however, is the range: as we saw in Figure 5.1, the spread is larger when the overall range within the group is larger. Thus, the evenness should be multiplied by the observed range to obtain the spread:

$$Spread \propto \frac{(x_{\max} - x_{\min})^2}{\max\,(x_i - x_{i-1})}$$

Finally, we can scale this so that it has a maximum value of 1: to do this, we divide by the maximum possible value, which is $(X_{\max} - X_{\min})^2/((X_{\max} - X_{\min})/(n - 1))$, or just $(n - 1)(X_{\max} - X_{\min})$, where $X_{\min}$ and $X_{\max}$ represent the minimum and maximum possible (as opposed to observed) values of X. If X is theoretically unbounded at one or both ends, then an arbitrarily chosen $X_{\min}$ and $X_{\max}$ would suffice, as long they are at least as low and high as any observed values of X across all groups, and are consistently applied. Therefore we can define the variety (spread) of a ratio or interval variable as:

$$D_{VR} = D_{VI} = \frac{(x_{\max} - x_{\min})^2}{(n-1)(X_{\max} - X_{\min}).\max(x_i - x_{i-1})} \qquad (\text{if } x_{\max} - x_{\min} > 0)$$

$$D_{VR} = D_{VI} = 0 \qquad (\text{otherwise})$$

It is necessary to specify the value 0 for minimum spread to avoid division by zero. The five groups described above have spread of 1, 0.8, 0.75, 0.45 and 0 respectively with this measure.

Calculation of spread for categorical variables is somewhat different. As noted earlier, I define spread of nominal variables as “the extent to which all possible categories are equally represented”. For ordinal variables, this still needs to take some account of range, unlike for nominal variables; for example, if a group of six people had two members each from three out of five possible categories of a nominal variable, it should not matter which three categories were represented: the spread would be the same whichever combination of categories was present. For a five-level ordinal variable, though, the range would be greater if the three categories represented were 1, 3 and 5, rather than 2, 3 and 4. Thus, an analogy to the measure of spread for continuous variables would be to multiply a measurement of evenness for categorical (including nominal) variables by the range of the (ordinal) variable. Therefore I move to consider nominal variables and the issue of spread before returning to the issue of ordinal variables.

The definition of spread for a nominal variable is “the extent to which all possible categories are equally represented within a unit”. This is distinguished from the earlier definition of range for nominal variables (measured as a simple count) because the number of unit members with each category matters. Again, though, this encompasses the notions of both range and evenness. Given the concept of diversity as variety, therefore (and particularly bearing in mind its historical position in the information/decision-making perspective), there are two particular principles of measurement: (i) that a group with more categories represented has greater spread than one with fewer categories represented (assuming relatively stable levels of spread), and

(ii) when the same number of categories are represented, the one with a more even distribution has greater spread.

Blau's and Teachman's indices are naturally prime candidates for this, given their preeminent role in measuring diversity of nominal variables. Certainly, Blau's index (in its unadjusted form) is generally larger when more categories are represented; indeed it can be proven that if one member of a group were removed and replaced by a new member from a new category, then Blau's index would increase (unless the member removed were the sole representative from his/her category, in which case it would stay the same) – see Appendix 3. Likewise, the same is true for Teachman's index – again, a proof is provided in Appendix 3. Both indices are also at their highest (for a given number of categories) when members are evenly distributed across all categories. Therefore, the choice between them comes down to deciding which captures the notion of spread more effectively. Note that the unadjusted forms of the indices are necessary to capture the "range" part of the definition.

As reported in the previous chapter (section 4.4.3), Teachman's index is more sensitive to small changes in a minority when there are only two categories present – e.g. Teachman's index is relatively higher when just 5% of the group belong to the minority category. It is also worth considering what might happen with more than two categories. A relatively simple example would involve three five-person groups, with up to four available categories. The first group has members in categories A, A, B, B, C. The second has members in categories A, A, A, B, C. The third has members in categories A, A, B, C, D. The first group, being more evenly distributed than the second, should have greater spread. However, this difference is relatively minimal compared with the greater spread of the third group, which has greater range (an extra category – which, considering diversity as a source of information is highly salient), as well as reasonably good spread. The values of Blau's and Teachman's indices for these groups are shown in Table 5.3.

Table 5.3: Blau's and Teachman's indices for three different groups

Team number	Distribution of categories	Blau's index	Teachman's index
(1)	A, A, B, B, C	0.64	1.05
(2)	A, A, A, B, C	0.56	0.95
(3)	A, A, B, C, D	0.72	1.33

It can clearly be seen that the difference in Blau's index between groups 1 and 2 is the same as that between groups 1 and 3 (both being 0.08), despite the apparent greater salience of the added category in group 3. In contrast, however, Teachman's index does reflect this greater salience, and the difference between groups 1 and 3 is nearly double that between groups 1 and 2. This is due to the entropy-based nature of the index, which reflects the total information available to the group. Therefore, I propose that Teachman's index is a more appropriate measure of spread for nominal variables, giving us:

$$D_{VN} = -\sum_{j=1}^{k} p_i \ln(p_i)$$

Within this section, therefore, it just remains to return to the case of ordinal variables, for which I argued earlier that spread should be the same as for nominal variables but scaled to represent the total range of the group. Therefore, I propose:

$$D_{VO} = -(x_{\max} - x_{\min}) \sum_{j=1}^{k} p_i \ln(p_i)$$

Note that I do not propose a scaling of this index to between 0 and 1 as Teachman's index itself is unbounded. This, and all other proposed indices for spread, are summarised in Table 5.5. As with the measure of spread for continuous variables, it is formed by multiplying the range by a measure of evenness; both parts are necessary to represent the notion of spread as defined earlier. Researchers should be careful to check, however, that this is not merely capturing a curvilinear (squared) effect of one or other of these original diversity indices, or an interaction effect between the two.

5.3.4 Measurement of disparity

Disparity is defined as "the extent to which there is inequality between unit members with respect to an asymmetrical variable *X*". As suggested by Harrison and Klein (2007), inequality is greatest when there is one member of the unit with the maximum possible value of *X*, and all other members have the minimum possible value. This is because the average distance between a member's own value of *X* and the largest observed value of *X* in the group is then as large as possible: if *X* represents pay, the average discrepancy between a group member's pay and the largest amount of pay in the group is the largest under this scenario. As with other types of diversity, it should be at its minimum when all members of the group have the same value (regardless of what value this is). The other desirable property to fit in with the definition is scale invariance (Allison, 1978), which is gained by comparing the ratio of the dispersion of a variable to its mean.

This definition is almost exactly the same as that provided by Harrison and Klein (2007), and therefore the two indices they recommend provide a good starting point. These two indices are the coefficient of variation (CV), which is equal to the standard deviation divided by the mean, and the Gini index, which is a scaled version of the CMD divided by the mean. The distinction between these two indices, then, is the same as the distinction between the standard deviation and the CMD – the former comparing individuals with the average position; the latter comparing differences between pairs of individuals. Unlike the definition of separation, the definition of disparity does not give such a clear steer as to which of these is more appropriate: it refers to the “inequality between unit members”, which could be captured by either method.

It is therefore worth looking in a little more detail at the difference between the two indices. Once again, I compare groups with different profiles, including those at the extreme ends of the spectrum. These are shown in Table 5.4. Group 1 has no variation at all, and therefore both indices give a value of zero. Group 2, which has maximum spread, should have relatively low disparity because the nature of high spread is that inequality is lower. Group 3, which has maximum separation, should have moderate levels of disparity, which then increase sharply in groups 4 and 5 – particularly group 5, as this is the maximum disparity possible, with huge inequality (one member holding all of the non-zero quantity).

Table 5.4: Coefficient of Variation (CV) and Gini Index for different groups

Team number	Values of group members	CV	Gini
(1)	1, 1, 1, 1, 1, 1	0	0
(2)	0, 1, 2, 3, 4, 5	0.68	0.39
(3)	0, 0, 0, 5, 5, 5	1.00	0.50
(4)	0, 0, 0, 0, 5, 5	1.41	0.67
(5)	0, 0, 0, 0, 0, 5	2.24	0.83

Here the difference between the indices is clear to see. The Gini index increases linearly through groups 3, 4, and 5, as the number of members with the highest value decreases. However, the CV increases to a greater degree for each member moving from the highest to the lowest value (this is due to the standard deviation being more sensitive to outliers – an undesirable property when measuring separation, but a desirable one for measuring disparity). This therefore better reflects the extreme disparity in the group, and I propose therefore:

$$D_{DR} = \frac{SD_x}{\bar{x}} = \frac{1}{\bar{x}}\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

However, it should be noted that this is *not* an appropriate measurement for interval variables. As discussed in Chapter 4, and described in detail by Bedeian and Mossholder (2000) and Harrison and Klein (2007), division by the mean is only meaningful when the value of the mean itself is intrinsically meaningful when related to zero: i.e. when ratio-measurement is used.

This begs the question of whether it is appropriate to measure disparity at all for a non-ratio variable. I contend that this depends on the method of calculation of the variable, and the theoretical appropriateness of the concept of disparity for that variable. Whereas calculating disparity of income or age (both ratio variables) may be theoretically meaningful due to the potential inequality created by these constructs, it may not make so much sense to calculate the disparity of job satisfaction, for example. Nevertheless, there may be interval variables for which disparity is meaningful; for example, if power or influence were measured on a Likert scale, or if a proxy were used for a variable that may be considered sensitive, such as salary. In this case, rather than dividing the standard deviation by the mean value (which is intrinsically meaningless), it would be more appropriate to divide by the difference of the mean value from the minimum possible value (which is equivalent to a mean value for ratio variables in the sense that it represents the difference from a "zero" position). Therefore I propose:

$$D_{DI} = \frac{SD_x}{(\bar{x} - x_{\min})}$$

where $x_{\min}$ is the minimum possible value of x, not the minimum observed value. For example, if relative power were measured on a scale from 1 to 7, the theoretical minimum would be 1, and this would be subtracted from the observed mean value before the standard deviation is divided by it. In cases where there is no theoretical minimum value, the minimum observed value *in all units combined* would make an adequate replacement; however, if this is used, values cannot be compared across studies.

Disparity is even less likely to be an appropriate index for ordinal variables than for interval variables. Since disparity measures the extent of inequality, one of the prerequisites for the variable being measured is that differences in its scores are meaningful. By definition, ordinal variables do not necessarily possess this property: the difference between 1 and 2 on a Likert scale, for example, could be greater than the difference between 2 and 5 in terms of the meaning of the underlying construct. The property of scale invariance that is desirable for a measure of disparity cannot be applied to ordinal variables. Therefore, unless it can be considered reasonable to make the assumption that the ordinal scale is approximately interval in nature (which is a strong assumption that would need to be defended explicitly), it would not be possible to calculate it for an ordinal variable at all. Therefore the quantity D_{DO} is left undefined.

For similar reasons, it is not appropriate to measure the disparity of a nominal variable. Because the definition of disparity relies on an asymmetrical continuum, it is impossible to measure the inequality of values that do not lie on such a continuum. Therefore the quantity D_{DN} is also undefined.

However, it is possible to conceive of a situation where a number of categories reduced to two distinct groups, with one being considered "greater" than the other in some quantity, e.g. power or status. An example might be job groups being classified as management or non-management (N.B. this could be the case for either nominal or ordinal variables). In this situation, disparity could be calculated using D_{DR} (the coefficient of variation), subject to the values of the two groups being 0 and 1, with the one with greater status having value 1. (N.B. The sometimes used measure for binary variables, % in one category, would provide the equivalent to the CMD – in that it does not assume a particular asymmetry between the categories.)

A summary of the indices proposed for measuring disparity, together with those for separation and variety (range and spread), is shown in Table 5.5.

Table 5.5: Summary of proposed indices for measurement of diversity

	Type of variable			
	Ratio	Interval	Ordinal	Nominal
Separation	$D_{SR} = D_{SI} = \frac{4}{N(N-1)}\sum_{i=1}^{N} ix_i - \frac{2(N+1)}{(N-1)}\bar{x}$ (Coefficient of Mean Difference)		$D_{SO} = 1 - \frac{4}{k-1}\sum_{i=1}^{k-1}\left(F_i - \frac{1}{2}\right)^2$	$D_{SN} = \frac{n}{n-1}\left(1 - \sum_{j=1}^{k} {p_j}^2\right)$ (Adjusted Blau's index)
Variety: range	$D_{RR} = D_{RI} = D_{RO} = x_{\max} - x_{\min}$ (range)			D_{RN} = Number of distinct categories represented
Variety: spread	$D_{VR} = D_{VI} = \frac{(x_{\max} - x_{\min})^2}{(n-1)(X_{\max} - X_{\min}).\max(x_i - x_{i-1})}$ (if $x_{\max} - x_{\min} > 0; D_{VR} = D_{VI} = 0$ otherwise)		$D_{VO} = -(x_{\max} - x_{\min})\sum_{j=1}^{k} p_i \ln{(p_i)}$	$D_{VN} = -\sum_{j=1}^{k} p_i \ln{(p_i)}$ (Teachman's index)
Disparity	$D_{DR} = \frac{SD_x}{\bar{x}}$ (Coefficient of variation)	$D_{DI} = \frac{SD_x}{(\bar{x} - x_{\min})}$	Not applicable	Not applicable

5.4 Nominal variables with distances

As noted in the previous section, although nominal variables are by their nature without quantitative value, there may still be some quantifiable differences between them. The previous example was one where categories may fall into two distinct broader groups which differed in terms of some definable quantity (e.g. status). But, even if this were not the case, it may be that the differences between the original categories were not thought to be quite equal.

The formula Blau's index (and hence D_{SN}) is based on the premise that categories of nominal variables are equally different from one another. This is not necessarily the case. For example, cultural diversity is often measured by applying Blau's index to nationality. Yet some pairs of nations are clearly more culturally similar (e.g. USA, Canada) than others (e.g. UK, China). In these cases, it makes less sense to attribute all distances between different nations as 1. Similarly, for other variables, the decision as to whether individuals belong to the same category or not may depend on the measurement scale – for example, occupational group amongst hospital staff may be measured broadly as doctors, nurses, administrators and others; alternatively a finer breakdown (which separates, for example, surgeons from oncologists) may be used. In the latter categorisation, a surgeon is likely to be considered more similar to an oncologist (as both have medical training) than to non-medical staff; however, this similarity would be missed completely using Blau's index.

This leaves the intriguing possibility of an index in which more precise "distances" between categories are taken into account. For separation, this measure would be very similar to that of D_{SR} above, but the distance between any pair of individuals is reliant on the distance between the categories they belong to. Possible methods for calculating these distances have been discussed by other authors, including

Greenberg (1956), who applied linguistic difference ratings to measure linguistic diversity, and Dawson and Brodbeck (2005), who suggested the use of GLOBE data (House et al., 2004) to measure cultural diversity. The index would be:

$$D_{Sdist} = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j<i} dist(x_i, x_j)$$

where *dist* (x_i, x_j) represents the distance between the categories of members *i* and *j*. In the Dawson and Brodbeck (2005) example these were calculated as the Euclidean distance between the nine GLOBE dimensions for each pair of countries in the study, but could be garnered from any sort of rating deemed appropriate. This is based on the same idea as the mean Euclidean distance as a diversity measure; the difference being that such a calculation usually takes data directly from group members' values on one or more continuous variables, whereas this proposed approach takes such distances from another source of data and applies them to nominal data collected from group members.

There is also an analogue of range for nominal variables which have distances available between them. This is somewhat more difficult to calculate (and it may be the case that the extra effort is not worthwhile). It first requires that a representation of the *k* categories be made in *n*-dimensional space, where $1 < n < k$. It may be that the distances were based on *n* dimensions in the first place (as with the cultural distances example above), in which case this is easy to achieve; otherwise, this can be approximated using multi-dimensional scaling. It then requires a calculation of the area (if $n = 2$), volume (if $n = 3$), or *n*-dimensional equivalent (if $n > 3$) between the location of the points represented by the group. This is computationally difficult and therefore is recommended only when there are extreme variations in the distances between categories. Likewise, spread could be measured as the product of separation and range, as D_{SO} suggests it should be for interval variables. The potential formulae for these indices are complex and not given here, as they are beyond the scope of this typology, but are left as a possibility for future research.

5.5 Hypotheses

The underlying hypothesis for the proposed research is that the diversity indices described better measure diversity than other indices used in the literature, and therefore more strongly predict outcomes. This would suggest the following hypotheses:

$H1_a$: Separation is measured more accurately by D_{SR}, D_{SI}, D_{SO} and D_{SN} for ratio, interval, ordinal and nominal variables respectively than by any other index;

$H1_b$: The indices D_{SR}, D_{SI}, D_{SO} and D_{SN} better predict outcomes of separation for ratio, interval, ordinal and nominal variables respectively than do any other index;

$H2_a$: Range is measured more accurately by D_{RR}, D_{RI}, D_{RO} and D_{RN} for ratio, interval, ordinal and nominal variables respectively than by any other index;

$H2_b$: The indices D_{RR}, D_{RI}, D_{RO} and D_{RN} better predict outcomes of range for ratio, interval, ordinal and nominal variables respectively than do any other index;

$H3_a$: Spread is measured more accurately by D_{VR}, D_{VI}, D_{VO} and D_{VN} for ratio, interval, ordinal and nominal variables respectively than by any other index;

$H3_b$: The indices D_{VR}, D_{VI}, D_{VO} and D_{VN} better predict outcomes of spread for ratio, interval, ordinal and nominal variables respectively than do any other index;

$H4_a$: Disparity is measured more accurately by D_{DR} and D_{DI} for ratio and interval variables respectively than by any other index.

Although these are hypotheses that are suggested by the typology, they are not hypotheses that can be tested directly, as doing so would require making non-confirmable assumptions about the correct indices for separation, range, spread and disparity, and the outcomes most associated with them. Therefore the "a" hypotheses will be tested on simulated data with the assumption that the typology is correct, and the tests will be that other indices do not measure the same properties as well. The "b" hypotheses will draw on assumptions from the literature about how each type of diversity is likely to be linked to certain outcomes, and will test this using real data. Although this does not provide complete tests of the hypotheses as stated, it is as close as it is possible to come because of the intractable confound between assumptions about measurement and assumptions about relationships (e.g. it would not be possible to use only real data to test $H1_a$ because it is impossible to know what the "true" value of diversity within a group is). These methods are discussed at greater length in chapter 6, with the rationale for the approach given greater attention in section 6.1 in particular.

It may be expected that there is also a $H4_b$ (The indices D_{DR} and D_{DI} better predict outcomes of disparity for ratio and interval variables respectively than do any other index). Although this would complete the implied matrix of hypotheses, it is not likely to be testable because of the lack of theoretical and empirical evidence for outcomes of disparity in a work group diversity context. Therefore this is left as a hypothesis to be considered in the future.

5.6 Chapter summary

In this chapter I have presented an extended typology of work group diversity, and in particular have derived what I consider to be the most appropriate indices for measuring each type of diversity with each type of data. I have then proposed hypotheses which state that these indices should better measure diversity than alternative indices (either other indices in the typology, or others used in the literature).

I now move on to test these hypotheses, and in the next chapter I describe my approach to doing this, and the methods I have used.

CHAPTER 6

METHODS

6.0 Chapter introduction

In this chapter I describe the methods used for testing the hypotheses stated in Chapter 5, which allude to the relative merits of the indices I have proposed for each type of diversity. First I outline the rationale for my approach, in particular the need for both simulated and real data. Next I give details of the simulation study: the design, the data generation techniques, and the methods of comparison used. Finally I give details of the real data analysis, including the sample and measures used in each of eight data sets, the relationships that might be expected to produce significant results according to the literature, and the methods used for comparing indices.

6.1 Rationale for methodological approach

As the nature of the hypotheses is of the quantitative measurement of group diversity, the need for quantitative methods is self-evident. However, there is an underlying fallacy that prevents the direct testing of hypotheses such as these: their truth depends on the existence of underlying relationships, which are based on theories that are not universally accepted. Therefore, any rejection of the hypotheses may not be due to the incorrect specification of the diversity indices, but could be due to the diversity theories themselves being wrong.

To minimise the potential effects of incorrect theoretical specification, I have employed a dual approach towards data analysis:

(1) Monte Carlo simulations were used to generate large numbers of data sets with known relationships between group diversity (as measured by each of the

indices in the typology in Chapter 5) and an outcome variable; these data sets are analysed with diversity correctly specified and incorrectly specified (i.e. measured by other indices), to show the effects of mis-specification of diversity type and use of inappropriate indices. This is used to test hypotheses $H1_a$, $H2_a$, $H3_a$ and $H4_a$.

(2) A number of real data sets were analysed, so that theoretically expected relationships are tested with the hypothesised indices, and other indices used in the literature; the expectation is that the hypothesised indices will, on average, produce stronger relationships than any other indices. This tests hypotheses $H1_b$, $H2_b$, and $H3_b$.

These two approaches are described in more detail in the following sections.

6.2 Monte Carlo simulations

There are, in total, 10 separate indices specified in section 3.4 (summarised in Table 6.1), across 14 different conditions (four variable types, four diversity types, with two combinations not meaningful). The following analysis strategy was adopted for each in turn.

For teams of size 3, 6, 10 and 20 (representing the range of team sizes commonly seen in management research), samples of 50, 100, 250 and 1,000 teams, and underlying correlations of 0.10, 0.30 and 0.50 (representing Cohen's (1988) small, medium and large effect sizes) and 0.00 – 64 different conditions in total – 100 data sets were simulated. For each of these, group diversity was calculated using the correct index and other incorrect indices as specified in Table 6.1 (which are based on those used in the literature and possible mis-specification of diversity type; they include all relevant indices as presented in Chapter 4[2]). The size of the correlation, and its statistical significance, is recorded in each case, enabling the calculation of both the average bias and the reduction in power for each possible mis-specification.

In order to cover the different types of data that may be found by researchers, different underlying data distributions were used. For ratio variables, data with underlying normal and uniform distributions, as well as data with a heavy skew (generated by a Beta distribution with parameters 0.8 and 2.0), were simulated. For interval data, the only difference between this and ratio data was its boundedness and arbitrary values, so a random Beta distribution was used to generate this on a 1-5 scale. For ordinal variables, the same three underlying distributions as for ratio variables were used, but with observed values being assigned to the nearest integer. For nominal variables, four distributions were used: two with binary variables (with a 50/50 and an 80/20 probability split respectively), one with four categories (with a 40/30/20/10 split) and one with ten categories (with a 20/15/15/10/10/10/5/5/5/5 split).

In this way, a total of 236,800 data sets were generated (6400 for each of the 37 combinations of variable and data type described above), with a total of 808,080,000 data points included. A total of 2,137,600 effect sizes were calculated and significance tests performed (nine for each of the data sets relating to diversity of ratio variables, ten for each for interval variables, 14 for each for ordinal variables, and five for each for nominal variables). The resulting data set includes summaries for 2,368 different conditions (14 diversity indices, four sample sizes, four team sizes, four effect sizes, and between one and four underlying data types), with one correct specification and between four and thirteen misspecifications for each. These data are then analysed to

---

[2] Note that r_{wg} and r_{wg}* are not included here, as they are principally used for justifying aggregation, which requires the (arbitrary) specification of a "null" distribution; without this, for a single item, it is a linear transformation of the variance, which is already included. a_{wg} is included, but only for ordinal variables as this is what it was designed for. Incorrect indices for ordinal variables include those used for continuous variables, as these are often misused as such in the literature.

determine whether changes in power and effect size bias due to misspecification are affected by size of team, sample size, and effect size (or combinations thereof): as such, the effects of using incorrect diversity indices – both overall and under different conditions – are determined.

Table 6.1: Diversity indices to be tested for each type of data

Variable type	Indices to be tested	Total no.
Ratio	D_{SR}, D_{RR} (range), D_{VR}, D_{DR} (Coefficient of variation), Standard deviation, AD_M, Variance, Euclidean distance, Gini index	9
Interval	D_{SI}, D_{RI} (range), D_{VI}, D_{DI}, Standard deviation, D_{DR} (Coefficient of variation), AD_M, Variance, Euclidean distance, Gini index	10
Ordinal	D_{SO}, D_{RO} (range), D_{VO}, l, LOV, D_{SI}, D_{DI}, Standard deviation, D_{DR} (Coefficient of variation), AD_M, Variance, a_{wg}, Euclidean distance, Gini index	14
Nominal	D_{SN} (adjusted Blau's index), D_{RN} (count), D_{VN} (Teachman's index), Blau's index (unadjusted), IQV	5

6.3 Analysis of real data sets

Eight real data sets were identified which include a variety of individual attributes and potential outcomes of diversity. All eight data sets were collected using self-completion questionnaires, supplemented by other sources of outcome data in most cases. The study of work group diversity was not the primary purpose of the collection of any of these, but was a secondary objective in all cases. The relevant sections of the data sets are described here, and the analysis conducted in section 6.4.

6.3.1 Health care team data sets

Three separate data sets, which were collected as part of a Department of Health-funded study to study the effectiveness of health care teams in different contexts (Borrill et al., 2001), were used. These include data on primary health care teams (PHCTs), community mental health teams (CMHTs), and breast cancer care teams (BCCTs).

For the PHCTs, databases of general practices (synonymous with PHCTs) were obtained from 19 health authorities in England, and 300 teams randomly selected. Letters explaining the wider research project were sent to the senior GP partner, senior health visitor and practice manager in each team, and follow-up phone calls made. In the end 133 teams agreed to take part, and paper questionnaires were sent out with reply-paid envelopes to the members of these teams. In 33 of these teams response rates were either nil or very low, so they were excluded from the sample. The final usable data set comprises 1137 individuals in 98 teams (a response rate of 54%).

For the CMHTs, chief executives of 101 community mental health trusts in four English regions were approached, and asked to encourage participation by individual teams. 58 of them provided details of CMHTs that they managed. Direct contact was made with 162 CMHTs, inviting participation in the questionnaire study. 113 teams agreed, and the final data set included 1446 individuals in 113 teams (a 75% response rate).

For the BCCT study, a sample of 113 such teams (of 190 across England, as listed in the Cancer Relief Macmillan Directory – The Macmillan Directory, 1996) was approached by sending invitations to participate to lead breast clinicians. 72 agreed to participate, and provided information about their teams, including details of the members of the team belonging to each clinical discipline. Questionnaires were then sent out to these members, and data were collected from 548 individuals in 72 teams (a 77% response rate).

Although the data sets are from different types of teams, much of the content of the questionnaires was common to all three types. Demographic data collected in each case includes age and sex; functional data includes occupational group and team tenure (occupational group measured in a salient way for the respective type of team, with categories as listed in Table 6.2). Summary statistics for these variables for the three data sets are shown in Table 6.2.

Table 6.2: Background characteristics of health care teams data sets

	Type of variable	PHCTs	CMHTs	BCCTs
Number of teams		98	113	72
Number of individuals		1137	1446	548
Size of teams		Mean 11.6, SD 7.1, range 3-38	Mean 12.8, SD 6.1, range 3-38	Mean 7.3, SD 2.5, range 3-14
Gender	Nominal	85% female, 15% male	67% female, 33% male	54% female, 46% male
Age	Ratio	Mean 42.7, SD 8.9, range 16-69	Mean 40.0, SD 8.4, range 19-63	Mean 45.7, SD 8.0, range 25-65
Occupational group	Nominal	8% GPs, 15% practice nurses, 11% district nurses, 3% midwives, 8% health visitors, 1% CPNs, 5% allied health professionals, 8% managers, 35% clerical	6% psychiatrists, 37% CPNs, 15% social workers, 7% occupational therapists, 4% psychologists, 6% support workers, 13% admin, 12% other	21% breast surgeons, 19% radiologists, 13% clinical oncologists, 3% medical oncologists, 21% histopathologists, 21% breast care nurses, 1% other
Team tenure	Ratio	Mean 7.7 years, SD 6.4 years, range 0.0-35.0 years	Mean 3.1 years, SD 3.0 years, range 0.1-30.0 years	Mean 5.9 years, SD 4.6 years, range 0.1-33.5 years

Outcome data used were team processes, team effectiveness and team innovation. Team processes were measured by the Team Climate Inventory (TCI; Anderson & West, 1998), and by reflexivity. The TCI includes four separate dimensions: participation, support for innovation, clarity of objectives, and emphasis on quality. Participation includes 12 items around the extent to which members feel information is freely shared, and are able to participate in decision making in the team; support for innovation includes eight items about the extent to which the team supports and encourages new ideas; clarity of objectives includes 11 items about the extent to which the team has clear, shared and achievable objectives; and emphasis on quality includes seven items about the extent to which the team has processes designed to maintain a good standard of performance and excellent outcomes. In order to capture diversity over an interval variable, climate strength (based on the overall TCI score: an average of all four scales) was also studied. Reflexivity was measured by Swift and West's (1998) eight item measure.

Team effectiveness was measured via self-report scales for all three data sets, with additional external ratings for PHCTs. The self-report measures were developed specifically for each type of team using a series of workshops with experts in the field (predominantly health care workers in their respective areas). The methods used were based on the ProMES system (Productivity Measurement and Enhancement System; Pritchard et al., 1988) which helps team members and stakeholders to derive and weight the outcome measures that are important for their specific setting. As a result, the measure for the PHCTs comprised 21 items, the CMHT measure 27 items and the BCCT measure 31 items. Full details of the procedures and measures are reported in Borrill et al. (2001) (for the PHCTs and CMHTs), and Haward et al. (2003) (for the BCCTs). The external ratings for PHCTs were made by staff in the local health authorities who were familiar with the team's work, but not part of the team themselves. They used the same 21 items as the self-report measures.

For team innovation, individual respondents were asked as part of the questionnaire to list changes or innovations that had taken place in their team within the previous 12 months. These were subsequently collated by team, anonymised and given to up to three expert raters (health service professionals who were familiar with a large number of teams and services), who rated each team for novelty, impact, radicalness and magnitude of innovations. Team innovation was also measured by a 5-item self-report scale (Borrill et al., 2001). Reliability data for these scales are shown in Table 6.3. A more in-depth description of all the teams and studies can be found in Borrill et al. (2001) and Haward et al. (2003).

Table 6.3: Reliability (Cronbach's alpha) of scales in health care teams data sets

	PHCTs	CMHTs	BCCTs
TCI: Participation	0.93	0.93	0.92
TCI: Support for innovation	0.90	0.91	0.90
TCI: Clarity of objectives	0.94	0.92	0.91
TCI: Emphasis on quality	0.90	0.88	0.88
TCI: Overall	0.96	0.96	0.96
Team effectiveness (self-report)	0.94	0.95	0.93
Team effectiveness (external ratings)	0.94	n/a	n/a
Innovation (self-report)	0.94	0.93	0.91
Innovation (external ratings)	0.91	0.91	0.82

6.3.2 Top management teams

As part of a study of the links between human resource management, staff attitudes and climate, and performance of UK manufacturing organisations, data were collected from top management teams. Companies were identified from sector databases, Chambers of Commerce and Trade Associations, and organisations from four sectors approached: engineering, plastics and rubber, electronics and electrical engineering, and food and drink. A representative sample of 111 companies was composed for the wider project, and in 67 of these, the CEO agreed that the top management team could be included in a survey. 388 members of these 67 top management teams responded to a postal questionnaire (a 68% response rate) – further details can be found in van Knippenberg et al. (2011).

Individual attributes collected include age, sex, educational level, team tenure, and functional background of team members: these are summarised in Table 6.4.

Table 6.4: Background characteristics of top management team data set

	Type of variable	Characteristics
Number of teams		67
Number of individuals		388
Size of teams		Mean 5.8, SD 2.1, range 3-13
Gender	Nominal	95% male, 5% female
Age	Ratio	Mean 44.6, SD 9.0, range 25-66
Educational level	Ordinal	1% No formal qualifications, 8% O Levels or equivalent, 7% A Levels or equivalent, 26% HNC/HND, 32% undergraduate degree, 26% postgraduate degree
Functional background	Nominal	20% management, 17% finance, 37% engineering, 11% production, 13% marketing/sales, 3% HRM
Team tenure	Ratio	Mean 7.7 years, SD 6.4 years, range 0.0-35.0 years

Outcome measures included team processes and organisational performance and innovation. Team processes were measured by the Team Climate Inventory (Anderson & West, 1998, as described in section 6.3.1). Cronbach's alphas were 0.90 for participation, 0.87 for support for innovation, and 0.91 for both clarity of objectives and emphasis on quality. In order to capture diversity over an interval variable, climate strength (based on the overall TCI score) was also studied (Cronbach's alpha = 0.95). Organisational performance was captured from financial returns and other publicly recorded data, and was measured by productivity (logarithm of value added per employee, standardised by industrial sector and retail price index), and profitability (logarithm of profit per employee, standardised by industrial sector and retail price index) (Nickell and van Reenen, 2001). Innovation was measured via a two-stage process. First, a detailed questionnaire was sent to at least one senior manager (selected by the CEO), who would answer questions about changes that had occurred in five areas of the organisation within the last two years – products, production techniques, production processes, work organisation and human resource management. The responses to these questions were then blind rated by three expert raters, on a scale from 1 to 5 for each of the areas.

6.3.3 Aston Team Performance Inventory

The Aston Team Performance Inventory (ATPI; West, Markiewicz and Dawson, 2005) was used to collect data in two separate healthcare samples. The ATPI is a 100-item measure covering 18 dimensions of team inputs, team processes, leadership processes and outcomes.

The first sample included 100 hospital-based nursing teams, selected in collaboration with the Royal College of Nursing from a variety of hospitals in England, and 1326 responses were received via a paper-based questionnaire (due to the devolved survey process a precise response rate cannot be calculated). For the second sample, mental health trusts were invited to participate in a wider project studying the effectiveness of multi-professional team working in mental health, for which the ATPI was collected as part of the second stage, along with a mental health-specific effectiveness measure that was developed in the first stage. Eleven trusts agreed to participate, and these trusts selected a locality within which all community-based adult or older adult mental health teams were invited to participate in the online questionnaire. 135 teams agreed, and 1500 responses were received, giving a response rate of 68%. Demographic attributes collected include age (measured ordinally, with categories “Under 30”, “30-39”, “40-49”, “50-59” and “60 or over”), sex, and ethnic background. These are summarised in Table 6.5.

Table 6.5: Background characteristics of ATPI data sets

	Type of variable	Nursing teams	Mental Health teams
Number of teams		100	133
Number of individuals		1326	1497
Size of teams		Mean 13.3, SD 8.1, range 3-52	Mean 7.3, SD 2.5, range 3-14
Gender	Nominal	88% female, 12% male	72% female, 28% male
Age	Ordinal	15% under 30, 24% 30-39, 35% 40-49, 23% 50-59, 3% 60 or over	7% under 30, 21% 30-39, 38% 40-49, 30% 50-59, 5% 60 or over
Occupational group	Nominal	(100% nurses)	32% CPNs, 11% social workers, 7% occupational therapists, 8% psychiatrists, 5% clinical psychologists, 6% support time recovery workers, 12% admin, 5% other nurses, 1% other medical, 14% other
Ethnic background	Nominal	95% White, 3% Asian, <1% Black, 1% Mixed, <1% other	87% White, 5% Asian, 3% Black, 2% Mixed, 2% other

Potential outcomes of diversity included ten of the ATPI scales. Five of these were team processes: objectives (comprising 3 items), reflexivity (4 items), participation (7 items), task focus (6 items), and team conflict (5 items). The creativity/innovation scale was omitted from this analysis as it was ambiguous as to whether positive or negative effects would be expected. The other five were outputs: team member satisfaction (6 items), attachment (3 items), team effectiveness (3 items), inter-team relationships (5 items) and team innovation (4 items). In addition, for the mental health teams the mental-health specific effectiveness measure was included, comprising 20 items (Richardson et al., 2010). Reliabilities of these scales are reported in Table 6.6.

Table 6.6: Reliability (Cronbach's alpha) of scales in ATPI data sets

	Nursing teams	Mental Health teams
Objectives	0.79	0.85
Reflexivity	0.83	0.82
Participation	0.89	0.90
Task focus	0.81	0.80
Team conflict	0.80	0.80
Creativity/innovation	0.77	0.82
Team member satisfaction	0.89	0.90
Team member attachment	0.89	0.86
Team effectiveness (ATPI)	0.72	0.81
Inter-team relationships	0.79	0.83
Innovation	0.84	0.89
Mental health team effectiveness	n/a	0.91

6.3.4 MSc students

Questionnaire data were collected at three distinct time points (once during each term of a one-year degree) from 389 Aston Business School MSc students in 71 syndicate groups. Students were studying a range of postgraduate business-related degrees, and were arranged in these syndicate groups at the start of their degrees, and expected to work in these teams throughout the taught element of their courses. Questionnaire were administered in a traditional (paper) format, and because completion of the questionnaire was a compulsory element of the course, the response rate was 100%. Individual attributes collected include age, sex, country of birth, and first language. These are summarised in Table 6.7. Outcomes of diversity included both team processes and outcomes. The process variable used was the 7-item group mutual trust dimension from the Team Climate for Learning inventory (Brodbeck, Guillaume & Winkler, 2010; Cronbach's alpha = 0.86); outcome data include commitment to the group (three items; Ellemers, Kortekaas & Ouwerkerk, 1999; Cronbach's alpha = 0.89), attendance at group meetings, and group performance (measured by overall mark awarded).

Table 6.7: Background characteristics of MSc students data set

	Type of variable	Characteristics
Number of teams		71
Number of individuals		389
Size of teams		Mean 5.4, SD 0.8, range 4-7
Gender	Nominal	48% male, 52% female
Age	Ratio	Mean 25.1, SD 4.6, range 20-50
Country of birth	Nominal	34% UK, 21% China, 12% Greece, 8% India, 4% Taiwan, 3% France, 2% Thailand, 2% Hong Kong, 1% Pakistan, 1% Nigeria, 29 other countries at <1% each
First language	Nominal	34% English, 26% Chinese, 13% Greek, 5% Hindi, 3% French, 2% Thai, 2% Punjabi, 2% Spanish, 2% Urdu, 1% Russian, 27 other languages at <1% each

6.3.5 Business Game students

Data were collected from 341 upper-level undergraduate students, from a range of business and management degrees, enrolled in a business game (BG) module. The module involved the students completing a complex and realistic computer-based simulation whereby a team of students formed the board of a European car company. The students were assigned at random into 65 work groups, each with between 3 and 7 members, which stayed together for the 10 week duration of the course. Groups met at least once a week for three hours to develop, present and write a report on a business plan. Because completion of the questionnaire was a compulsory element of the course, the response rate was 100%. Individual attributes collected include age, sex, country of birth and ethnic background. These are summarised in Table 6.8.

Potential outcomes of diversity included one group process variable: mutual trust from the Team Climate for Learning inventory (Brodbeck et al., 2010; 7 items; Cronbach's alpha = 0.86), overall group performance on the game itself (scored as an academic mark that counted towards the students' degrees), and absenteeism from group meetings.

Table 6.8: Background characteristics of Business Game data set

	Type of variable	Characteristics
Number of teams		65
Number of individuals		341
Size of teams		Mean 5.3, SD 0.8, range 3-7
Gender	Nominal	47% male, 53% female
Age	Ratio	Mean 20.0, SD 2.1, range 18-45
Country of birth	Nominal	88% UK, 4% China, 20 other countries at <1% each
Ethnic background	Nominal	61% White, 29% Asian, 2% Black, 6% Chinese, 2% Other

6.4 Analysis conducted with real data sets

Because of the inconclusive nature of results from the work group diversity literature, analysis was limited to those relationships where the evidence is more consistent. It is relatively easy to distinguish between separation and variety (spread or range) in this way because of the differing theoretical perspectives and direction of results, but it is more difficult to distinguish between spread and range.

Separation is most commonly associated with the social categorisation perspective of diversity. The most conclusive results from the literature on separation suggest that diversity of non-task related variables is associated with poorer group processes (including more conflict) and performance. Additionally, climate strength has been considered as a measure of separation (Harrison & Klein, 2007) and so this is also included. The relationships in the identified data sets that are most likely to be attenuated by incorrect measurement of separation are therefore:

Health care teams:	Diversity of age (ratio) and sex (nominal) predicting five team processes and team effectiveness (measured as self-report in all three data sets, and additionally by external raters in PHCTs); also climate strength (interval) predicting team effectiveness [42 relationships in total]
Top management teams:	Diversity of age (ratio) and sex (nominal) predicting four team processes and organisational productivity and profitability; also climate strength (interval) predicting productivity and profitability [14 relationships in total]
ATPI teams:	Diversity of age (ordinal), ethnic background (nominal) and sex (nominal) predicting five group processes (besides creativity) and four outcomes (other than innovation), with an additional effectiveness outcome for mental health teams [57 relationships in total]
MSc students:	Diversity of age (ratio), sex, first language and country of birth (all nominal) predicting mutual trust (at three time points), commitment to the group (two time points), attendance and group performance (three time points

each) [44 relationships in total]

BG students: Diversity of age (ratio), sex, ethnic background and country of birth (all nominal) predicting mutual trust (three time points), performance (one time point) and absenteeism (three time points) [28 relationships in total]

In total, these data sets account for a total of 185 different relationships, allowing a good sample size to test whether these are, indeed, significantly stronger when separation is measured with the hypothesised index than with other indices.

Variety (incorporating both range and spread) is more commonly associated with the information/decision-making perspective of diversity. The most conclusive results from the literature on variety suggest that diversity of task-related variables is associated with improved creativity and innovation. The relationships in the identified data sets that are most likely to be attenuated by incorrect measurement of variety are therefore:

Health care teams: Diversity of tenure (ratio) and occupational group (nominal) predicting team innovation (measured both as one self-report dimension and four externally-rated dimensions) [30 relationships in total]

Top management teams: Diversity of team tenure (ratio), educational background (ordinal) and functional background (nominal) predicting five dimensions of organisational innovation [15 relationships in total]

ATPI teams: For mental health teams only, diversity of occupational group predicting team innovation [1 relationship]

MSc students: None

BG students: None

Although a smaller data set than for separation, this still provides a total of 46 different relationships to test whether these are, indeed, significantly stronger when variety is measured with the hypothesised index than with other indices.

As the literature makes no firm conclusions about the differences between range and spread, hypotheses $H2_b$ and $H3_b$ are both be tested in this way: any differences found between results for range and spread are treated as exploratory findings to be analysed further in the future.

In each case, regression analysis is used to test the relationship between diversity and the relevant outcome, controlling for group size (and for climate level when climate strength is the diversity variable – this is considered good practice in climate strength research, see e.g. Gonzalez-Romá, Peiró & Tordera., 2002), and the effect size taken for further analysis will be the standardised regression coefficient. Pairwise t-tests and repeated measures ANOVA are then used to compare the effect sizes generated with the "correct" (hypothesised) diversity index, and other possible indices that are used for the measurement.

6.5 Chapter summary

I have described in this chapter the detailed methods that were chosen to test the hypotheses, including selection and analysis of both simulated and real data. In the following two chapters, I describe the results of the analyses conducted: Chapter 7 contains the results of the analysis using simulated data, and Chapter 8 the results using the real data sets. These are then brought together and discussed in Chapter 9.

CHAPTER 7

RESULTS OF SIMULATION STUDY

7.0 Chapter introduction

In this chapter I describe the results of the simulation study comparing different diversity indices' relationship with outcomes. For each type of diversity, and each data type within that, between 5 and 14 diversity indices were compared across up to 256 different conditions (with 100 replications in each cell). The effect sizes produced by each index, and the statistical significance, were recorded and compared, and analysed to determine how much support was available for hypotheses 1a, 2a, 3a and 4a.

7.1 Structure of chapter and analysis

For each of the conditions specified in section 6.2, each of the various diversity indices identified was correlated with the simulated outcome data, and the effect size and significance level of these correlations recorded. The method of comparison of indices used in this chapter is to examine the relative effect sizes – i.e. observing for what indices and under what conditions there appears to be a systematic bias in the correlation estimated – and the estimated power, i.e. the proportion of times a result was found to be significant.

Many of the same indices were tested across the different types of diversity. Therefore, to avoid confusion, the chapter is structured with a main section for each type of diversity (separation, range, spread and disparity) in turn. These sections – 7.2, 7.3, 7.4 and 7.5 – examine hypotheses 1a, 2a, 3a and 4a respectively. Within each of these sections, sub-sections will look at the four types of variable (ratio, interval, ordinal, nominal). With each of these subsections, I display a table that summarises the estimated effect size and power for each index in turn. As there are up to 256 different conditions for each index within each variable, I do not attempt to display the data for each, but instead give an overall average for each of the four actual effect sizes, and then summarise for each of the other factors (team size, sample size, data type) for the example effect size of 0.30. Patterns were similar for other effect sizes in each case.

7.2 Measurement of separation: Hypothesis 1a

7.2.1 Separation for ratio variables

Table 7.2.1 shows the average effect size and estimated power for each of the nine diversity indices identified as possible for ratio variables. The index used to generate the data was D_{SR}, and it can be seen that the average estimated effect size for this index is always exactly as predicted.

There is a wide variety in the estimated effect size when other indices are used, however. The standard deviation and Euclidean distance come very close to matching the performance of D_{SR}, both in effect size and power. This is somewhat reassuring given these are the two indices recommended by Harrison and Klein (2007) for measuring separation, but also unsurprising as the formulae are related to that for D_{SR}. The variance and AD_M indices are not far behind, and the decreases in effect size and power for these indices are also minimal. However, other indices are not so good. The range (D_{RR}) does not perform so badly, with around 15% attenuation of effect size; the D_{VR} (measuring spread) is somewhat worse, with approximately 33% attenuation of effect size overall and substantial decreases in power. But the coefficient of variation (called D_{DR} in my typology) performs very badly indeed, with actual effect sizes of 0.50 being estimated at only 0.16 – a 68% decrease (that holds for other effect sizes too). Even worse is the Gini index, which actually estimates negative effect sizes on average, and as such would lead to a lot of incorrect conclusions being drawn, particularly as 38% of those for an actual effect size of 0.50 were statistically significant. So it would appear that mis-specifying separation as spread could produce misleading results, but mis-specifying it as disparity could give even less correct conclusions.

Team size and sample size do not greatly affect the estimated effect sizes for most indices, even though the power is obviously affected by sample size. However, team size can lead to different estimates for some of the indices. In particular, the use of D_{RR} (range), D_{VR} (spread) and D_{DR} (coefficient of variation) with larger team sizes will lead to smaller effect sizes and power if D_{SR} is the correct index. Therefore the mis-specification of separation as another type of diversity is more worrying when the team sizes are larger.

Likewise, most indices are not greatly affected by the underlying data distribution, with some notable exceptions. The D_{VR} (spread) index leads to greater attenuation of effect sizes when data are uniform. The coefficient of variation is hopeless (with zero effect sizes) when underlying data are normal. The same is true for the Gini index, but having skewed data leads to the negative effect size estimates mentioned previously.

Table 7.2.1: Separation for ratio variables – Average effect size (and power) for different indices

	D_{SR}**	D_{RR} (range)	D_{VR}	D_{DR} (CV)	SD	AD_M	Variance	Euclidean	Gini
Effect size = 0.5	0.50 (0.99)	0.44 (0.96)	0.33 (0.85)	0.16 (0.48)	0.49 (0.99)	0.48 (0.99)	0.48 (0.99)	0.49 (0.99)	-0.12 (0.38)
Effect size = 0.3	0.30 (0.86)	0.26 (0.80)	0.20 (0.65)	0.09 (0.34)	0.30 (0.86)	0.29 (0.85)	0.29 (0.85)	0.30 (0.86)	-0.07 (0.27)
Effect size = 0.1	0.10 (0.38)	0.09 (0.32)	0.07 (0.24)	0.03 (0.12)	0.10 (0.37)	0.10 (0.37)	0.10 (0.37)	0.10 (0.37)	-0.02 (0.10)
Effect size = 0	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)
Team size = 3	0.30 (0.86)	0.30 (0.86)	0.27 (0.82)	0.13 (0.45)	0.30 (0.86)	0.30 (0.86)	0.29 (0.85)	0.30 (0.86)	-0.06 (0.25)
Team size = 6	0.30 (0.87)	0.28 (0.83)	0.22 (0.71)	0.09 (0.36)	0.30 (0.86)	0.29 (0.85)	0.29 (0.85)	0.30 (0.86)	-0.08 (0.28)
Team size = 10	0.30 (0.86)	0.26 (0.79)	0.18 (0.62)	0.09 (0.31)	0.30 (0.86)	0.29 (0.86)	0.29 (0.86)	0.30 (0.86)	-0.07 (0.27)
Team size = 20	0.30 (0.86)	0.21 (0.70)	0.13 (0.46)	0.07 (0.26)	0.29 (0.85)	0.29 (0.85)	0.29 (0.85)	0.30 (0.86)	-0.07 (0.28)
Sample size = 50	0.30 (0.58)	0.26 (0.46)	0.20 (0.32)	0.09 (0.14)	0.29 (0.57)	0.29 (0.56)	0.29 (0.56)	0.30 (0.57)	-0.07 (0.12)
Sample size = 100	0.30 (0.87)	0.26 (0.75)	0.20 (0.53)	0.09 (0.24)	0.30 (0.86)	0.29 (0.85)	0.29 (0.85)	0.30 (0.86)	-0.07 (0.21)
Sample size = 250	0.30 (1.00)	0.26 (0.98)	0.20 (0.81)	0.10 (0.42)	0.30 (1.00)	0.29 (1.00)	0.29 (1.00)	0.30 (1.00)	-0.07 (0.32)
Sample size = 1000	0.30 (1.00)	0.26 (1.00)	0.20 (0.94)	0.09 (0.57)	0.30 (1.00)	0.29 (1.00)	0.29 (1.00)	0.30 (1.00)	-0.07 (0.43)
Data – normal	0.30 (0.87)	0.27 (0.82)	0.23 (0.74)	0.00 (0.05)	0.30 (0.86)	0.29 (0.86)	0.29 (0.85)	0.30 (0.86)	0.00 (0.05)
Data – uniform	0.30 (0.86)	0.25 (0.77)	0.15 (0.51)	0.19 (0.66)	0.30 (0.86)	0.29 (0.85)	0.29 (0.85)	0.30 (0.86)	-0.03 (0.12)
Data – heavy skew	0.30 (0.86)	0.26 (0.80)	0.21 (0.70)	0.09 (0.32)	0.30 (0.85)	0.29 (0.86)	0.29 (0.85)	0.30 (0.85)	-0.19 (0.64)

Figures in table represent mean effect size found with simulated data (figures in parentheses indicate estimated power, i.e. proportion of significant effects found)
** indicates “correct” index, i.e. that used to generate the data
For differing team size, sample size and data type, only those simulations with effect size 0.30 are shown

Table 7.2.2: Separation for interval variables – Average effect size (and power) for different indices

	D_{SI}**	D_{RI} (range)	D_{VI}	D_{DI}	SD	CV	AD_M	Variance	Euclidean	Gini
Effect size = 0.5	0.50 (0.99)	0.43 (0.97)	0.34 (0.89)	0.36 (0.92)	0.49 (0.99)	0.43 (0.96)	0.49 (0.99)	0.48 (0.99)	0.49 (0.99)	0.07 (0.66)
Effect size = 0.3	0.30 (0.87)	0.26 (0.80)	0.21 (0.69)	0.21 (0.71)	0.30 (0.86)	0.26 (0.81)	0.29 (0.86)	0.29 (0.86)	0.30 (0.87)	0.01 (0.47)
Effect size = 0.1	0.10 (0.38)	0.09 (0.31)	0.07 (0.23)	0.05 (0.19)	0.10 (0.37)	0.08 (0.26)	0.10 (0.37)	0.10 (0.37)	0.10 (0.37)	-0.02 (0.15)
Effect size = 0	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.04)	0.00 (0.06)
Team size = 3	0.30 (0.87)	0.30 (0.87)	0.28 (0.85)	0.23 (0.72)	0.30 (0.87)	0.28 (0.83)	0.30 (0.87)	0.29 (0.87)	0.30 (0.87)	-0.04 (0.24)
Team size = 6	0.31 (0.88)	0.28 (0.85)	0.21 (0.74)	0.17 (0.69)	0.30 (0.87)	0.24 (0.82)	0.30 (0.88)	0.29 (0.87)	0.30 (0.88)	0.03 (0.67)
Team size = 10	0.30 (0.87)	0.25 (0.79)	0.18 (0.64)	0.23 (0.76)	0.30 (0.86)	0.27 (0.83)	0.29 (0.85)	0.29 (0.87)	0.30 (0.87)	0.02 (0.50)
Team size = 20	0.29 (0.85)	0.21 (0.69)	0.16 (0.55)	0.20 (0.69)	0.29 (0.84)	0.26 (0.76)	0.29 (0.84)	0.29 (0.84)	0.29 (0.85)	0.03 (0.46)
Sample size = 50	0.30 (0.60)	0.26 (0.44)	0.21 (0.32)	0.14 (0.21)	0.30 (0.58)	0.25 (0.42)	0.30 (0.60)	0.29 (0.58)	0.30 (0.60)	-0.14 (0.23)
Sample size = 100	0.30 (0.87)	0.26 (0.77)	0.21 (0.57)	0.27 (0.75)	0.30 (0.87)	0.28 (0.82)	0.29 (0.85)	0.29 (0.87)	0.30 (0.87)	0.09 (0.28)
Sample size = 250	0.30 (1.00)	0.26 (0.98)	0.21 (0.89)	0.27 (0.99)	0.30 (1.00)	0.29 (1.00)	0.29 (1.00)	0.29 (1.00)	0.30 (1.00)	0.10 (0.46)
Sample size = 1000	0.30 (1.00)	0.26 (1.00)	0.19 (1.00)	0.16 (0.91)	0.29 (1.00)	0.24 (0.99)	0.29 (1.00)	0.29 (1.00)	0.30 (1.00)	-0.02 (0.90)

Figures in table represent mean effect size found with simulated data (figures in parentheses indicate estimated power, i.e. proportion of significant effects found)
** indicates “correct” index, i.e. that used to generate the data
For differing team size and sample size, only those simulations with effect size 0.30 are shown

Table 7.2.3: Separation for ordinal variables – Average effect size (and power) for different indices

	D_{SO}**	D_{RO} (range)	D_{VO}	*l*	*LOV*	D_{SI}	D_{DI}	SD	CV
Effect size = 0.5	0.50 (0.99)	0.40 (0.92)	0.40 (0.93)	0.49 (0.99)	0.48 (0.99)	0.50 (0.99)	0.22 (0.69)	0.49 (0.99)	0.37 (0.92)
Effect size = 0.3	0.30 (0.86)	0.24 (0.75)	0.24 (0.75)	0.30 (0.85)	0.29 (0.84)	0.30 (0.86)	0.14 (0.49)	0.29 (0.85)	0.23 (0.73)
Effect size = 0.1	0.10 (0.38)	0.08 (0.29)	0.08 (0.29)	0.10 (0.37)	0.10 (0.36)	0.10 (0.38)	0.04 (0.15)	0.10 (0.37)	0.08 (0.27)
Effect size = 0	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)
Team size = 3	0.30 (0.85)	0.30 (0.85)	0.27 (0.82)	0.29 (0.84)	0.30 (0.85)	0.30 (0.85)		0.30 (0.84)	0.26 (0.79)
Team size = 6	0.30 (0.87)	0.27 (0.83)	0.26 (0.80)	0.30 (0.87)	0.29 (0.85)	0.30 (0.87)	0.21 (0.48)	0.29 (0.86)	0.23 (0.75)
Team size = 10	0.30 (0.86)	0.23 (0.74)	0.23 (0.74)	0.30 (0.85)	0.29 (0.84)	0.30 (0.86)	0.18 (0.50)	0.30 (0.85)	0.22 (0.70)
Team size = 20	0.30 (0.86)	0.16 (0.58)	0.20 (0.65)	0.30 (0.85)	0.28 (0.84)	0.30 (0.86)	0.09 (0.51)	0.29 (0.85)	0.20 (0.68)
Sample size = 50	0.29 (0.57)	0.24 (0.42)	0.24 (0.41)	0.29 (0.55)	0.28 (0.53)	0.29 (0.57)	0.14 (0.20)	0.29 (0.55)	0.23 (0.36)
Sample size = 100	0.30 (0.86)	0.24 (0.67)	0.24 (0.69)	0.30 (0.86)	0.29 (0.84)	0.30 (0.86)	0.15 (0.36)	0.30 (0.85)	0.23 (0.63)
Sample size = 250	0.30 (1.00)	0.25 (0.92)	0.25 (0.92)	0.30 (1.00)	0.29 (1.00)	0.30 (1.00)	0.14 (0.58)	0.30 (1.00)	0.23 (0.92)
Sample size = 1000	0.30 (1.00)	0.24 (0.98)	0.24 (0.99)	0.30 (1.00)	0.29 (1.00)	0.30 (1.00)	0.14 (0.84)	0.30 (1.00)	0.23 (1.00)
Data – normal	0.30 (0.86)	0.25 (0.79)	0.26 (0.81)	0.30 (0.86)	0.29 (0.83)	0.30 (0.86)	0.22 (0.74)	0.30 (0.85)	0.26 (0.81)
Data – uniform	0.30 (0.86)	0.21 (0.67)	0.20 (0.65)	0.29 (0.85)	0.29 (0.85)	0.30 (0.86)	0.14 (0.53)	0.29 (0.85)	0.22 (0.72)
Data – heavy skew	0.30 (0.86)	0.26 (0.79)	0.26 (0.80)	0.30 (0.85)	0.29 (0.85)	0.30 (0.86)	-0.08 (0.21)	0.29 (0.84)	0.20 (0.66)

(table continues on next page)

Table 7.2.3 (continued)

	AD_M	Variance	a_{wg}	Euclidean	Gini
Effect size = 0.5	0.48 (0.99)	0.48 (0.99)	-0.45 (0.97)	0.49 (0.99)	0.11 (0.54)
Effect size = 0.3	0.29 (0.84)	0.29 (0.84)	-0.27 (0.80)	0.30 (0.85)	0.07 (0.37)
Effect size = 0.1	0.10 (0.36)	0.10 (0.36)	-0.09 (0.32)	0.10 (0.37)	0.02 (0.11)
Effect size = 0	0.00 (0.04)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)
Team size = 3	0.29 (0.84)	0.28 (0.84)		0.30 (0.84)	0.13 (0.49)
Team size = 6	0.29 (0.85)	0.29 (0.85)	-0.29 (0.81)	0.30 (0.86)	0.08 (0.30)
Team size = 10	0.29 (0.84)	0.29 (0.85)	-0.28 (0.81)	0.30 (0.86)	0.05 (0.31)
Team size = 20	0.28 (0.84)	0.29 (0.85)	-0.26 (0.79)	0.29 (0.85)	0.01 (0.38)
Sample size = 50	0.28 (0.53)	0.28 (0.54)	-0.27 (0.47)	0.29 (0.56)	0.06 (0.14)
Sample size = 100	0.29 (0.84)	0.29 (0.84)	-0.28 (0.75)	0.30 (0.86)	0.07 (0.23)
Sample size = 250	0.29 (1.00)	0.29 (1.00)	-0.28 (0.98)	0.30 (1.00)	0.07 (0.42)
Sample size = 1000	0.29 (1.00)	0.29 (1.00)	-0.27 (1.00)	0.30 (1.00)	0.07 (0.69)
Data – normal	0.29 (0.84)	0.29 (0.85)	-0.29 (0.85)	0.30 (0.85)	0.17 (0.60)
Data – uniform	0.29 (0.84)	0.29 (0.84)	-0.28 (0.82)	0.29 (0.85)	0.06 (0.20)
Data – heavy skew	0.29 (0.84)	0.29 (0.84)	-0.22 (0.73)	0.30 (0.85)	-0.02 (0.31)

Figures in table represent mean effect size found with simulated data (figures in parentheses indicate estimated power, i.e. proportion of significant effects found)
** indicates "correct" index, i.e. that used to generate the data
For differing team size, sample size and data type, only those simulations with effect size 0.30 are shown

Table 7.2.4: Separation for nominal variables – Average effect size (and power) for different indices

	D_{SN}**	D_{RN} (Count)	D_{VN} (Teachman)	Blau	IQV
Effect size = 0.5	0.46 (0.96)	0.33 (0.78)	0.45 (0.95)	0.46 (0.96)	0.46 (0.96)
Effect size = 0.3	0.29 (0.84)	0.20 (0.64)	0.29 (0.83)	0.29 (0.84)	0.29 (0.84)
Effect size = 0.1	0.10 (0.38)	0.07 (0.25)	0.10 (0.37)	0.10 (0.38)	0.10 (0.38)
Effect size = 0	0.00 (0.05)	0.00 (0.04)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)
Team size = 3	0.30 (0.87)	0.30 (0.86)	0.30 (0.87)	0.30 (0.87)	0.30 (0.87)
Team size = 6	0.30 (0.85)	0.25 (0.78)	0.29 (0.84)	0.30 (0.85)	0.30 (0.85)
Team size = 10	0.29 (0.84)	0.17 (0.58)	0.28 (0.83)	0.29 (0.84)	0.29 (0.84)
Team size = 20	0.27 (0.80)	0.09 (0.34)	0.26 (0.79)	0.27 (0.80)	0.27 (0.80)
Sample size = 50	0.28 (0.51)	0.20 (0.32)	0.27 (0.50)	0.28 (0.51)	0.28 (0.51)
Sample size = 100	0.29 (0.84)	0.20 (0.56)	0.28 (0.83)	0.29 (0.84)	0.29 (0.84)
Sample size = 250	0.30 (1.00)	0.21 (0.79)	0.29 (1.00)	0.30 (1.00)	0.30 (1.00)
Sample size = 1000	0.30 (1.00)	0.21 (0.90)	0.29 (1.00)	0.30 (1.00)	0.30 (1.00)
Binary data (uneven)	0.30 (0.86)	0.22 (0.68)	0.30 (0.86)	0.30 (0.86)	0.30 (0.86)
Binary data (even)	0.29 (0.83)	0.14 (0.45)	0.28 (0.83)	0.29 (0.83)	0.29 (0.83)
4 categories	0.30 (0.86)	0.22 (0.69)	0.29 (0.84)	0.30 (0.86)	0.30 (0.86)
10 categories	0.28 (0.80)	0.24 (0.74)	0.27 (0.79)	0.28 (0.80)	0.28 (0.80)

Figures in table represent mean effect size found with simulated data (figures in parentheses indicate estimated power, i.e. proportion of significant effects found)
** indicates "correct" index, i.e. that used to generate the data
For differing team size, sample size and data type, only those simulations with effect size 0.30 are shown

7.2.2 Separation for interval variables

Table 7.2.2 shows the average effect size and estimated power for each of the ten diversity indices identified for interval variables. The index used to generate the data was DSI, and the overall average estimated effect size for this index is always exactly as predicted (with some very slight variation by team size).

There is again a wide variety in the estimated effect size when other indices are used, although not quite so wide as for ratio variables. The standard deviation, Euclidean distance and AD_M are again very close to matching the performance of D_{SI} in both effect size and power. The variance is almost as good. The coefficient of variation, although being attenuated by around 15%, performs far better than it did for ratio data, being similar to the range in its effect. This is likely to be due to the restricted range of interval data (in this simulation), meaning that there is less variety in the error created by its incorrect use. The D_{VI} and D_{DI} (spread and disparity for interval variables respectively) are more greatly affected, with decreases of around 30% in effect size and substantially less power than the correct index. Again, though, the Gini index is very poor in comparison, and sometimes leads to negative estimates where positive effects should be found.

Team size and sample size do not greatly affect the estimated effect sizes for most indices. However, team size can lead to different estimates for some of the indices. In particular, the use of D_{RI} (range) and D_{VI} (spread) with larger team sizes will lead to smaller effect sizes and power if D_{SI} is the correct index, and use of D_{DI} will lead to less stability in effect size. Therefore, as with ratio variables, the mis-specification of separation is more problematic with larger teams. The use of the Gini index appears particularly problematic when teams are very small (3 members) or when the number of teams in the sample is smaller.

7.2.3 Separation for ordinal variables

Table 7.2.3 shows the average effect size and estimated power for each of the 14 diversity indices identified as possible for ordinal variables. The index used to generate the data was DSO, and it can be seen that the average estimated effect size for this index is always exactly as predicted.

Results are not dissimilar for those found for ratio and interval variables for those indices that are the same. In particular, the standard deviation, AD_M, variance and Euclidean difference provide results that are nearly as good as the D_{SO} index (which of course was based on a cumulative frequency of levels of the ordinal variable). Additionally, the D_{SI} index (which is the hypothesised measure of separation for both ratio and interval variables) performs identically to D_{SO}, as was implied in section 5.3.1. Therefore it seems unnecessary to use a separate index for ordinal variables if the only purpose is correlation rather than description.

Measures for range and spread (D_{RO} and D_{SO}) appear to attenuate the correct effect size by about 20% with similar decreases in power, while measures of disparity (D_{DI}, the coefficient of variation, and the Gini index) perform much worse. a_{wg} is again not so bad, but effect sizes are attenuated by around 10%. Meanwhile other indices derived for ordinal variables, *l* and *LOV*, are nearly as good as D_{SO} and are on a par with the standard deviation and Euclidean distance.

As was found for ratio and interval variables, for most indices which perform well in estimation of effect size there was no significant variation in this by team size, sample size or even data type. There were substantial differences for some other indices however – in particular, range, spread and disparity indices correlated with the outcome far less when team sizes were larger. In addition, the disparity indices resulted in much smaller correlations when there was a heavy skew in the underlying data. Also, a_{wg} severely underestimated correlations when data were very skewed.

7.2.4 Separation for nominal variables

Table 7.2.4 shows the average effect size and estimated power for each of the five diversity indices identified as possible for nominal variables. The index used to generate the data was D_{SN} (or the adjusted Blau's index).

Even though there were five indices proposed, three of these perform identically. The difference between the adjusted Blau's index, D_{SN}, and the original Blau's index, only becomes prominent when teams in the sample are of different sizes: otherwise one index is just a linear multiple of the others, so correlations would be identical. In this simulation, teams were kept to a single size within each replication and therefore there is no differentiation between correlations using the two indices. For further information about why the distinction is important, Biemann and Kearney (2009) give examples using differing team sizes.

Similarly, the Index of Qualitative Variation (IQV) is a linear transformation of Blau's index. The multiplier, which is a function of the number of possible categories, would always remain constant for any given attribute within a sample, and therefore correlations involving the IQV would always be identical to those using Blau's index; the two differ only when describing the actual level of diversity. Therefore I consider only the three distinct indices – D_{SN}, D_{RN} and D_{VN} – going forward.

It is worth noting that even when the correct index (D_{SN}) is used, correlations are underestimated whenever team sizes are larger, sample sizes are smaller, and/or more categories are present in the underlying data. The correlations using D_{VN} (Teachman's index) are almost as high as those using D_{SN} in all cases. In general correlations are about 0.01 lower on average, and power usually within 2%. Thus the effect of mis-specifying separation as spread (i.e. using Teachman's index rather than the adjusted Blau's index) is probably minimal. The same cannot be said for using D_{RN} (count), where correlations are attenuated by around a third on average, and the effect is larger when team sizes increase. Therefore mis-specifying separation as range could have detrimental effects on the estimation and testing of correlations.

7.2.5 Separation – summary

Hypothesis 1a predicted that separation would measured more accurately by D_{SR}, D_{SI}, D_{SO} and D_{SN} for ratio, interval, ordinal and nominal variables respectively than by any other index. For ratio, interval and ordinal variables, there was a fairly consistent pattern that emerged. The hypothesised indices for separation were almost matched by two other common indices, the standard deviation and the Euclidean distance. In addition, the variance, AD_M, a_{wg}, l and LOV indices (the latter three for ordinal variables in particular) produced results that were only a little worse. In contrast, the use of measures of variety – range or spread – led to moderate attenuation of estimated effects, while indices intended for disparity (the coefficient of variation; D_{DI}, the hypothesised index of disparity for ordinal variables; and the Gini index) gave very different results with severely underestimated effects, sometimes even in the wrong direction. Overall this suggests that mis-specifying separation as range or spread, or even more so as disparity, can have a large deleterious effect on results obtained, although the use of other common indices for separation is unlikely to have a large bearing on findings.

For nominal variables the effect of mis-specifying separation as spread was less serious, with only slight underestimation in most cases. However, mis-specifying separation as range did lead to more severe underestimation of correlations.

Overall, there was some support for hypothesis 1a: mis-specification of the type of diversity would lead to much poorer results, although use of some indices instead of the hypothesised indices would only produce negligible attenuation of results.

7.3 Measurement of range: Hypothesis 2a

7.3.1 Range for ratio variables

Table 7.3.1 shows the average effect size and estimated power for each of the nine diversity indices identified as possible for ratio variables. The index used to generate the data was D_{RR} (or simply the range), and it can be seen that the average estimated effect size for this index is (almost) exactly as predicted.

There is quite a wide variety in the estimated effect size when other indices are used, however. Indices that measured separation well (D_{SR}, standard deviation, variance, Euclidean difference, AD_M) measure range less well, with a 10-20% attenuation in effect sizes common, as well as corresponding decreases in power. The index for spread performs even worse, with attenuation of around 40%, suggesting the distinction between range and spread is indeed important. Indices that were intended for measuring disparity, i.e. the coefficient of variation (D_{DR}) and the Gini index, were much worse still, with the Gini index again often estimating negative effects instead of positive ones.

Using the range leads to very stable effect sizes across different team sizes, sample sizes and data types. The same cannot be said for other indices, with larger team sizes creating attenuated estimates of effects for all other cases. There was little difference due to sample size, however, other than the obvious changes in power. There were some slight differences due to underlying data type, but these were most prominent for indices that actually measure disparity, where uniform or skewed data resulted in higher estimated effects than normal data.

7.3.2 Range for interval variables

Table 7.3.2 shows the average effect size and estimated power for each of the ten diversity indices identified as possible for interval variables. The index used to generate the data, D_{RI}, is exactly the same as that for ratio variables, and therefore it is not a surprise to see most results are very similar (the only difference being the different data generation method).

The only real difference is that the coefficient of variation actually improves its performance – as with separation, this is likely to be because of the restricted range of the variables within this simulation – while the hypothesised index for disparity of interval variables, D_{DI}, performs somewhat worse, with attenuation of effect sizes of around 20% for small teams, up to nearly 50% in larger teams. In this latter case average power was reduced from about 85% (for an effect size of 0.30) to around 57%.

Table 7.3.1: Range for ratio variables – Average effect size (and power) for different indices

	D_{SR}	D_{RR} (range)**	D_{VR}	D_{DR} (CV)	SD	AD_M	Variance	Euclidean	Gini
Effect size = 0.5	0.43 (0.96)	0.49 (0.99)	0.31 (0.82)	0.15 (0.48)	0.44 (0.96)	0.39 (0.93)	0.43 (0.96)	0.43 (0.96)	-0.11 (0.35)
Effect size = 0.3	0.26 (0.80)	0.30 (0.86)	0.19 (0.62)	0.09 (0.35)	0.27 (0.81)	0.24 (0.75)	0.26 (0.80)	0.26 (0.80)	-0.06 (0.25)
Effect size = 0.1	0.09 (0.33)	0.10 (0.38)	0.06 (0.22)	0.03 (0.11)	0.09 (0.34)	0.08 (0.29)	0.09 (0.33)	0.09 (0.33)	-0.02 (0.09)
Effect size = 0	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)
Team size = 3	0.30 (0.87)	0.30 (0.87)	0.28 (0.82)	0.12 (0.45)	0.30 (0.87)	0.30 (0.87)	0.29 (0.85)	0.30 (0.87)	-0.07 (0.27)
Team size = 6	0.28 (0.84)	0.30 (0.87)	0.22 (0.72)	0.10 (0.37)	0.29 (0.84)	0.26 (0.80)	0.28 (0.83)	0.28 (0.84)	-0.07 (0.27)
Team size = 10	0.25 (0.78)	0.30 (0.87)	0.16 (0.57)	0.08 (0.33)	0.26 (0.80)	0.22 (0.72)	0.26 (0.79)	0.26 (0.78)	-0.06 (0.24)
Team size = 20	0.21 (0.69)	0.29 (0.85)	0.09 (0.36)	0.06 (0.25)	0.22 (0.72)	0.18 (0.61)	0.22 (0.71)	0.21 (0.70)	-0.05 (0.21)
Sample size = 50	0.26 (0.47)	0.30 (0.58)	0.19 (0.30)	0.09 (0.13)	0.27 (0.49)	0.24 (0.40)	0.26 (0.46)	0.26 (0.47)	-0.07 (0.11)
Sample size = 100	0.26 (0.75)	0.30 (0.87)	0.19 (0.48)	0.09 (0.22)	0.27 (0.77)	0.24 (0.66)	0.26 (0.75)	0.26 (0.75)	-0.06 (0.17)
Sample size = 250	0.26 (0.97)	0.30 (1.00)	0.19 (0.74)	0.09 (0.40)	0.27 (0.98)	0.24 (0.93)	0.26 (0.97)	0.26 (0.97)	-0.07 (0.28)
Sample size = 1000	0.26 (1.00)	0.30 (1.00)	0.19 (0.96)	0.09 (0.65)	0.27 (1.00)	0.24 (1.00)	0.26 (1.00)	0.26 (1.00)	-0.06 (0.42)
Data – normal	0.27 (0.82)	0.30 (0.87)	0.20 (0.67)	0.00 (0.05)	0.28 (0.83)	0.25 (0.77)	0.27 (0.83)	0.27 (0.83)	0.00 (0.05)
Data – uniform	0.25 (0.76)	0.30 (0.85)	0.17 (0.57)	0.16 (0.57)	0.25 (0.76)	0.22 (0.70)	0.24 (0.75)	0.24 (0.75)	-0.03 (0.11)
Data – heavy skew	0.27 (0.81)	0.30 (0.87)	0.19 (0.62)	0.11 (0.42)	0.28 (0.83)	0.25 (0.77)	0.27 (0.81)	0.27 (0.81)	-0.16 (0.58)

Figures in table represent mean effect size found with simulated data (figures in parentheses indicate estimated power, i.e. proportion of significant effects found)
** indicates “correct” index, i.e. that used to generate the data
For differing team size, sample size and data type, only those simulations with effect size 0.30 are shown

Table 7.3.2: Range for interval variables – Average effect size (and power) for different indices

	D_{SI}	D_{RI} (range)**	D_{VI}	D_{DI}	SD	CV	AD_M	Variance	Euclidean	Gini
Effect size = 0.5	0.42 (0.95)	0.48 (0.99)	0.31 (0.79)	0.28 (0.77)	0.43 (0.96)	0.37 (0.88)	0.38 (0.91)	0.41 (0.96)	0.42 (0.95)	-0.09 (0.35)
Effect size = 0.3	0.26 (0.78)	0.30 (0.85)	0.20 (0.65)	0.19 (0.62)	0.27 (0.80)	0.24 (0.75)	0.24 (0.75)	0.26 (0.78)	0.26 (0.79)	-0.02 (0.33)
Effect size = 0.1	0.09 (0.32)	0.10 (0.40)	0.07 (0.20)	0.07 (0.23)	0.09 (0.34)	0.08 (0.31)	0.08 (0.29)	0.09 (0.33)	0.09 (0.34)	0.00 (0.10)
Effect size = 0	0.00 (0.06)	0.00 (0.06)	0.00 (0.05)	0.00 (0.06)	0.00 (0.05)	0.00 (0.06)	0.00 (0.06)	0.00 (0.05)	0.00 (0.05)	0.00 (0.06)
Team size = 3	0.30 (0.86)	0.30 (0.86)	0.27 (0.80)	0.24 (0.69)	0.30 (0.86)	0.28 (0.83)	0.30 (0.85)	0.29 (0.85)	0.30 (0.86)	-0.02 (0.26)
Team size = 6	0.28 (0.83)	0.30 (0.86)	0.22 (0.71)	0.19 (0.64)	0.29 (0.84)	0.26 (0.78)	0.26 (0.80)	0.28 (0.83)	0.28 (0.84)	-0.08 (0.42)
Team size = 10	0.25 (0.78)	0.30 (0.85)	0.18 (0.61)	0.16 (0.57)	0.27 (0.82)	0.23 (0.75)	0.23 (0.74)	0.25 (0.76)	0.26 (0.79)	-0.02 (0.36)
Team size = 20	0.21 (0.67)	0.30 (0.85)	0.12 (0.46)	0.16 (0.57)	0.23 (0.71)	0.20 (0.62)	0.18 (0.60)	0.22 (0.69)	0.22 (0.68)	0.03 (0.30)
Sample size = 50	0.25 (0.45)	0.29 (0.56)	0.19 (0.29)	0.15 (0.24)	0.26 (0.48)	0.22 (0.37)	0.23 (0.39)	0.25 (0.43)	0.26 (0.46)	-0.07 (0.13)
Sample size = 100	0.26 (0.72)	0.30 (0.85)	0.20 (0.52)	0.18 (0.43)	0.27 (0.76)	0.24 (0.66)	0.24 (0.66)	0.26 (0.72)	0.26 (0.73)	-0.02 (0.38)
Sample size = 250	0.26 (0.98)	0.30 (1.00)	0.20 (0.79)	0.21 (0.81)	0.27 (0.99)	0.24 (0.95)	0.24 (0.94)	0.26 (0.98)	0.26 (0.98)	0.00 (0.28)
Sample size = 1000	0.27 (1.00)	0.30 (1.00)	0.20 (1.00)	0.22 (1.00)	0.28 (1.00)	0.26 (1.00)	0.25 (1.00)	0.27 (1.00)	0.27 (1.00)	0.01 (0.55)

Figures in table represent mean effect size found with simulated data (figures in parentheses indicate estimated power, i.e. proportion of significant effects found)
** indicates "correct" index, i.e. that used to generate the data
For differing team size and sample size, only those simulations with effect size 0.30 are shown

Table 7.3.3: Range for ordinal variables – Average effect size (and power) for different indices

	D_{SO}	D_{RO} (range)**	D_{VO}	l	*LOV*	D_{SI}	D_{DI}	SD	CV
Effect size = 0.5	0.39 (0.90)	0.46 (0.94)	0.43 (0.93)	0.38 (0.89)	0.34 (0.84)	0.39 (0.90)	0.17 (0.52)	0.40 (0.90)	0.32 (0.84)
Effect size = 0.3	0.24 (0.73)	0.28 (0.82)	0.26 (0.78)	0.23 (0.72)	0.21 (0.67)	0.24 (0.73)	0.10 (0.37)	0.24 (0.74)	0.20 (0.65)
Effect size = 0.1	0.08 (0.29)	0.09 (0.37)	0.09 (0.33)	0.08 (0.28)	0.07 (0.24)	0.08 (0.29)	0.03 (0.12)	0.08 (0.29)	0.06 (0.23)
Effect size = 0	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.04)	0.00 (0.05)
Team size = 3	0.30 (0.86)	0.30 (0.86)	0.27 (0.82)	0.30 (0.86)	0.30 (0.86)	0.30 (0.86)	0.00 (0.48)	0.30 (0.86)	0.26 (0.80)
Team size = 6	0.27 (0.82)	0.30 (0.86)	0.27 (0.82)	0.26 (0.81)	0.23 (0.75)	0.27 (0.82)	0.17 (0.40)	0.27 (0.83)	0.22 (0.71)
Team size = 10	0.23 (0.75)	0.30 (0.86)	0.27 (0.82)	0.22 (0.73)	0.19 (0.65)	0.23 (0.75)	0.12 (0.35)	0.24 (0.76)	0.18 (0.64)
Team size = 20	0.14 (0.50)	0.23 (0.70)	0.21 (0.67)	0.14 (0.49)	0.11 (0.41)	0.14 (0.50)	0.06 (0.23)	0.15 (0.53)	0.12 (0.45)
Sample size = 50	0.23 (0.41)	0.27 (0.52)	0.25 (0.44)	0.23 (0.39)	0.20 (0.33)	0.23 (0.41)	0.08 (0.14)	0.24 (0.43)	0.19 (0.29)
Sample size = 100	0.23 (0.67)	0.27 (0.80)	0.25 (0.74)	0.23 (0.65)	0.21 (0.57)	0.23 (0.67)	0.11 (0.25)	0.24 (0.68)	0.20 (0.54)
Sample size = 250	0.24 (0.89)	0.29 (0.96)	0.26 (0.94)	0.23 (0.88)	0.21 (0.83)	0.24 (0.89)	0.11 (0.44)	0.24 (0.90)	0.20 (0.83)
Sample size = 1000	0.24 (0.97)	0.29 (1.00)	0.27 (1.00)	0.23 (0.96)	0.21 (0.95)	0.24 (0.97)	0.11 (0.63)	0.25 (0.97)	0.20 (0.95)
Data – normal	0.25 (0.78)	0.30 (0.86)	0.28 (0.84)	0.25 (0.77)	0.22 (0.70)	0.25 (0.78)	0.18 (0.65)	0.26 (0.79)	0.23 (0.73)
Data – uniform	0.20 (0.64)	0.25 (0.74)	0.21 (0.68)	0.19 (0.62)	0.17 (0.58)	0.20 (0.64)	0.06 (0.35)	0.20 (0.64)	0.15 (0.53)
Data – heavy skew	0.26 (0.78)	0.30 (0.86)	0.27 (0.82)	0.25 (0.77)	0.23 (0.73)	0.26 (0.78)	0.00 (0.09)	0.27 (0.80)	0.21 (0.69)

(table continues on next page)

Table 7.3.3 (continued)

	AD_M	Variance	a_{wg}	Euclidean	Gini
Effect size = 0.5	0.35 (0.85)	0.38 (0.89)	-0.31 (0.87)	0.39 (0.89)	0.10 (0.51)
Effect size = 0.3	0.21 (0.68)	0.23 (0.73)	-0.19 (0.70)	0.24 (0.73)	0.06 (0.33)
Effect size = 0.1	0.07 (0.25)	0.08 (0.28)	-0.06 (0.25)	0.08 (0.29)	0.02 (0.10)
Effect size = 0	0.00 (0.05)	0.00 (0.04)	0.00 (0.04)	0.00 (0.04)	0.00 (0.05)
Team size = 3	0.29 (0.85)	0.28 (0.84)	0.00 (0.81)	0.30 (0.86)	0.13 (0.47)
Team size = 6	0.24 (0.77)	0.26 (0.81)	-0.25 (0.76)	0.27 (0.82)	0.07 (0.28)
Team size = 10	0.20 (0.67)	0.23 (0.74)	-0.21 (0.72)	0.23 (0.74)	0.03 (0.28)
Team size = 20	0.12 (0.41)	0.15 (0.51)	-0.14 (0.50)	0.15 (0.51)	0.00 (0.28)
Sample size = 50	0.21 (0.34)	0.23 (0.40)	-0.19 (0.36)	0.23 (0.41)	0.05 (0.11)
Sample size = 100	0.21 (0.58)	0.23 (0.65)	-0.19 (0.60)	0.23 (0.67)	0.06 (0.18)
Sample size = 250	0.21 (0.84)	0.23 (0.89)	-0.19 (0.88)	0.24 (0.89)	0.06 (0.36)
Sample size = 1000	0.21 (0.95)	0.24 (0.96)	-0.19 (0.96)	0.24 (0.97)	0.06 (0.67)
Data – normal	0.22 (0.70)	0.25 (0.77)	-0.23 (0.77)	0.25 (0.78)	0.14 (0.52)
Data – uniform	0.18 (0.59)	0.19 (0.61)	-0.13 (0.59)	0.20 (0.63)	0.05 (0.20)
Data – heavy skew	0.23 (0.74)	0.26 (0.79)	-0.21 (0.73)	0.26 (0.79)	-0.02 (0.27)

Figures in table represent mean effect size found with simulated data (figures in parentheses indicate estimated power, i.e. proportion of significant effects found)
** indicates "correct" index, i.e. that used to generate the data
For differing team size, sample size and data type, only those simulations with effect size 0.30 are shown

Table 7.3.4: Range for nominal variables – Average effect size (and power) for different indices

	D_{SN}	D_{RN} (Count)**	D_{VN} (Teachman)	Blau	IQV
Effect size = 0.5	0.32 (0.74)	0.38 (0.79)	0.35 (0.76)	0.32 (0.74)	0.32 (0.74)
Effect size = 0.3	0.19 (0.62)	0.24 (0.69)	0.21 (0.66)	0.19 (0.62)	0.19 (0.62)
Effect size = 0.1	0.06 (0.25)	0.08 (0.32)	0.07 (0.28)	0.06 (0.25)	0.06 (0.25)
Effect size = 0	0.00 (0.04)	0.00 (0.04)	0.00 (0.04)	0.00 (0.04)	0.00 (0.04)
Team size = 3	0.30 (0.86)	0.30 (0.87)	0.30 (0.87)	0.30 (0.86)	0.30 (0.86)
Team size = 6	0.22 (0.71)	0.26 (0.76)	0.24 (0.74)	0.22 (0.71)	0.22 (0.71)
Team size = 10	0.17 (0.56)	0.22 (0.65)	0.20 (0.61)	0.17 (0.56)	0.17 (0.56)
Team size = 20	0.09 (0.34)	0.16 (0.48)	0.12 (0.42)	0.09 (0.34)	0.09 (0.34)
Sample size = 50	0.18 (0.32)	0.22 (0.42)	0.20 (0.37)	0.18 (0.32)	0.18 (0.32)
Sample size = 100	0.19 (0.54)	0.22 (0.66)	0.21 (0.60)	0.19 (0.54)	0.19 (0.54)
Sample size = 250	0.20 (0.76)	0.24 (0.81)	0.22 (0.80)	0.20 (0.76)	0.20 (0.76)
Sample size = 1000	0.20 (0.85)	0.26 (0.87)	0.23 (0.86)	0.20 (0.85)	0.20 (0.85)
Binary data (uneven)	0.19 (0.63)	0.23 (0.70)	0.21 (0.65)	0.19 (0.63)	0.19 (0.63)
Binary data (even)	0.10 (0.33)	0.11 (0.34)	0.10 (0.34)	0.10 (0.33)	0.10 (0.33)
4 categories	0.22 (0.71)	0.30 (0.86)	0.26 (0.80)	0.22 (0.71)	0.22 (0.71)
10 categories	0.26 (0.79)	0.30 (0.87)	0.28 (0.84)	0.26 (0.79)	0.26 (0.79)

Figures in table represent mean effect size found with simulated data (figures in parentheses indicate estimated power, i.e. proportion of significant effects found)
** indicates “correct” index, i.e. that used to generate the data
For differing team size, sample size and data type, only those simulations with effect size 0.30 are shown

7.3.3 Range for ordinal variables

Table 7.3.3 shows the average effect size and estimated power for each of the 14 diversity indices identified as possible for ordinal variables. Again, the index used to generate the data, D_{RO}, is exactly the same as that for ratio and interval variables, and therefore it is not a surprise to see most results are somewhat similar.

A slight difference this time, however, is that indices which more accurately capture separation (D_{SO}, D_{SI}, standard deviation, AD_M, and Euclidean distance) are slightly worse performing, with effect size attenuation usually in the 20-30% range, while the measure of spread, D_{VO}, does a little better, with attenuation of only around 12-13%. Thus although it is still important to distinguish between range and spread, the distinction between range and separation appears more substantial for ordinal variables.

In terms of the ordinal-specific indices, l is again similar to the other measures of separation, while *LOV* gives slightly greater attenuation, with around 30% of the effect size being lost when this is used instead of range. a_{wg} is slightly worse still.

Again, larger team sizes lead to greater attenuation, although this time this also applies to range itself – when team size is 20 in the simulation, effects are underestimated here too. This is likely to be because more samples will be generated with the maximum range possible in more teams, and it is therefore a form of range restriction (Cook & Campbell, 1979; Linn, 1968). In real studies this is less likely to be a problem, as it would imply that all teams in a sample could have the full range on the attribute in question – in which case this would not be an appropriate sample to study the effects of range as a type of diversity.

7.3.4 Range for nominal variables

Table 7.3.4 shows the average effect size and estimated power for each of the five diversity indices identified as possible for nominal variables. The index used to generate the data was D_{RN} (or a simple count of the number of categories). As in section 7.2.4, results for D_{SN}, Blau's index and the IQV are identical; I leave these in the tables for the sake of completeness but do not discuss the latter two indices any further.

It can be seen from the table that use of the D_{RN} index would itself lead to underestimation of the correct effect, and that this is particularly the case when team sizes are larger, sample sizes are smaller or when there were fewer categories represented (particularly for evenly spread binary data). This is obviously due to the fact that, for an attribute where only two categories are possible (e.g. sex), and particularly when the two categories are equally likely, it is highly probable that many (or indeed all) teams would have both categories represented, which would lead to no effect being estimated. In reality, the use of D_{RN} as an index, or even range as a concept, in such scenarios is theoretically unlikely. Therefore it makes most sense to consider the more likely scenarios of multinomial variables (i.e. 4 categories and 10 categories, in the last two rows of the table).

From these it can be seen that D_{VN} (Teachman's index) underestimates effects slightly, with around 6-12% attenuation in the correlation estimated and a 3-6% drop in statistical power. Thus mis-specifying range as spread would have a small effect on conclusions only. D_{SN} (adjusted Blau's index) is somewhat less good, with up to 28% attenuation of effect sizes and up to a 15% drop in power. Thus mis-specifying range as separation would have a moderate negative effect on conclusions.

7.3.5 Range – summary

Hypothesis 2a predicted that range would be measured more accurately by D_{RR}, D_{RI}, D_{RO} and D_{RN} for ratio, interval, ordinal and nominal variables respectively than by any other index. In all cases, the hypothesised index was significantly better than any other index. However, the extent to which other indices underestimated effects depended on the type of variable.

For ratio and interval variables, use of indices that capture separation had a slightly unfavourable effect, whereas the index for spread had a more damaging effect. For ordinal variables, this distinction was reversed. In all three cases, use of indices more commonly applied to disparity created huge underestimates, and for the Gini index these could even be in the opposite direction.

For nominal variables the mis-specification of range as spread would have a small detrimental effect on findings, whereas the mis-specification as separation would have a slightly more serious effect.

Thus if range is the correct type of diversity for a hypothesis, the use of any other diversity index will create attenuated effects and lower power – to a greater or lesser degree for different indices and variable types. Hypothesis 2a is therefore fully supported.

7.4 Measurement of spread: Hypothesis 3a

7.4.1 Spread for ratio variables

Table 7.4.1 shows the average effect size and estimated power for each of the nine diversity indices identified as possible for ratio variables. The index used to generate the data was D_{VR}, and it can be seen that the average estimated effect size for this index is exactly as predicted.

For all other indices, however, there is a significant underestimation of effects. Indices which capture separation and range typically result in 35-40% attenuation. For indices designed to capture disparity this is even worse, with the coefficient of variation leading to 90% attenuation of correlations, and huge decreases in power. As with separation and range, the Gini index often produces negative effect sizes.

Once more, larger team sizes are associated with greater attenuation of effects, although not for the "correct" index, D_{VR}, which is stable across all variations in team and sample size and underlying data type. Sample size does not greatly affect estimates for any index, and data type only for disparity indices, where highly skewed data leads to negative associations with the Gini index, and zero associations with the coefficient of variation (the same is true for normal data in the latter case). Uniform data usually produce slightly smaller effect sizes and power for most indices.

Table 7.4.1: Spread for ratio variables – Average effect size (and power) for different indices

	D_{SR}	D_{RR} (range)	D_{VR}**	D_{DR} (CV)	SD	AD_M	Variance	Euclidean	Gini
Effect size = 0.5	0.33 (0.85)	0.31 (0.82)	0.50 (0.99)	0.05 (0.28)	0.30 (0.80)	0.29 (0.79)	0.29 (0.79)	0.30 (0.80)	-0.11 (0.36)
Effect size = 0.3	0.20 (0.65)	0.19 (0.62)	0.30 (0.87)	0.03 (0.18)	0.18 (0.60)	0.18 (0.59)	0.17 (0.58)	0.18 (0.60)	-0.07 (0.25)
Effect size = 0.1	0.07 (0.23)	0.06 (0.21)	0.10 (0.38)	0.01 (0.07)	0.06 (0.21)	0.06 (0.20)	0.06 (0.20)	0.06 (0.21)	-0.02 (0.09)
Effect size = 0	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)
Team size = 3	0.28 (0.83)	0.28 (0.83)	0.30 (0.87)	0.10 (0.38)	0.27 (0.81)	0.25 (0.79)	0.25 (0.80)	0.26 (0.81)	-0.07 (0.26)
Team size = 6	0.22 (0.72)	0.22 (0.73)	0.30 (0.89)	0.03 (0.14)	0.20 (0.66)	0.18 (0.62)	0.19 (0.64)	0.20 (0.66)	-0.07 (0.26)
Team size = 10	0.17 (0.60)	0.16 (0.58)	0.30 (0.86)	0.00 (0.10)	0.15 (0.53)	0.15 (0.51)	0.14 (0.50)	0.15 (0.53)	-0.07 (0.25)
Team size = 20	0.13 (0.46)	0.09 (0.35)	0.30 (0.85)	-0.02 (0.10)	0.11 (0.40)	0.12 (0.42)	0.11 (0.39)	0.11 (0.41)	-0.06 (0.22)
Sample size = 50	0.20 (0.33)	0.19 (0.31)	0.30 (0.61)	0.03 (0.09)	0.18 (0.30)	0.18 (0.28)	0.18 (0.28)	0.18 (0.30)	-0.07 (0.10)
Sample size = 100	0.20 (0.52)	0.18 (0.48)	0.30 (0.86)	0.03 (0.10)	0.18 (0.45)	0.17 (0.44)	0.17 (0.43)	0.18 (0.46)	-0.06 (0.17)
Sample size = 250	0.20 (0.81)	0.19 (0.75)	0.30 (1.00)	0.03 (0.18)	0.18 (0.74)	0.18 (0.72)	0.17 (0.71)	0.18 (0.74)	-0.07 (0.30)
Sample size = 1000	0.20 (0.94)	0.19 (0.96)	0.30 (1.00)	0.03 (0.34)	0.18 (0.92)	0.17 (0.90)	0.17 (0.91)	0.18 (0.92)	-0.07 (0.42)
Data – normal	0.23 (0.74)	0.21 (0.66)	0.30 (0.86)	0.00 (0.05)	0.22 (0.71)	0.22 (0.71)	0.21 (0.70)	0.22 (0.72)	0.00 (0.05)
Data – uniform	0.15 (0.51)	0.17 (0.60)	0.30 (0.87)	0.08 (0.28)	0.13 (0.45)	0.12 (0.41)	0.12 (0.42)	0.13 (0.45)	-0.03 (0.11)
Data – heavy skew	0.21 (0.71)	0.18 (0.61)	0.30 (0.86)	0.00 (0.21)	0.19 (0.64)	0.19 (0.64)	0.18 (0.62)	0.19 (0.65)	-0.16 (0.59)

Figures in table represent mean effect size found with simulated data (figures in parentheses indicate estimated power, i.e. proportion of significant effects found)
** indicates "correct" index, i.e. that used to generate the data
For differing team size, sample size and data type, only those simulations with effect size 0.30 are shown

Table 7.4.2: Spread for interval variables – Average effect size (and power) for different indices

	D_{SI}	D_{RI} (range)	D_{VI}**	D_{DI}	SD	CV	AD_M	Variance	Euclidean	Gini
Effect size = 0.5	0.35 (0.92)	0.32 (0.85)	0.50 (0.99)	0.15 (0.61)	0.32 (0.86)	0.26 (0.76)	0.31 (0.86)	0.30 (0.83)	0.32 (0.86)	-0.09 (0.58)
Effect size = 0.3	0.21 (0.66)	0.19 (0.61)	0.31 (0.88)	0.11 (0.46)	0.19 (0.60)	0.16 (0.55)	0.18 (0.61)	0.18 (0.58)	0.19 (0.61)	-0.01 (0.36)
Effect size = 0.1	0.07 (0.25)	0.06 (0.22)	0.10 (0.39)	0.03 (0.15)	0.06 (0.22)	0.05 (0.20)	0.06 (0.22)	0.06 (0.21)	0.06 (0.23)	-0.01 (0.19)
Effect size = 0	0.00 (0.04)	0.00 (0.04)	0.00 (0.04)	0.00 (0.04)	0.00 (0.05)	0.00 (0.04)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)
Team size = 3	0.28 (0.84)	0.28 (0.84)	0.31 (0.88)	0.22 (0.75)	0.27 (0.81)	0.25 (0.80)	0.26 (0.78)	0.26 (0.80)	0.27 (0.81)	-0.02 (0.08)
Team size = 6	0.24 (0.76)	0.23 (0.74)	0.31 (0.90)	0.12 (0.45)	0.21 (0.70)	0.20 (0.65)	0.21 (0.70)	0.20 (0.66)	0.21 (0.71)	0.01 (0.56)
Team size = 10	0.15 (0.49)	0.15 (0.51)	0.31 (0.87)	0.00 (0.14)	0.12 (0.41)	0.07 (0.28)	0.12 (0.42)	0.12 (0.40)	0.13 (0.42)	-0.08 (0.52)
Team size = 20	0.16 (0.55)	0.10 (0.36)	0.30 (0.87)	0.10 (0.48)	0.14 (0.48)	0.12 (0.49)	0.15 (0.52)	0.14 (0.48)	0.15 (0.49)	0.03 (0.29)
Sample size = 50	0.19 (0.29)	0.18 (0.25)	0.32 (0.64)	0.16 (0.25)	0.16 (0.22)	0.17 (0.27)	0.16 (0.22)	0.15 (0.19)	0.16 (0.23)	0.10 (0.17)
Sample size = 100	0.24 (0.66)	0.21 (0.55)	0.31 (0.89)	0.00 (0.25)	0.22 (0.58)	0.16 (0.41)	0.22 (0.60)	0.21 (0.55)	0.22 (0.59)	-0.18 (0.49)
Sample size = 250	0.19 (0.69)	0.18 (0.65)	0.30 (1.00)	0.12 (0.55)	0.17 (0.61)	0.14 (0.60)	0.16 (0.61)	0.16 (0.60)	0.17 (0.62)	0.00 (0.21)
Sample size = 1000	0.22 (1.00)	0.21 (1.00)	0.30 (1.00)	0.16 (0.78)	0.21 (1.00)	0.18 (0.94)	0.20 (1.00)	0.20 (1.00)	0.21 (1.00)	0.02 (0.57)

Figures in table represent mean effect size found with simulated data (figures in parentheses indicate estimated power, i.e. proportion of significant effects found)
** indicates “correct” index, i.e. that used to generate the data
For differing team size and sample size, only those simulations with effect size 0.30 are shown

Table 7.4.3: Spread for ordinal variables – Average effect size (and power) for different indices

	D_{SO}	D_{RO} (range)	D_{VO}**	l	*LOV*	D_{SI}	D_{DI}	SD	CV
Effect size = 0.5	0.40 (0.93)	0.44 (0.96)	0.49 (0.98)	0.39 (0.92)	0.35 (0.88)	0.40 (0.93)	0.15 (0.56)	0.37 (0.91)	0.27 (0.78)
Effect size = 0.3	0.24 (0.76)	0.27 (0.81)	0.30 (0.85)	0.24 (0.75)	0.21 (0.70)	0.24 (0.76)	0.09 (0.38)	0.23 (0.73)	0.17 (0.58)
Effect size = 0.1	0.08 (0.30)	0.09 (0.34)	0.10 (0.38)	0.08 (0.29)	0.07 (0.26)	0.08 (0.30)	0.03 (0.12)	0.08 (0.28)	0.06 (0.19)
Effect size = 0	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)
Team size = 3	0.28 (0.83)	0.28 (0.83)	0.30 (0.86)	0.27 (0.83)	0.28 (0.83)	0.28 (0.83)	0.00 (0.39)	0.26 (0.80)	0.22 (0.72)
Team size = 6	0.26 (0.80)	0.28 (0.83)	0.30 (0.85)	0.25 (0.78)	0.22 (0.73)	0.26 (0.80)	0.19 (0.41)	0.24 (0.77)	0.18 (0.61)
Team size = 10	0.24 (0.75)	0.28 (0.82)	0.30 (0.86)	0.23 (0.74)	0.20 (0.67)	0.24 (0.75)	0.10 (0.38)	0.22 (0.72)	0.15 (0.53)
Team size = 20	0.20 (0.66)	0.25 (0.76)	0.28 (0.83)	0.19 (0.64)	0.16 (0.56)	0.20 (0.66)	0.04 (0.32)	0.19 (0.63)	0.13 (0.45)
Sample size = 50	0.24 (0.42)	0.27 (0.48)	0.29 (0.55)	0.24 (0.43)	0.21 (0.36)	0.24 (0.42)	0.07 (0.16)	0.23 (0.40)	0.17 (0.25)
Sample size = 100	0.24 (0.68)	0.26 (0.77)	0.29 (0.85)	0.23 (0.66)	0.21 (0.59)	0.24 (0.68)	0.10 (0.23)	0.22 (0.64)	0.17 (0.42)
Sample size = 250	0.24 (0.93)	0.27 (0.99)	0.30 (1.00)	0.24 (0.92)	0.22 (0.87)	0.24 (0.93)	0.09 (0.42)	0.23 (0.91)	0.17 (0.71)
Sample size = 1000	0.24 (1.00)	0.27 (1.00)	0.30 (1.00)	0.24 (0.98)	0.21 (0.98)	0.24 (1.00)	0.09 (0.70)	0.23 (0.98)	0.17 (0.93)
Data – normal	0.26 (0.81)	0.29 (0.85)	0.30 (0.87)	0.26 (0.81)	0.23 (0.75)	0.26 (0.81)	0.19 (0.65)	0.26 (0.80)	0.23 (0.74)
Data – uniform	0.19 (0.66)	0.24 (0.75)	0.28 (0.82)	0.18 (0.62)	0.17 (0.58)	0.19 (0.66)	0.04 (0.24)	0.17 (0.60)	0.12 (0.44)
Data – heavy skew	0.27 (0.81)	0.28 (0.83)	0.30 (0.86)	0.27 (0.81)	0.24 (0.77)	0.27 (0.81)	-0.08 (0.24)	0.26 (0.79)	0.16 (0.55)

(table continues on next page)

Table 7.4.3 (continued)

	AD_M	Variance	a_{wg}	Euclidean	Gini
Effect size = 0.5	0.33 (0.86)	0.36 (0.89)	-0.29 (0.83)	0.37 (0.90)	0.14 (0.58)
Effect size = 0.3	0.20 (0.66)	0.22 (0.71)	-0.18 (0.63)	0.22 (0.72)	0.09 (0.41)
Effect size = 0.1	0.07 (0.24)	0.07 (0.27)	-0.06 (0.22)	0.08 (0.27)	0.03 (0.12)
Effect size = 0	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)
Team size = 3	0.24 (0.78)	0.25 (0.78)	0.00 (0.72)	0.26 (0.80)	0.15 (0.52)
Team size = 6	0.21 (0.70)	0.23 (0.75)	-0.26 (0.67)	0.24 (0.76)	0.11 (0.38)
Team size = 10	0.19 (0.65)	0.21 (0.70)	-0.18 (0.62)	0.22 (0.71)	0.07 (0.34)
Team size = 20	0.15 (0.52)	0.19 (0.60)	-0.16 (0.52)	0.19 (0.62)	0.03 (0.39)
Sample size = 50	0.20 (0.32)	0.22 (0.38)	-0.18 (0.29)	0.23 (0.39)	0.09 (0.15)
Sample size = 100	0.20 (0.54)	0.22 (0.60)	-0.18 (0.49)	0.22 (0.62)	0.09 (0.25)
Sample size = 250	0.20 (0.83)	0.22 (0.89)	-0.18 (0.79)	0.23 (0.90)	0.09 (0.46)
Sample size = 1000	0.20 (0.96)	0.22 (0.97)	-0.18 (0.95)	0.23 (0.98)	0.09 (0.77)
Data – normal	0.22 (0.73)	0.25 (0.79)	-0.25 (0.79)	0.25 (0.79)	0.17 (0.60)
Data – uniform	0.15 (0.52)	0.16 (0.56)	-0.08 (0.48)	0.17 (0.59)	0.10 (0.37)
Data – heavy skew	0.23 (0.74)	0.25 (0.77)	-0.18 (0.62)	0.25 (0.79)	0.00 (0.26)

Figures in table represent mean effect size found with simulated data (figures in parentheses indicate estimated power, i.e. proportion of significant effects found)
** indicates “correct” index, i.e. that used to generate the data
For differing team size, sample size and data type, only those simulations with effect size 0.30 are shown

Table 7.4.4: Spread for nominal variables – Average effect size (and power) for different indices

	D_{SN}	D_{RN} (Count)	D_{VN} (Teachman)**	Blau	IQV
Effect size = 0.5	0.46 (0.96)	0.38 (0.84)	0.47 (0.96)	0.46 (0.96)	0.46 (0.96)
Effect size = 0.3	0.29 (0.84)	0.23 (0.71)	0.30 (0.85)	0.29 (0.84)	0.29 (0.84)
Effect size = 0.1	0.10 (0.38)	0.08 (0.29)	0.10 (0.39)	0.10 (0.38)	0.10 (0.38)
Effect size = 0	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)
Team size = 3	0.30 (0.87)	0.30 (0.87)	0.31 (0.88)	0.30 (0.87)	0.30 (0.87)
Team size = 6	0.30 (0.85)	0.27 (0.82)	0.30 (0.86)	0.30 (0.85)	0.30 (0.85)
Team size = 10	0.29 (0.83)	0.21 (0.66)	0.30 (0.85)	0.29 (0.83)	0.29 (0.83)
Team size = 20	0.27 (0.80)	0.14 (0.47)	0.28 (0.81)	0.27 (0.80)	0.27 (0.80)
Sample size = 50	0.29 (0.54)	0.23 (0.42)	0.29 (0.56)	0.29 (0.54)	0.29 (0.54)
Sample size = 100	0.29 (0.82)	0.23 (0.65)	0.29 (0.83)	0.29 (0.82)	0.29 (0.82)
Sample size = 250	0.29 (1.00)	0.23 (0.84)	0.30 (1.00)	0.29 (1.00)	0.29 (1.00)
Sample size = 1000	0.29 (1.00)	0.23 (0.92)	0.30 (1.00)	0.29 (1.00)	0.29 (1.00)
Binary data (uneven)	0.30 (0.87)	0.23 (0.73)	0.30 (0.87)	0.30 (0.87)	0.30 (0.87)
Binary data (even)	0.28 (0.80)	0.15 (0.47)	0.28 (0.80)	0.28 (0.80)	0.28 (0.80)
4 categories	0.29 (0.85)	0.26 (0.78)	0.30 (0.86)	0.29 (0.85)	0.29 (0.85)
10 categories	0.29 (0.84)	0.29 (0.84)	0.30 (0.86)	0.29 (0.84)	0.29 (0.84)

Figures in table represent mean effect size found with simulated data (figures in parentheses indicate estimated power, i.e. proportion of significant effects found)
** indicates “correct” index, i.e. that used to generate the data
For differing team size, sample size and data type, only those simulations with effect size 0.30 are shown

7.4.2 Spread for interval variables

Table 7.4.2 shows the average effect size and estimated power for each of the ten diversity indices identified as possible for interval variables. The index used to generate the data, D_{VI}, is exactly the same as that for ratio variables, and therefore it is not a surprise to see most results are similar.

As with range, the coefficient of variation (although poor) is less bad than the equivalent for interval variables, D_{DI}, which produces correlations attenuated by around 70%. Other results follow the same patterns as those for ratio variables, with if anything slightly greater underestimation in general.

7.4.3 Spread for ordinal variables

Table 7.3.3 shows the average effect size and estimated power for each of the 14 diversity indices identified as possible for ordinal variables. The index used to generate the data, D_{VO}, is not the same as those for ratio and interval variables, and therefore the pattern of results is a little different.

What remains the same, however, is that correlations with the correct index produce almost exactly the correct results – albeit with very slight underestimation for larger team sizes, smaller sample sizes, or uniform underlying data. Other indices always underestimate the effects to a much greater extent. The closest to D_{VO} is actually the range (or D_{RO}), where the attenuation is only around 10% and average power is only decreased slightly. Thus it would appear that the mis-specification of spread as range is less serious than the mis-specification of range as spread for ordinal variables. For those indices capturing separation, including the ordinal-specific indices, there is a consistent pattern of 20-40% attenuation and associated levels of reduced power. Once again, measures of disparity perform worst, although in this case the Gini index is little worse than the D_{DI} and does not tend to create correlations in the opposite direction.

Larger team sizes again lead to greater attenuation – although only for sizes greater than 10 for range and spread indices. Uniform underlying data results in smaller estimated effect sizes in most cases too, although for the D_{DI} index, heavily skewed data can result in negative estimates of a positive effect. For the Gini index this time the average effect with highly skewed data was zero.

7.4.4 Spread for nominal variables

Table 7.4.4 shows the average effect size and estimated power for each of the five diversity indices identified as possible for nominal variables. The index used to generate the data was D_{VN}, or Teachman's index. As in section 7.2.4 and 7.3.4, I do not consider the last two columns, as performance is identical to that of D_{SN}.

D_{VN} itself produces generally accurate estimates of the correct effect, albeit slightly less so for larger effect sizes, larger team sizes and evenly distributed binary data – as in previous sections, this is due to the possible range restriction under these conditions. The use of D_{SN} (adjusted Blau's index) leads to only slight attenuation of effect sizes and minimally decreased power, suggesting that the mis-specification of spread as separation is not very problematic (the reverse having been found true in section 7.2.4). The use of D_{RN}, or a count, instead though could lead to more serious underestimation of correlations: over 20% attenuation in many cases, with this effect being worse with larger team sizes and with binary data.

7.4.5 Spread – summary

Hypothesis 3a predicted that spread would be measured more accurately by D_{VR}, D_{VI}, D_{VO} and D_{VN} for ratio, interval, ordinal and nominal variables respectively than by any other index. In all cases, the hypothesised index was significantly better than any other index. However, the extent to which other indices underestimated effects depended on the type of variable.

For ratio and interval variables, incorrect use of separation or range indices had a similar level of effect to each other, with a moderate level of attenuation and reduction in power. For ordinal variables, the use of range instead of spread was less detrimental than the use of a separation index. This suggests that the distinction between range and spread is less salient for ordinal variables. In all three cases, use of indices more commonly applied to disparity created huge underestimates.

For nominal variables, the correct specification of separation or spread is not too detrimental to effect sizes or statistical power, with only slight attenuation found. The use of range instead of spread, however, was a bit more problematic.

Thus if spread is the correct type of diversity for a hypothesis, the use of any other diversity index will lead to biased results, but Blau's index instead of Teachman's index for nominal variables produces only slight biases.

Overall, there was clear support for hypothesis 3a, although the extent to which the use of the correct index made a difference depended on the type of data being studied, with nominal data being the least affected.

7.5 Measurement of disparity: Hypothesis 4a

7.5.1 Disparity for ratio variables

Table 7.5.1 shows the average effect size and estimated power for each of the nine diversity indices identified as possible for ratio variables. The index used to generate the data was D_{DR} (the coefficient of variation). Unlike in previous cases, the use of this correct index could lead to an overestimate of the effect size; examination of the different factors suggests that it is when data are normally distributed that this overestimation occurs, and no such problems exist for other data distributions. This is likely to be because of the higher concentration of values close to zero for normal data, leading to some instability in the results. It could be said that this is an unrealistic scenario for a variable where disparity was the construct of interest, so this should not be of great concern.

All other indices, however, greatly underestimate effects. This is least prominent with the Gini index; although this is designed to measure disparity, it still leads to high levels of attenuation, especially with larger actual effects. Correlations of 0.50 are reduced by around a third on average, while this of 0.30 are reduced by over 20%. Smaller correlations are not so affected. Every other index, however, leads to underestimations of more than 50%, and frequently around 70%. This is most pronounced for the D_{VR} index, which shrinks correlations by 90% – exactly the same level as is the case in reverse.

Team size and sample size do not have a major effect of the level of underestimation with any index besides range, where the estimates get smaller with larger team sizes (again, this is probably due to range restriction of the construct of interest). Underlying data distribution does have a substantial effect, however. The level of attenuation is generally smallest with uniform data and greatest with normally distributed data – the exception to the latter being the Gini index, where with normal data it almost matches the performance of the hypothesised index, yielding only 0.01 in the estimated effect size and 1% of power.

7.5.2 Disparity for interval variables

Table 7.5.2 shows the average effect size and estimated power for each of the ten diversity indices for interval variables. The index used to generate the data, D_{DI}, is an adjusted form of the coefficient of variation. Despite this, results are quite different (which, in itself, demonstrates why use of the coefficient of variation with data on an arbitrary scale is inappropriate).

The extent of underestimation is far less in almost all cases. For most indices the level of attenuation of the correlations is around 20-30%. The main exceptions to this are the D_{VI} (spread), which reduces effects by at least half, and the coefficient of variation itself, which has attenuation of only 10-15% in general, and only slightly reduced power. The Gini index has around the same effect sizes produced as for ratio variables, meaning that it is now outperformed by several other indices.

Team size appears to have a curvilinear effect for the incorrect indices. Generally the most attenuation is found for medium team sizes (6 and 10), with smaller and larger teams leading to greater (but still underestimated) effects. There is also a general trend that larger sample sizes produce slightly larger estimated correlations than smaller sample sizes, but this is not a substantial effect.

Table 7.5.1: Disparity for ratio variables – Average effect size (and power) for different indices

	D_{SR}	D_{RR} (range)	D_{VR}	D_{DR} (CV)**	SD	AD_M	Variance	Euclidean	Gini
Effect size = 0.5	0.16 (0.49)	0.15 (0.49)	0.05 (0.28)	0.51 (0.96)	0.18 (0.54)	0.16 (0.51)	0.17 (0.54)	0.17 (0.53)	0.34 (0.79)
Effect size = 0.3	0.10 (0.35)	0.09 (0.35)	0.03 (0.19)	0.33 (0.84)	0.11 (0.40)	0.10 (0.37)	0.11 (0.39)	0.11 (0.39)	0.23 (0.60)
Effect size = 0.1	0.03 (0.13)	0.03 (0.11)	0.01 (0.08)	0.13 (0.41)	0.03 (0.14)	0.03 (0.13)	0.03 (0.14)	0.03 (0.14)	0.10 (0.26)
Effect size = 0	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)
Team size = 3	0.13 (0.45)	0.13 (0.45)	0.10 (0.38)	0.33 (0.85)	0.13 (0.46)	0.13 (0.47)	0.12 (0.45)	0.13 (0.46)	0.18 (0.45)
Team size = 6	0.10 (0.36)	0.10 (0.37)	0.03 (0.14)	0.33 (0.84)	0.11 (0.39)	0.10 (0.37)	0.11 (0.39)	0.11 (0.39)	0.23 (0.61)
Team size = 10	0.09 (0.31)	0.08 (0.32)	0.00 (0.11)	0.33 (0.85)	0.10 (0.37)	0.09 (0.33)	0.10 (0.37)	0.10 (0.37)	0.25 (0.67)
Team size = 20	0.08 (0.29)	0.07 (0.27)	-0.01 (0.11)	0.33 (0.83)	0.10 (0.36)	0.08 (0.32)	0.09 (0.36)	0.09 (0.35)	0.26 (0.69)
Sample size = 50	0.10 (0.15)	0.10 (0.14)	0.03 (0.09)	0.34 (0.61)	0.11 (0.17)	0.11 (0.16)	0.11 (0.17)	0.11 (0.17)	0.23 (0.33)
Sample size = 100	0.10 (0.25)	0.09 (0.23)	0.03 (0.11)	0.33 (0.82)	0.11 (0.28)	0.10 (0.26)	0.11 (0.27)	0.11 (0.28)	0.23 (0.47)
Sample size = 250	0.10 (0.42)	0.09 (0.40)	0.03 (0.19)	0.33 (0.95)	0.11 (0.48)	0.10 (0.44)	0.10 (0.47)	0.10 (0.47)	0.23 (0.70)
Sample size = 1000	0.09 (0.58)	0.09 (0.64)	0.03 (0.35)	0.33 (0.99)	0.11 (0.66)	0.10 (0.62)	0.10 (0.67)	0.10 (0.66)	0.23 (0.91)
Data – normal	0.00 (0.05)	0.00 (0.06)	0.00 (0.05)	0.39 (0.78)	0.00 (0.05)	0.00 (0.05)	0.00 (0.06)	0.00 (0.05)	0.38 (0.77)
Data – uniform	0.20 (0.67)	0.17 (0.59)	0.08 (0.29)	0.30 (0.87)	0.20 (0.69)	0.19 (0.67)	0.20 (0.68)	0.20 (0.68)	0.19 (0.62)
Data – heavy skew	0.09 (0.33)	0.11 (0.42)	0.00 (0.21)	0.30 (0.88)	0.12 (0.45)	0.11 (0.39)	0.12 (0.44)	0.12 (0.44)	0.12 (0.42)

Figures in table represent mean effect size found with simulated data (figures in parentheses indicate estimated power, i.e. proportion of significant effects found)
** indicates "correct" index, i.e. that used to generate the data
For differing team size, sample size and data type, only those simulations with effect size 0.30 are shown

Table 7.5.2: Disparity for interval variables – Average effect size (and power) for different indices

	D_{SI}	D_{RI} (range)	D_{VI}	D_{DI}**	SD	CV	AD_M	Variance	Euclidean	Gini
Effect size = 0.5	0.37 (0.84)	0.34 (0.83)	0.23 (0.74)	0.50 (0.99)	0.38 (0.88)	0.46 (0.95)	0.36 (0.86)	0.37 (0.88)	0.38 (0.87)	0.34 (0.84)
Effect size = 0.3	0.15 (0.56)	0.16 (0.57)	0.07 (0.39)	0.30 (0.86)	0.17 (0.60)	0.25 (0.74)	0.16 (0.57)	0.17 (0.59)	0.17 (0.59)	0.15 (0.53)
Effect size = 0.1	0.07 (0.25)	0.06 (0.23)	0.04 (0.14)	0.10 (0.38)	0.07 (0.25)	0.09 (0.33)	0.07 (0.25)	0.07 (0.26)	0.07 (0.25)	0.06 (0.18)
Effect size = 0	0.00 (0.05)	0.00 (0.06)	0.00 (0.05)	0.00 (0.06)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)
Team size = 3	0.20 (0.68)	0.20 (0.68)	0.17 (0.58)	0.30 (0.86)	0.20 (0.69)	0.26 (0.77)	0.21 (0.69)	0.19 (0.66)	0.20 (0.69)	0.10 (0.38)
Team size = 6	0.12 (0.49)	0.13 (0.49)	0.03 (0.39)	0.30 (0.88)	0.14 (0.51)	0.23 (0.64)	0.12 (0.49)	0.14 (0.50)	0.14 (0.51)	0.15 (0.57)
Team size = 10	0.12 (0.55)	0.13 (0.59)	0.01 (0.29)	0.30 (0.84)	0.15 (0.63)	0.25 (0.77)	0.13 (0.56)	0.14 (0.61)	0.14 (0.61)	0.16 (0.63)
Team size = 20	0.17 (0.54)	0.16 (0.53)	0.06 (0.31)	0.30 (0.87)	0.19 (0.58)	0.26 (0.79)	0.17 (0.53)	0.19 (0.57)	0.19 (0.56)	0.17 (0.54)
Sample size = 50	0.08 (0.15)	0.09 (0.10)	-0.02 (0.08)	0.31 (0.62)	0.11 (0.15)	0.22 (0.35)	0.09 (0.15)	0.10 (0.15)	0.10 (0.16)	0.15 (0.22)
Sample size = 100	0.18 (0.49)	0.17 (0.42)	0.09 (0.25)	0.29 (0.83)	0.19 (0.52)	0.25 (0.68)	0.18 (0.48)	0.19 (0.48)	0.19 (0.50)	0.14 (0.29)
Sample size = 250	0.16 (0.72)	0.18 (0.80)	0.09 (0.46)	0.30 (1.00)	0.18 (0.76)	0.25 (0.93)	0.16 (0.71)	0.18 (0.74)	0.18 (0.74)	0.15 (0.61)
Sample size = 1000	0.19 (0.89)	0.19 (0.97)	0.11 (0.78)	0.30 (1.00)	0.21 (0.98)	0.27 (1.00)	0.19 (0.93)	0.20 (0.97)	0.20 (0.97)	0.15 (0.99)

Figures in table represent mean effect size found with simulated data (figures in parentheses indicate estimated power, i.e. proportion of significant effects found)
** indicates "correct" index, i.e. that used to generate the data
For differing team size and sample size, only those simulations with effect size 0.30 are shown

7.5.3 Disparity – summary

Hypothesis 4a predicted that disparity would be measured more accurately by D_{DR} and D_{DI} for ratio and interval variables respectively than by any other index. For both ratio and nominal variables, the hypothesised index outperforms all other indices. However, with ratio data this effect is severe, whereas for interval data it is more moderate. In both cases, though, it is clear that the misspecification of disparity as any other type of diversity can lead to severely biased results and a lack of power for testing effects. Thus hypothesis 4a is clearly supported.

7.6 Chapter summary

Overall, the results suggest that there are usually clear differences between types of diversity, even if not between all indices. For ratio and interval variables, D_{SN}, standard deviation, AD_M, variance and Euclidean distance all appear to give adequate measurements of separation. However, indices that are intended for measuring range or spread often give very different results, suggesting the specification of diversity type is important. This is even clearer with indices intended for disparity, which can give highly erroneous results if used inappropriately.

For ordinal variables, many of the same patterns held true, with those indices derived specifically for ordinal variables performing reasonably as measures of separation. The distinction between range and spread was less important for ordinal variables, however.

For nominal variables, there were clear differences between range and the other two diversity types applicable (separation and spread), although the indices hypothesised for separation and spread, i.e. adjusted Blau's index and Teachman's index, performed very similarly to each other, indicating that the difference was not so salient for nominal data.

In terms of the hypotheses tested by these simulations, they were all supported in part or in full. Hypothesis 1a predicted that separation would measured more accurately by D_{SR}, D_{SI}, D_{SO} and D_{SN} for ratio, interval, ordinal and nominal variables respectively than by any other index. In reality, for ratio, interval and ordinal variables the hypothesised indices certainly measured separation more accurately than indices for range, spread or disparity; however, some other common indices that were not part of the typology measured it almost as well – particularly the standard deviation, variance, AD_M and Euclidean distance, where there was relatively little to choose between these indices. For nominal variables there was only a slight advantage of D_{SN} over Teachman's index (D_{VN}), but a significant advantage over D_{RN}.

Hypothesis 2a predicted that range would be measured more accurately by D_{RR}, D_{RI}, D_{RO} and D_{RN} for ratio, interval, ordinal and nominal variables respectively than by any other index. This was wholeheartedly supported for all types of variable.

Hypothesis 3a predicted that spread would be measured more accurately by D_{VR}, D_{VI}, D_{VO} and D_{VN} for ratio, interval, ordinal and nominal variables respectively than by any other index. This was clearly supported for ratio, interval and ordinal variables, but for nominal variables there was only a slight advantage of the hypothesised index, D_{VN} (Teachman's index) over the adjusted Blau's index (D_{SN}).

Hypothesis 4a predicted that disparity would be measured more accurately by D_{DR} and D_{DI} for ratio and interval variables respectively than by any other index. This was clearly the case for ratio variables. For interval variables it was also true that correlations were strongest with D_{DI} used as the disparity index; however, the relatively strong performance of D_{DR} (the coefficient of variation) in this case does call into question the wisdom of trying to measure disparity for interval variables.

Having tested the relative merits of the indices with simulated data, and found in general that the typology is supported, I now move to the more complex issue of real data. In Chapter 8 I test the remaining hypotheses (1b, 2b and 3b) with eight real team data sets to determine whether the proposed indices are more strongly related to actual team outcomes than other indices.

CHAPTER 8

RESULTS FROM ANALYSIS OF REAL DATA

8.0 Chapter introduction

In this chapter I describe the results of analysis of eight real data sets, testing 239 possible relationships by comparing the results when using different diversity indices. These relationships included some with dependent variables hypothesised to be outcomes of separation (with negative relationships suggested by the literature), and others hypothesised to be outcomes of variety, whether range or spread (with positive relationships suggested by the literature). Relationships were compared both in terms of average effect size and statistical significance.

8.1 Structure of chapter and analysis

The eight real data sets, described in Chapter 6, were analysed by testing each of 239 possible relationships between diversity and outcomes, with the same range of indices as used in Chapter 7. I address these in separate sections, dividing each into continuous, ordinal and nominal variables for the diversity attributes. Section 8.2 studies measurement of separation, and therefore examines hypothesis 1b. Section 8.3 studies measurement of variety, incorporating both range and spread, and therefore examines hypotheses 2b and 3b.

In keeping with results found in the literature, there was a very wide range of effects found, both positive and negative, and as a result many of the averaged effects are close to zero with relatively few results found to be significant; this is not unusual for a summary of results somewhat akin to a meta-analysis, and tallies with the existing meta-analyses of group diversity described earlier. Pairwise non-parametric (Wilcoxon) tests are therefore used to examine the significance of one index versus another. Also, in one data set (the MSc students), very large correlations in the unexpected (positive) direction were found for age diversity. There are various reasons why this might be, but the relative range restriction of age in this data might be one cause of that. In any case, these results were found to be so out of keeping with the rest of the data that I removed them from the analysis so that they would not exert any undue influence on the overall results.

Tables in this chapter show summaries across types of attribute and diversity. Full tables showing all effect sizes considered are shown in appendix 4. Correlations between diversity indices for four example variables from the Top Management Teams data set are shown in appendix 5 (this data set was chosen because it included all four data types).

8.2 Relationships involving separation: Hypothesis 1b

8.2.1 Separation for continuous variables

The hypothesised indices for separation for ratio and interval variables, D_{SR} and D_{SI} were identical in formula: therefore I consider the comparison between this index and other possible indices together. Table 8.2.1 shows the average effect size (standardised regression coefficient) and estimated power (proportion of effects that were statistically significant) for each index applied to age diversity and climate strength. In total, there were 32 different relationships tested for age diversity (with age as a ratio variable) across the data sets, and 25 tested for climate strength (which included climate level as a control variable). All analyses included group size as a control variable.

Table 8.2.1: Separation for continuous and ordinal variables – Average effect size (and power) for different indices

	Age diversity (ratio)	Climate strength (interval)	Age diversity (ordinal)
No. of effects	25	6	19
D_{SR} (D_{SI})	-0.05 (0.09)	-0.06 (0.00)	0.07 (0.26)
D_{RR} (D_{RI}, range)	0.02 (0.03)*	-0.04 (0.17)	0.07 (0.32)**
D_{VR} (D_{VI})	-0.03 (0.06)*	-0.05 (0.00)	
D_{DR} (CV)	-0.03 (0.03)	-0.02 (0.00)	0.07 (0.37)
SD	-0.03 (0.03)*	-0.04 (0.00)	*0.04 (0.16)***
AD_M	-0.05 (0.13)	-0.05 (0.00)	0.05 (0.05)
Variance	-0.03 (0.06)*	-0.02 (0.00)*	0.05 (0.05)
Euclidean distance	-0.03 (0.06)*	-0.04 (0.00)	*0.04 (0.16)***
Gini index	0.01 (0.03)	0.21 (0.17)*	0.10 (0.53)
D_{DI}		-0.01 (0.00)	0.06 (0.47)
a_{wg}			-0.05 (0.21)
D_{SO}			0.05 (0.26)
D_{VO}			0.10 (0.37)**
l			0.06 (0.21)
LOV			0.06 (0.32)**

Figures in table represent mean regression coefficient size found across relevant data sets (figures in parentheses indicate proportion of significant effects found)
D_{SR} (D_{SI}) index hypothesised to have most negative effect for ratio and interval variables, D_{SO} for ordinal variables
* indicates a significantly different ($p < .05$) average effect size from the D_{SR} (D_{SI}) index
** indicates a significantly different ($p < .05$) average effect size from the D_{SO} index
Italics indicate statistical significance is in the opposite predicted from that hypothesised.

Overall, it can be seen that the D_{SR}/D_{SI} index has the joint lowest relationship with outcomes, along with the AD_M index. This fits in with the hypothesis, which suggests the most negative relationships should occur with this index. For age diversity the AD_M index actually produced one more significant result, whereas for climate strength the D_{SI} index gave one more statistically significant result. Therefore there appears to be little to choose between these two indices.

Comparison with other indices, however, is somewhat more conclusive. For age diversity, the D_{SR} index produces consistently more negative effects than D_{RR}, D_{VR}, standard deviation, variance or Euclidean distance ($p < .05$ in all cases). The differences with the coefficient of variation and Gini index were not statistically significant however; this may suggest that disparity of age could have similar effects to those of age separation, or perhaps that the variety of effects using these indices is far greater (as seen in the results in Chapter 7). The estimated power was substantially lower in both cases though.

For climate strength, the results were not so conclusive, with only the variance and Gini index significantly different from the D_{SI} index in terms of effect sizes produced. Thus for interval variables (or climate in particular) there appears to be far less differentiation between indices than for ratio variables. It is noteworthy, however, that the Gini index (one of the three indices not to be significantly different for age separation) produced far higher effects than any other index, and that its one significant result was caused by a (theoretically counterintuitive) positive relationship with the outcome. These conclusions are based on only six relationships, however.

Overall, then, the D_{SR}/D_{SI} index gives better (more theoretically consistent) results than any other index besides the AD_M index, from which it is not very distinguishable.

8.2.2 Separation for ordinal variables

The final column of Table 8.2.1 shows the average effect size and estimated power for each diversity index applied to age diversity when it was measured ordinally (as it was in two of the data sets). There were 19 different outcome variables across these two data sets and therefore a total of 19 different relationships for each index.

Overall, it can be seen that the D_{SO} index – hypothesised to give the most negative results – is amongst those with the lowest average effects (note that a_{wg}, as a measure of agreement rather than diversity, has effects in the “opposite” direction).

However, two indices – the standard deviation and Euclidean distance – give even more negative effects on average (and significantly so, even though the discrepancy in effect size is very small indeed). These two indices are not intended for use with ordinal variables, even though they are sometimes (inappropriately) employed. When comparing with other indices that are specifically intended for ordinal variables, a different pattern emerges: D_{RO} (range), D_{VO} (spread) and Leik's (1966) *LOV* index all result in significantly higher effect sizes than D_{SO}; a_{wg} and Blair and Lacy's (1996) *l* measure perform very similarly to D_{SO}, with marginally higher effect sizes and lower estimated power.

From this it would appear that D_{SO} is amongst the best measures of separation for ordinal variables; if the choice is confined to those specifically designed for ordinal variables then it seems to be the best, although some measures designed for continuous variables appear to work just as well in this case.

8.2.3 Separation for nominal variables

Table 8.2.2 shows the average effect size and estimated power for each nominal diversity index across the eight data sets. In total, there were 62 different outcomes for sex diversity, 26 for ethnic diversity, 18 for nationality diversity and 11 for language diversity; all 117 effects are summarised in the final column of the table.

As noted in Chapter 7, the effects for Blau's index and the IQV are actually identical (but are left in for the sake of completeness). Unlike in Chapter 7, however, D_{SN} and Blau's index do not always give the same results, as there was some variation in the sizes of teams within each data set.

Table 8.2.2: Separation for nominal variables – Average effect size (and power) for different indices

	Sex diversity	Ethnic diversity	Nationality diversity	Language diversity	Overall
No. of effects	62	26	18	11	117
D_{SN}	-0.01 (0.05)	-0.09 (0.12)	-0.17 (0.28)	-0.17 (0.45)	-0.07 (0.14)
D_{RN} (range)	-0.01 (0.08)	-0.02 (0.00)*	-0.14 (0.17)	-0.09 (0.09)*	-0.04 (0.08)*
D_{VN}	0.00 (0.10)*	-0.06 (0.04)*	-0.16 (0.28)	-0.14 (0.27)*	-0.05 (0.13)*
Blau	0.00 (0.08)*	-0.08 (0.08)	-0.17 (0.28)	*-0.18 (0.45)**	-0.06 (0.15)*
IQV	0.00 (0.08)*	-0.08 (0.08)	-0.17 (0.28)	*-0.18 (0.45)**	-0.06 (0.15)*

Figures in table represent mean regression coefficient size found across relevant data sets (figures in parentheses indicate proportion of significant effects found)
D_{SN} index hypothesised to have most negative effect
* indicates a significantly different ($p < .05$) average effect size from the D_{SN} index
Italics indicate statistical significance is in the opposite predicted from that hypothesised.

Overall, it can be seen that D_{SN} – hypothesised to have the most negative effects – does so, although the differences are small, particularly between D_{SN} and Blau's index. Indeed, Blau's index actually produces one more significant result than D_{SN} overall. However, the difference between D_{SN}, D_{RN} and D_{VN} is slightly clearer. D_{VN} produces less negative results for sex diversity, ethnic diversity and language diversity. D_{RN} produces less negative results for ethnic diversity and language diversity. There were no significant differences amongst the 18 relationships involving nationality diversity. However, the totality of results suggests that the hypothesised differences between D_{SN} and other indices are supported, even if the differences are sometimes small.

8.2.4 Separation – summary

For continuous (ratio and interval) variables, the hypothesised D_{SR}/D_{SI} index was found to give more negative (and therefore theoretically consistent) results than any other index besides the AD_M index. There was very little to distinguish between the D_{SR}/D_{SI} index and AD_M. Thus hypothesis 1b, which included the prediction that the D_{SR}/D_{SI} index would be the best predictor for continuous variables, was largely supported. For ordinal variables the hypothesised index D_{SO} was the best of the ordinal-specific indices, although there was evidence of some indices designed for continuous variables being equally appropriate. For nominal variables, the hypothesised index D_{SN} (the adjusted Blau's index) was a more negative predictor than other possible indices, albeit with small differences. Overall, then, hypothesis 1b (which stated that the indices D_{SR}, D_{SI}, D_{SO} and D_{SN} better predict outcomes of separation for ratio, interval, ordinal and nominal variables respectively than do any other index) is largely supported by the data.

8.3 Relationships involving variety (range and spread): Hypotheses 2b and 3b

8.3.1 Variety for continuous variables

As with separation, the hypothesised indices for both range (D_{RR} and D_{RI}) and spread (D_{VR} and D_{VI}) were identical in formula for both ratio and interval variables; however, in my samples there were no examples of interval variables where variety would be the appropriate form of diversity to use. Therefore I study both range and spread of tenure as forms of variety expected to have positive relationships with outcomes; existing theory would not predict how these would differ from each other, and so I compare each with all other indices.

Table 8.3.1 shows the average effect size (standardised regression coefficient) and estimated power (proportion of effects that were statistically significant) for each index applied to tenure diversity. In total, there were 20 different relationships tested. Again, all analyses included group size as a control variable.

Table 8.3.1: Variety for continuous and ordinal variables – Average effect size (and power) for different indices

	Tenure diversity (ratio)	Educational diversity (ordinal)
No. of effects	20	5
D_{SR} (D_{SI})	0.01 (0.05)*	0.08 (0.00)**
D_{RR} (D_{RI}, range)	0.09 (0.10)**	0.12 (0.00)**
D_{VR} (D_{VI})	-0.01 (0.15)*	
D_{DR} (CV)	0.05 (0.00)*,**	0.12 (0.00)**
SD	0.04 (0.05)*	0.11 (0.00)**
AD_M	0.03 (0.05)*	0.10 (0.00)
Variance	0.03 (0.05)*	0.10 (0.00)**
Euclidean distance	0.04 (0.05)*	0.11 (0.00)**
Gini index	*0.17 (0.25)*,***	*0.28 (0.40)**
D_{DI}		0.12 (0.00)**
a_{wg}		-0.09 (0.00)
D_{SO}		0.10 (0.00)
D_{VO}		0.18 (0.40)*
l		0.10 (0.00)
LOV		0.11 (0.00)

Figures in table represent mean regression coefficient size found across relevant data sets (figures in parentheses indicate proportion of significant effects found)
D_{RR} (D_{RI}) and D_{VR} (D_{VI}) index hypothesised to have most positive effect
* indicates a significantly different ($p < .05$) average effect size from the D_{RR} (D_{RI}) or D_{RO} index
** indicates a significantly different ($p < .05$) average effect size from the D_{VR} (D_{VI}) or D_{VO} index
Italics indicate statistical significance is in the opposite predicted from that hypothesised.

It can be seen that the Gini index actually has the most positive index with outcomes, rather than either of the two hypothesised indices (D_{RR} or D_{VR}), and also produced the highest proportion of significant results. It has been clear from analysis presented in this and the previous chapter that the Gini index works somewhat differently from most other indices, and often produces unusual results. This will be discussed in more detail later.

Other than this, however, the range (D_{RR}) has the most positive relationships with outcomes, and this is significantly stronger than all other indices besides the Gini index. The hypothesised spread index, however, actually has the most negative results, and is a significantly weaker than D_{RR}, D_{DR} (the coefficient of variation) or the Gini index.

This would appear to suggest, therefore, that it is range rather than spread that has the more positive relationships with outcomes, and that simply measuring the range in a purely mathematical sense is the best way to capture that (the Gini index notwithstanding).

8.3.2 Variety for ordinal variables

The final column of Table 8.3.1 shows the average effect size and estimated power for each diversity index applied to educational diversity when it was measured ordinally. Only five different outcomes (all forms of innovation) were available, all in the same data set, and therefore the conclusions that can be drawn from this part are more limited than the previous sections.

As with tenure diversity, the most positive results were found using the Gini index, which represents an anomaly. Unlike tenure diversity, however, the next most positive results were not with the range index (D_{RO}) but with the spread index (D_{VO}). These effects were significantly larger than many other indices, although the differences were not significant for the ordinal-specific indices a_{wg}, D_{SO}, l or LOV, or for AD_M. Unlike these other indices, though, D_{VO} did produce 40% (2 out of 5) significant results – the only index besides the Gini index to do so. The range, however, was on a par with most other indices, and was not significantly better at predicting positive outcomes than any.

Table 8.3.2: Variety for nominal variables – Average effect size (and power) for different indices

	Functional (occupational) diversity
No. of effects	21
D_{SN}	-0.07 (0.14)*,**
D_{RN} (range)	0.10 (0.29)**
D_{VN}	0.04 (0.14)*
Blau	0.01 (0.14)*,**
IQV	0.01 (0.14)*,**

Figures in table represent mean regression coefficient size found across relevant data sets (figures in parentheses indicate proportion of significant effects found)
D_{RN} or D_{VN} index hypothesised to have most positive effect
* indicates a significantly different ($p < .05$) average effect size from the D_{RN} index
** indicates a significantly different ($p < .05$) average effect size from the D_{VN} index

8.3.3 Variety for nominal variables

Table 8.3.2 shows the average effect size and estimated power for each nominal diversity index. The only attribute to be appropriately specified as variety was functional (occupational) diversity, and in total, there were 21 different relationships across the data sets involving this variable. As before, the effects for Blau's index and the IQV are identical. Range and spread, measured by D_{RN} and D_{VN} respectively, are hypothesised to produce the most positive results, although it is not predicted which would produce the highest effects.

It is clear to see from the table, however, that range (D_{RN}) has the highest average effect, and also the highest proportion of statistically significant results. Its mean effect size is also significantly higher than those of all other indices. Spread (D_{VN}, or Teachman's index) produces the next highest results – also significantly higher than the other indices – but the estimated effect sizes and power lag some way behind those of range.

8.3.4 Variety – summary

For the continuous variable tenure, the range (D_{RR}) was found to give more positive (and therefore theoretically consistent) relationships with outcomes of variety than any other index besides the Gini index. For the ordinal variable educational background, however, the hypothesised spread index (D_{VO}) gave the most consistently positive results other than the Gini index. Spread for continuous variables was considerably weaker than range, whereas range for ordinal variables was a little weaker than spread. For the nominal variable functional background, range gave the most positive results, with spread significantly weaker than range but still significantly better than other indices, including the widely used Blau's index (or the adjusted version of Blau's index).

Between them, these results that both range and spread could be seen as more important, depending on the variable (and theory) in question. Crucially, there are significant differences between the two versions of variety across all types of variable, suggesting that these two versions of variety that I have introduced are indeed different from each other; equally, either could be seen as the more important predictor of outcomes such as innovation, depending on the attribute in question. Thus hypotheses 2b (the indices D_{RR}, D_{RI}, D_{RO} and D_{RN} better predict outcomes of range for ratio, interval, ordinal and nominal variables respectively than do any other index) and 3b (the indices D_{VR}, D_{VI}, D_{VO} and D_{VN} better predict outcomes of spread for ratio, interval, ordinal and nominal variables respectively than do any other index) are generally supported as far as can be determined by a limited number of data sets with no clear predictions about the differential effects of range and spread.

A slight spanner is thrown into the works by the Gini index, which appears to predict these outcomes even better than the hypothesised range and spread indices. This will be discussed further in Chapter 9.

8.4 Chapter summary

Overall, the analysis conducted in this chapter provides support for the hypotheses and therefore for the new typology. The proposed indices for separation provide at least as negative effects as any other indices, with higher statistical power, and therefore hypothesis 1b is largely supported. The proposed indices for either range or spread usually provide the highest positive effects with innovation outcomes, with different variables showing differential effects between these two new constructs. This suggests that further work will be necessary to disentangle the situations in which range or spread would be expected to have the largest relationships with outcomes.

Although the results presented in this chapter and the previous one generally support the hypotheses, there were instances where this support was not as conclusive or definitive as might have been desired. Therefore in the next chapter I summarise all the results, integrating both the simulated and real data analysis, and discuss what can be concluded for the practice of studying and measuring diversity in the future.

CHAPTER 9

DISCUSSION AND CONCLUSIONS

9.0 Chapter introduction

The aim of this chapter is to summarise the typology I have derived and presented, and to integrate the results of the analysis of simulated and real data sets to form overall conclusions about the relative merits of diversity indices and support for the hypotheses. It then presents implications arising from these results for both theory and researchers working with such indices, and concludes by evaluating the limitations of the work and suggesting directions for future research.

9.1 Overall summary and integration of findings

9.1.1 The new typology

In this thesis I have sought to clarify issues around the definition and measurement of work group diversity by providing a typology of four distinct definitions for diversity, and ten indices (covering 14 pairs of definition and data type) that researchers can use to measure it. The need for such a typology was provided by the lack of consistency in definition and measurement in the work group diversity literature, which may have contributed to the sheer variety in results found for outcomes of work group diversity. I have then tested the proposed indices by comparing them with other possible methods of measurement in a large simulated data set and eight real data sets, with findings largely supporting the hypotheses that they measure the appropriate forms of diversity better than other possible indices.

In particular, I have built on the work by Harrison and Klein (2007) who provided definitions for three types of diversity – separation, variety and disparity – and who suggested two indices apiece for each. Harrison and Klein's work was the first systematic attempt to clarify definitions of diversity, and to recommend how these should be measured. Nevertheless, there were four limitations in their paper that I have sought to address. First, and most seriously, they aligned their definitions of separation and variety not only to theoretical approaches but also to types of variable: specifically, separation was defined for continuous variables only and variety for categorical variables only. This meant that, for example, there was no index defined for measuring separation of nationality: an attribute that could easily be a basis of social categorisation. This has led to inappropriate indices being used in the literature: for example, Kearney and Gebert (2009) used Blau's index to measure age diversity (as variety), when such an index is psychometrically inappropriate for a continuous variable. I have addressed this by providing indices for each type of diversity for each of the four major data types: ratio, continuous, ordinal and nominal.

A second and related point is that relating definitions of diversity so closely to theoretical perspectives leaves researchers unclear what to do with a more integrative model, such as van Knippenberg, De Dreu and Homan's (2004) Categorization-Elaboration Model, which specifically combines theory from both the social categorization and information/decision making perspectives. I had therefore suggested that it is the definition of diversity itself that determines the method of measurement, rather than the broader theoretical perspective. Third, Harrison and Klein's (2007) definition of variety actually incorporated two related but separate constructs, which I have called range (the total range or number of categories represented by a group) and spread (the extent to which group members are well spread over a large range). I have defined these separately and provided suggested indices for both. Finally, Harrison and Klein presented two possible indices for each diversity definition without giving a rationale for choosing one over the other. I have therefore proposed indices that are specifically aligned to the actual definition of the diversity type, and gone on to test these with both simulated and real data.

The consideration of the type of data – ratio, interval, ordinal or nominal – is one which should be of critical importance for all researchers, but has received relatively little attention to this point. In general, there is complete understanding of the differences between continuous and nominal data (although, as mentioned above, there are examples of authors with continuous data using indices designed for nominal data); however, two issues remain. First is the use of ordinal data – common amongst organisational researchers for variables such as educational background, and sometimes for approximating continuous variables such as age – for which specific diversity indices have seldom appeared in the work group diversity literature. I have drawn from the sociological literature, and in particular the work of Blair and Lacy (1996, 2000) to help address this. Second is the distinction between ratio and interval variables, the latter being frequently found in questionnaire scales. For three of the diversity types (separation, range and spread) the distinction is irrelevant, with the same index being applied to both types of data. For disparity, however, the distinction is crucial, as the most commonly used index – the coefficient of variation – only makes sense when used with ratio data. This was the subject of previous work by Bedeian and Mossholder (2000), who warned against the inappropriate use of the index, and these warnings were repeated by Harrison and Klein (2007). Despite this, it persistently shows up in the literature being used for interval (or even ordinal) data (e.g. Hmieleski & Ensley, 2007; Hooper & Martin, 2008; Jehn & Bezrukova, 2004). The testing of this typology therefore represents an attempt to quantify the effects of such misuse and mis-specification, being one of the first of its kind to do so – Roberson, Sturman and Simons

(2007) having used a similar approach but limited to simulated data in multilevel situations with fewer indices.

An inherent problem with the testing of such a typology, however, is that the use of real data relies on making assumptions about what the actual relationships should be; given the variety of results within the literature and the different theories used, this is not straightforward. In fact, it would not have been surprising to find a lack of any clear results coming from the data available. That many clear differences were observed is an indication that there are consistent differentiations between indices, even though the variation of effects between variables and data sets is huge. This is further supported by the results from simulated data, which demonstrate more clearly the effects of mis-specification of diversity type or incorrect use of indices. I will now summarise the results for separation, range, spread and disparity in turn, stating what the findings mean for measurement of these constructs.

9.1.2 Measurement of separation

Separation was hypothesised to be measured by three possible indices; the first (for ratio and interval variables, and denoted as D_{SR} and D_{SI} accordingly) being the coefficient of mean difference as originally proposed by Helmert (1876), as well as the pure form of the mean Euclidean distance for a single dimension, and representing the average difference between all pairs of group members. For ordinal variables the index D_{SO} was the $1 - l^2$ measure given by Blair and Lacy (2000), whereas for nominal variables the proposed measure D_{SN} was the adjusted form of Blau's index, as given by Harrison and Klein (2007). (Harrison and Klein actually proposed it as a measure of variety; again, though, this more accurately measures differences between pairs of individuals.) Results from the two studies in Chapters 7 and 8 were fairly consistent in most regards. In both cases, the D_{SR}/D_{SI} index was amongst the best at predicting outcomes as expected. The incorrect use of a range, spread or disparity index could lead to far weaker effects, with lower statistical power. However, some other indices – notably the AD_M index – performed virtually as well as the D_{SR}/D_{SI} index; in the simulation study the standard deviation and Euclidean distance actually matched the performance of the hypothesised index, although in the real data sets these gave slightly weaker effects which were less likely to be statistically significant. Therefore, across the studies, the D_{SR}/D_{SI} index was the best performing, although its advantage over AD_M, standard deviation and Euclidean distance – the latter two being those proposed by Harrison and Klein (2007) – was small and it is unlikely that researchers would lose much accuracy by using one of these indices instead.

For ordinal data also, the use of indices designed for range, spread or disparity led to far weaker effects and lower power. Amongst the other indices tested, the standard deviation, Euclidean distance, AD_M and variance actually worked almost as well as the hypothesised D_{SO} index in both studies. This suggests that the use of indices designed for continuous data may not be disastrous for research results. However, a weakness of this conclusion is that the values of an ordinal variable could be changed or separated/collapsed, which would lead to potentially large changes to indices designed for continuous variables, but less so for those for ordinal variables. This puts the onus on the researcher to use the correct study design to start with: ensuring each successive category of an ordinal variable is distinct from adjacent ones, but that all salient differences are captured. Amongst those indices that are specifically designed for ordinal data, D_{SO} was almost matched in performance by a_{wg}, l and LOV in the simulation studies, but not so much with the real data (although fewer effects were available to test for ordinal data). On the basis of the collective evidence, however, D_{SO} would appear to be the best index to use to measure separation in ordinal variables.

For nominal variables, the hypothesised index was only marginally better than the others. The simulated data revealed small disadvantages of using Teachman's index rather than D_{SN} for testing separation, although more significant disadvantages were found if the range index D_{RN} were used instead. Real data analysis showed slightly stronger (i.e. more negative) effects with outcomes for D_{SN}, so there is support for the hypothesis that this is the index to use, although the disadvantages of using Teachman's index, or the unadjusted Blau's index, are likely to be minimal; the large number of papers in the literature including analysis using one of these two indices would unlikely to be changed much if a different index were used instead. Overall, however, hypotheses 1a and 1b were supported.

9.1.3 Measurement of range and spread

The distinction between range and spread was demonstrated to be an important one empirically as well as theoretically. Indeed, across the studies, the difference between range and spread appeared to be as large as that between either and separation. Range is a particularly easy construct to measure: it is simply the mathematical range of the sample, either the difference between minimum and maximum values (continuous and ordinal variables) or the number of distinct categories represented (nominal variables). Despite this simplicity, though, it has been used relatively rarely in the work group diversity literature: this is all the more surprising given the results found. For ratio and interval variables, the use of separation indices (including those not specifically hypothesised) instead of range led to slightly weaker effects in the simulation studies, and the use of a spread index to far weaker effects. In the real data sets, range was the index most consistently positively related to innovation outcomes – more so than either separation or spread. As it happened, the Gini index gave even more positive results in the real data, but simulated data showed that the use of this index led to highly attenuated correlations, so this appears to be an anomalous effect and the Gini index should be ignored for all purposes other than disparity. For ordinal data some of these effects appeared to be reversed: the use of a spread index was less damaging than the use of a separation index in the simulations, and spread had a larger relationship with innovation in the real data set. The differences in the real data analysis are likely to be because of the variables being used: educational variety and firm innovation in top management teams provided all of the effects, and as increased range in this scenario is likely to refer to having a team member of low educational background, this may not in itself be as likely to lead to innovation as having members from a selection of different educational backgrounds (e.g. Dahlin, Weingart & Hinds, 2005). Likewise, for nominal variables, much weaker effects tended to be found in the simulated data if a different index were used, and the largest relationships with innovation outcomes were found in the real data by using the range index. Thus hypothesis 2a is fully supported, and hypothesis 2b as supported as possible given the lack of theoretically distinct predictions for outcomes of range and spread.

The measurement of spread, however, was somewhat more complex. For ratio and continuous variables the proposed index D_{VR}/D_{VI} was a new one, which takes into account both the range of the group and the evenness of data within that range. For nominal variables, D_{VN} was Teachman's index, as this describes both range and evenness within its definition. For ordinal variables a combination of the two was used: D_{VO} was Teachman's index multiplied by the range. For all indices the simulation studies showed that the use of alternative indices resulted in weaker effects being found and less statistical power, although for nominal data Blau's index and its adjusted version were nearly as good. Using a separation index for continuous or ordinal data was far worse, with the use of range for a continuous variable equally as bad at predicting results accurately. As far as real data goes, spread was not as strong a predictor of innovation for nominal variables as range was, although it was a stronger predictor than any separation index; for ordinal data, however, spread was a superior predictor of innovation outcomes compared with range. Amongst continuous data there was very little effect of spread on such outcomes. This may be because spread of tenure (the attribute in question for all effects tested) is a less theoretically important factor for innovation, or because of a weakness in the index itself. Overall, however, hypothesis 3a is fully supported, and hypothesis 3b partly supported (with some doubts around the indices for ratio and interval variables).

9.1.4 Measurement of disparity

Disparity is a somewhat different construct from the others listed, and my definition is not changed from that given by Harrison and Klein (2007). The definition lends itself to the use of the coefficient of variation (CV), so that is the index chosen and denoted D_{DR}. As noted previously, however, this is only appropriate to be used for ratio variables. Therefore two interesting questions that arise here are: (i) can disparity be measured for interval variables, and (ii) how does this compare with the Gini index, which was recommended by Harrison and Klein alongside the CV? These are questions that have to be answered purely by examining the simulated data, as no obvious hypothesis involving real data could be found.

The simulated data show that, amongst ratio data, no other index comes anywhere close to predicting effects accurately in place of the CV. The Gini index is the next best, but even then effects are attenuated by up to 30% with a severe loss of power. Combined with other results about the Gini index (which show that it can lead to larger effect sizes than predicted indices with real data, despite simulations showing that its use is likely to attenuate effects), and its formula (involving division by the mean, making it unsuitable for any form of diversity other than disparity of ratio variables), this suggests that it should not be considered seriously for use by researchers of work group diversity, and thus this refines the recommendations of Harrison and Klein (2007).

For interval data, an adjusted form of the CV was hypothesised, with the minimum possible (or observed) value of the attribute in question subtracted from the mean before it was divided into the standard deviation. This was intended to "normalise" the data, in the sense that it would then operate on a scale starting at zero (which would give it the meaning required for disparity). Simulations showed that this was not quite matched by any other index, although the coefficient of variation was only a little worse, with around 10% attenuation of effect sizes and slightly larger drops in power. Thus, if disparity of an interval variable is meaningful, use of the CV does not appear to do much damage to potential effects. It remains a question, however, as to whether the adjustment used is meaningful. For a variable that is spread evenly over a scale from, say, 1 to 5, then converting this to a scale from 0 to 4 may be sensible. But if a variable has a hypothetical range of 1 to 5, but in reality most values are grouped in the range 3-4, then this subtraction seems very arbitrary, as different relative values would be produced if the scale were reversed, for example. Therefore despite hypothesis 4 being supported by the data, I advise caution and careful consideration before applying disparity to an interval variable.

9.2 Theoretical implications

This thesis offers contributions to knowledge both in terms of understanding of different forms of diversity, and in terms of providing a guide for how it should be measured. The latter can be considered a theoretical contribution in some regards, but as the users of this research are likely to be researchers themselves, more so than practitioners in the workplace, I consider that under the heading of practical implications (section 9.3). In this section I discuss how my work can help researchers consider different conceptualisations of diversity, particularly examining the distinction between range and spread, and highlighting how the conceptualisation should be determined independently of the data.

Theoretically, the work builds on that of Harrison and Klein (2007), and offers two significant expansions to their typology. First is the idea that the type of diversity is, to a large extent, independent of the data type of the attribute under consideration. For example, the concept of separation should not be confined to variables measured along a continuum, as suggested by Harrison and Klein, but could just as easily apply to a categorical variable such as nationality. The key issue for separation is that it refers to the collective amount of differences between pairs of individuals within a group. Harrison and Klein (2007) related this to social categorisation, which fits this definition in many respects: if each individual within a group compares him/herself with each other member, then it is the total summation of these comparisons that should be measured. Certainly, though, research based on this perspective has not been limited to continuous variables; for example, Kooij-de Bode, van Knippenberg and Ginkel (2008) base their predictions about ethnic diversity and group decision making on social categorisation theory, as do Earley and Mosakowski (2000) for their study of national diversity and team performance. Many other such examples exist: however, the diversity type need not be linked directly to the theoretical perspective. Indeed, this cannot be the case if a more integrative model is used; for example, van Knippenberg, Homan and De Dreu's (2004) Categorization-Elaboration Model, which incorporates both aspects of social categorization and also the elaboration of information associated with the information/decision making perspective of diversity. Of more importance is the precise definition of diversity as given by the researcher.

Likewise the two forms of variety, range and spread, are not confined to categorical variables but can be applied to continuous variables as well. This may be particularly important when it is hypothesised that the variety of a construct such as tenure (e.g. Tihanyi et al., 2000) or educational level (e.g. Jehn & Bezrukova, 2004) is hypothesised to have a positive relationship with outcomes through process such as information sharing.

In contrast, disparity by definition relates to an asymmetrical variable, and therefore this is not independent of data type. Indeed, the construction of the proposed index is derived from situations where zero is a meaningful value (i.e. ratio variables), and therefore a question remains over whether interval variables – even when they are asymmetrical – could be appropriate attributes for disparity.

The second theoretical contribution is to further delineate types of diversity. In particular, I have introduced the concepts of range and spread to replace the single concept of variety. The difference between the two is akin to asking the question – is it enough to have a single member of a group with a certain value (which would imply greater range), or is a group only considered more diverse when all members are spread evenly over that range (which would imply greater spread)? It is quite possible that different theoretical viewpoints would favour one approach over the other. For example, Pelled, Ledford and Mohrman (1999) argue that a person with a distinct tenure or educational level (not necessarily the highest or lowest) may be viewed as a valued task resource, which suggests spread of these attributes is linked to their outcome, inclusion. On the other hand, Somech and Drach-Zahavy (2007) argue that a greater number of functional backgrounds within a team – i.e. range – will increase innovation. The differences between these approaches are small, and rely on careful definition of concepts and clear theory. The use of range as a concept also tallies with the theory of minority influence, which suggests that a single person in a minority position in a group can be as influential – or even more so – than a larger minority (Moscovici & Nemeth, 1974; Clark & Maass, 1990).

Indeed, the distinction between range and spread is illustrated by the results found. For the relationship between tenure diversity and innovation, range was found to be most important, and spread not important at all. For the relationship between educational background diversity and innovation, spread was found to be most important, and range not very important at all. For functional background diversity, range was found to be the most important diversity type, with spread also important but less so than range. The result for tenure suggests that a mix of experience and fresh ideas may be the most beneficial aspect, with what comes in between not so crucial: the minority influence of one new team member, say, could be highly valuable for innovation. The same appears to be the case for functional background, although the results for spread of functional background suggest that this will be stronger when all groups are more equally represented in the team. Conversely, for educational background, having one member of the team with a PhD or who dropped out of school before the age of 16 does not appear to increase overall innovativeness; however, a more even spread of people with school qualifications, first degrees and higher degrees would seem to lead to greater informational resources and perspectives brought to a team. Future research should carefully consider the theoretical implications of studying range or spread for different attributes and outcomes: for a particular diversity attribute, what is the mechanism by which a particular outcome is likely to be affected?

9.3 Practical implications

The main practical contribution of this thesis is to offer work group diversity researchers a clear guide to choosing an appropriate method of operationalisation. Whilst some previous research has offered some recommendations about comparing different forms of measurement (e.g. Allison, 1978; Harrison & Sin, 2005; Harrison & Klein, 2007; Biemann & Kearney, 2009; Roberson, Sturman & Simons, 2007), or described particular indices in more detail (e.g. Bedeian & Mossholder, 2000; Brown & Hauenstein, 2005; Burke, Finkelstein & Dusig, 1999), this is the first typology which recommends a single specific index for each diversity type and data type. The fact that these recommendations are based on matching the measurement process to the precise definition should assure researchers that each proposed index is appropriate, and that they can avoid the tendency to choose an index based on prior literature alone – which sometimes leads to inappropriate choices (e.g. Klein et al., 2001; Jehn & Bezrukova, 2004). Researchers can be further reassured by the results of the analysis of both simulated and real data, which supported the proposals.

One associated practical implication of this, though, is the need for researchers to define diversity more carefully. This aligns with Harrison and Klein's (2007) recommendation, and suggests that scholars should not use the term "diversity" loosely, but instead consider which precise definition matches their theoretical arguments.

The analysis conducted also sheds light on Harrison and Klein's (2007) recommendations for indices, and other commonly used indices in the literature. For example, I have shown that one recommended index for disparity (the coefficient of variation) is far more appropriate than the other (the Gini index). For separation, Harrison and Klein recommended the use of either the standard deviation or the Euclidean distance. While neither matches the definition of separation quite as well as the proposed D_{SR} index (which is nearly identical to the rarely-used Coefficient of Mean Difference), both produce similar results to the D_{SR} index with effect sizes only marginally smaller. This implies that researchers using one of these two indices – as many have in the past – are unlikely to find substantively different results from those that would have been achieved had D_{SR} been used. The same is true for the AD_M index and the variance (and therefore also the r_{WG} index). The same cannot be said for the coefficient of variation, however; as warned by Bedeian and Mossholder (2000) amongst others, this can lead to very different results if used inappropriately – i.e. attempting to measure anything but disparity for ratio variables. This has potential implications for understanding the many findings in the literature that used the coefficient of variation. Effect sizes may have been incorrectly estimated; incorrect conclusions may have been drawn about whether or not an effect was significant; and the interpretation of what diversity means in those cases (disparity as opposed to separation, for example) may be flawed. For example, Choi, Price and Vinokur (2003) used the coefficient of variation to conclude that diversity in education amongst group members was associated with increased job search efficacy, and they attribute this to the increased source of information and in-depth understanding of job markets and job search strategies in such groups; this argument is more in line with variety (range or spread) rather than disparity. An interpretation more closely aligned with the coefficient of variation might be that when there is one individual in a group with much higher educational status than the rest, he/she may be able to help, inspire or influence those of lower educational status. This may not fit with the authors' theory, and may not be correct; however, without using a more appropriate index (e.g. range), there is limited support for their precise argument. There is still, of course, a high chance that similar results would have been found in this and other studies with a more appropriate index, but this is far from certain.

In terms of nominal variables, I have given close attention to the differences between Blau's (1977) index and Teachman's (1980) index. Both have been used widely in the literature, often under different names, and occasionally researchers have used more than one. Harrison and Klein (2007) did not offer any reason for choosing one over the other, but I have formulated such a rationale and therefore given researchers a guide as to when each should be used. In addition, I have included the number of categories represented as a measure of range: the results of the analysis, particularly that of real data sets, suggest that this seldom-used index (Fay et al., 2006, being one exception) could well be an important one in the diversity researchers' armoury. However, the close relationship between Blau's and Teachman's indices suggest that most findings in the literature would probably not have been significantly different if authors had chosen the other index instead.

Another finding throughout my analysis is the relatively low power under certain circumstances. It can be seen from the simulation that this unsurprisingly depends on sample size, but even with a moderately good sample (e.g. 100 teams, representing a substantial data collection effort) the power can be very low under some circumstances. This is further borne out by the analysis of real data sets, which vary between 67 and 133 teams in size; in all cases, the proportion of effect sizes found significant was below 50%, and usually well below. This suggests that there is a substantial amount of luck in whether or not an effect is found by a researcher to be significant; of course, the best remedy for this is to collect larger data sets, but practically this is often not possible. It does however suggest that part of the reason diversity findings have been so haphazard in the literature is because studies cannot be relied upon to generate the findings that researchers may expect.

9.4 Limitations

There are of course a number of limitations with this study; some are theoretical, others empirical. From the theoretical perspective, although I have distinguished between range and spread as versions of diversity, it remains possible that other definitions of diversity beyond the four in my typology could be useful. If this is the case, my advice to the researchers who are looking for how to measure such a construct would be to take the same approach as I have done: find (or construct) an index for which the properties mirror those of the diversity definition. It is likely to be relatively rare that the need for new indices occurs however; I would expect the current four types of diversity to encompass most situations and theories used by work group diversity researchers.

Another theoretical limitation is that I have only created a definitional difference between range and spread, and not made theoretical predictions about how the two would differentially predict outcomes; consequently I have been restricted to observing how the two types of indices work for different attributes in the real data analysed. However, the definitions of range and spread should be relatively clear, and researchers should be able to use these to determine which is the more appropriate for their own particular theory and study. Subsequent theoretical advances and empirical research will hopefully show how the two differ in relationships with other variables.

A philosophical limitation of the testing of the typology of indices is that, whether real or simulated data are used, the accuracy of the typology cannot be completely verified because either it is not known what the actual relationship between diversity and outcomes is (in the real data sets); whilst in the simulation the data have to be generated so that one particular index is the "correct" version of diversity, and that this will naturally appear to be better than all other indices, which does not necessarily reflect a real situation. The use of the two methods together overcomes this as far as possible: the simulated data allows us to see what the effect of mis-specification of a diversity type or index would be, as we assume that the chosen index is the "correct" specification; the real data, conversely, shows us which indices are more closely related to outcomes in actual samples, despite the fact that we do not know for certain which diversity type should be related to which outcome, but are basing assumptions on the somewhat patchy diversity literature. The limitation is further mitigated by the support for the hypotheses from both methods, which demonstrate overall agreement with the typology.

Another limitation is the data that were analysed. The real data sets were confined to eight available data sets which, although typical of many found in organisation research in some respects, included five health care data sets, two student data sets, were relatively homogeneous in size and did not cover all variables in the diversity literature (even though all of the most common attributes were covered). The simulated data – despite the fact that over 800 million data points were generated – do not cover all possibilities in terms of data type or sample. For example, each replication of the simulation generated data from teams with exactly the same size. However, it is only ever possible to cover a finite number of possibilities, and the simulations broadly cover the range of data observed in the literature.

A final limitation is that I have only studied linear relationships between diversity and outcomes. Increasingly, curvilinear results are predicted and found in the literature (e.g. Carpenter & Fredrickson, 2001), and moderated effects of diversity are now highly common (van Knippenberg & Schippers, 2007). Future research may look at the effects of using different indices for these non-linear relationships.

9.5 Future research directions

Three areas for future research stand out. First is the "completion" of the analysis I have done on real data sets, both by expanding tests to other data sets with different variables included, and (more importantly) by including specific tests of disparity, and of the distinction between range and spread. These are more difficult for different reasons. The tests of disparity would be harder because few organisational data sets appear to have the right sort of variables for this (many researchers, e.g. Carpenter & Fredrickson, 2001; Ely, 2004; Tihanyi et al., 2000, have used the coefficient of variation with a variable such as tenure, but it is not clear that they really intended to measure disparity as defined by Harrison and Klein, 2007). Tests for the distinction between range and spread would require convincing theoretical arguments to be made for differential outcomes first.

A second area where further research would be helpful is the effect of incomplete team data on the measurement of diversity. This is something that has been examined in the past for some indices (Newman & Sin, 2009), but a systematic study would be welcome.

A final area for future research is utilising the fact that, for a nominal variable, not all categories are equally different from each other. This was mentioned in Chapter 5, which highlighted Dawson & Brodbeck's (2005) work on using culture scales to differentiate countries when measuring cultural diversity. There are a number of ways in which such differentiation could be done: e.g. using existing data sources, measuring other variables as proxies, or getting expert ratings of differences between categories. For separation, the resulting indices would be relatively straightforward, as exemplified in Chapter 5. However, for other types of diversity, constructing indices would require a more sophisticated approach that would still need to be developed.

9.6 Overall conclusion

In this thesis I have developed an extended version of Harrison and Klein's (2007) typology of work group diversity, by clarifying definitions, distinguishing between two forms of variety, and proposing indices to measure each of the four types of diversity with each possible type of data. Tests of the relative merits of the proposed indices using both simulated and real data revealed broad support for them, and therefore for the typology also.

It is to be hoped that diversity researchers will be able to use the typology to choose an appropriate form of measurement for work group diversity, particularly in the cases where few clear guidelines existed previously. More generally, the differences found between results with different indices suggest that researchers should consider very carefully which form of diversity most closely fits their theory.

REFERENCES

Agresti, A., & Agresti, B. (1978). Statistical analysis of qualitative variation. *Sociological Methodology, 10,* 204-237.

Alagna, S., Reddy, D., & Collins, D. (1982). Perceptions of functioning in mixed-sex and male medical training groups. *Journal of Medical Education, 57,* 801-803.

Alexander, J., Nuchols, B., Bloom, J., & Lee, S. Y. (1995). Organizational demography and turnover: An examination of multiform and nonlinear heterogeneity. *Human Relations, 48,* 1455-1480.

Allen, N., Stanley, D., Williams, H. & Ross, S. (2007). Assessing the impact of nonresponse on work group diversity effects. *Organizational Research Methods, 10,* 262–86.

Allison, P. D. (1978). Measures of inequality. *American Sociological Review, 43*, 865-880.

Ancona, D. G., & Caldwell, D. F. (1992). Demography and design: Predictors of new product team performance. *Organization Science,* 3: 321-341.

Anderson, N. & West, M. A. (1998). Measuring climate for work group innovation: development and validation of the Team Climate Inventory. *Journal of Organizational Behavior, 19,* 235-258.

Bantel, K., & Jackson, S. (1989). Top management and innovations in banking: Does the composition of the team make a difference? *Strategic Management Journal, 10,* 107-124.

Bardo, J. W., Yeager, S. J. and Klingsporn, M. J. (1982). Preliminary assessment of format-specific central tendency and leniency error in summated rating scales. *Perceptual and Motor Skills, 54,* 227-234.

Barrick, M. R., Stewart, G. L., Neubert, M. J., & Mount, M. K. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology, 83,* 377-391.

Beal, D. J., & Dawson, J. F. (2007). On the use of Likert-type scales in multilevel data. *Organizational Research Methods, 10,*657-672.

Bedeian, A. G., & Mossholder, K. W. (2000). On the use of the coefficient of variation as a measure of diversity. *Organizational Research Methods, 3,* 285-297.

Berry, K. J., & Mielke, P. W. (1992). Assessment of variation in ordinal data. *Perceptual and Motor Skills, 74,* 63-66.

Biemann, T., & Kearney, E. (2010). Size does matter: How varying group sizes in a sample affect the most common measures of group diversity. *Organizational Research Methods, 13,* 582-599.

Blair, J., & Lacy, M. G. (1996). Measures of variation for ordinal data as functions of the cumulative distribution. *Perceptual and Motor Skills, 82,* 411-418.

Blair, J., & Lacy, M. G. (2000). Statistics of ordinal variation. *Sociological Methods and Research, 28,* 251-280.

Blau, P. M. (1977). *Inequality and heterogeneity*. New York: Free Press.

Borrill, C. S., Carletta, J., Carter, A. J., Dawson, J. F., Garrod, S., Rees, A., Richards, A., Shapiro, D., & West, M. A. (2001). *The Effectiveness of Health Care Teams in the National Health Service*. Final Report to the Department of Health: ISBN 1854494252.

Bowers, C. A., Pharmer, J. A., & Salas, E. (2000). When member homogeneity is needed in work teams. A meta-analysis. *Small Group Research*, *31,* 305-27.

Brodbeck, F. C., Guillaume, Y. F. R., & Winkler, M. (2010). *Team Climate for Learning in Higher Education*. Aston Business School Working Paper.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 3,* 296–322.

Brown, R. D., & Hauenstein, N. M. A. (2005). Interrater agreement reconsidered: An alternative to the r_{WG} indices. *Organizational Research Methods, 8,* 165-184.

Bunderson, J.S., & Sutcliffe, K. (2002). Comparing alternative conceptualizations of functional diversity in management teams: Process and performance effects. *Academy of Management Journal, 45,* 875–893.

Burke, M. J., & Dunlap, W. P. (2002). Estimating interrater agreement with the average deviation index: A user's guide. *Organizational Research Methods, 5,* 159-172.

Burke, M. J., Finkelstein, L. M., & Dusig, M. S. (1999). On average deviation indices for estimating interrater agreement. *Organizational Research Methods, 2,* 49-68.

Byrne, D. (1971). *The attraction paradigm*. New York: Academic Press.

Calman, K., & Hine, D. (1995). *A policy framework for commissioning cancer services*. London: Department of Health.

Cannon, M. D., & Edmondson, A. C. (2001). Confronting failure: Antecedents and consequences of shared beliefs about failure in organizational work groups. *Journal of Organizational Behavior, 22,* 161-177.

Carpenter, M. A., & Fredrickson, J. W. (2001). Top management teams, global strategic posture, and the moderating role of uncertainty. *Academy of Management Journal, 44,* 533-545.

Chatman, J. A., & Flynn, F. J. (2001). The influence of demographic heterogeneity on the emergence and consequences of cooperative norms in work teams. *Academy of Management Journal, 44*, 956-974.

Chatman J. A., & O'Reilly, C. A. (2004). Asymmetric reactions to work group sex diversity among men and women. *Academy of Management Journal, 47,* 193-208.

Chattopadhyay, P. (1999). Beyond direct and symmetrical effects: The influence of demographic dissimilarity on organizational citizenship behavior. *Academy of Management Journal, 42,* 273-87.

Chattopadhyay, P., & George, E. (2001). Examining the effects of work externalization through the lens of social identity theory. *Journal of Applied Psychology, 86,* 781-88.

Chattopadhyay, P., George, E., & Lawrence, S. A. (2004). Why does dissimilarity matter? Exploring self-categorization, self-enhancement, and uncertainty reduction. *Journal of Applied Psychology, 89,* 892-900.

Choi, J. N. (2007). Group composition and employee creative behaviour in a Korean electronics company: Distinct effects of relational demography and group diversity. *Journal of Occupational and Organizational Psychology, 80,* 213–234.

Choi, J. N., Price, R. H., & Vinokur, A. D. (2003). Self-efficacy changes in groups: Effects of diversity, leadership and group climate. *Journal of Organizational Behavior, 24,* 357-372.

Clark, R. D., & Maass, A. (1990). The effects of majority size on minority influence. *European Journal of Social Psychology, 20,* 99-117.

Clore, G. L., & Byrne, D. A. (1974). A reinforcement-affect model of attraction. In T. L. Huston (Ed.), *Foundations of interpersonal attraction*, 143-170. New York: Academic Press.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37-46.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.

Coulter, P. B. (1984). Distinguishing inequality and concentration: The exponentiation principle. *Political Methodology, 10,* 323-335.

Coulter, P. B. (1989). *Measuring inequality: A methodological handbook.* Boulder, Colorado: Westview Press.

Cox, T., & Blake, S. (1991). Managing cultural diversity: Implications for organizational competitiveness. *Academy of Management Executive, 5(3)*, 45-56.

Cummings, A., Zhou, J., & Oldham, G. (1993). Demographic differences and employee work outcomes: Effects of multiple comparison groups. Paper presented at the Annual Meeting of the Academy of Management, Atlanta, GA.

Curseu, P. L., Schruijer, S. G. L., & Boros, S. (2007). The effects of groups' variety and disparity on groups' cognitive complexity. *Group Dynamics: Theory, Research and Practice, 11,* 187-206.

Dahlin, K. B., Weingart, L. R., & Hinds, P. J. (2005). Team diversity and information use. *Academy of Management Journal, 48,* 1107-1123.

Dawson, J. F., & Brodbeck, F. (2005). *Not all differences are the same: Measuring diversity using distances between categories*. Paper presented at the Annual Meeting of the Academy of Management, Honolulu, HI, August 2005.

Dawson, J. F., Gonzalez-Romá, V., Davis, A., & West, M. A. (2008). Organizational climate and climate strength in UK hospitals. *European Journal of Work and Organizational Psychology, 17,* 89-111.

Drach-Zahavy, A., & Somech, A. (2001). Understanding team innovation: The role of team processes and structures. *Group Dynamics: Theory, Research, and Practice*, *5,* 111-123.

Earley, C. (1989). East meets West meets Mideast: Further explorations of collectivistic and individualistic work groups. *Academy of Management Journal, 36,* 319-348.

Earley, P. C., & Mosakowski, E. (2000). Creating hybrid team cultures: An empirical test of transnational team functioning. *Academy of Management Journal, 43,* 26-49.

Eddington, A. (1914). *Stellar movements and the structure of the universe*. London: Macmillan.

Eisenhardt, K., & Schoonhoven, K. (1990). Organizational growth: Linking founding team strategy, environment, and growth amongst U.S. semiconductor ventures, 1978-1988. *Administrative Science Quarterly, 39,* 203-238.

Ellemers, N., Kortekaas, P., & Ouwerkerk, J. (1999). Self-categorisation, commitment to the group and group self-esteem as related but distinct aspects of social identity. *European Journal of Social Psychology, 29,* 371-389.

Ely, R. J. (2004). A field study of group diversity, participation in diversity education programs, and performance. *Journal of Organizational Behavior*, *25,* 755-780.

Fay, D., Borrill, C.S., Amir, Z., Haward, R., & West, M. A. (2006). Getting the most out of multidisciplinary teams: A multi-sample study of team innovation in health care. *Journal of Occupational and Organizational Psychology, 79,* 553-567.

Ferrier, W. J. (2001). Navigating the competitive landscape: The drivers and consequence of competitive aggressiveness. *Academy of Management Journal, 44,* 858-877.

Ferris, G. R., & Judge, T. A. (1991). Personnel/human resources management: A political influence perspective. *Journal of Management, 17,* 447-488.

Fisher, R. (1920). A mathematical examination of the methods of determining the accuracy of observation by the mean error and the mean square error. *Monthly Notes of the Royal Astronomical Society, 80,* 758-770.

Flatt, S. (1996). Developing innovative strategies: How top management teams bring creativity and implementation into the firm. Paper presented at the Strategic Management Society Conference, Phoenix, AZ.

Foo, M. W., Sin, H.-P., & Yeong, L. P. (2006). Effects of team inputs and intrateam processes on perceptions of team viability and member satisfaction in nascent ventures. *Strategic Management Journal, 27,* 389-399.

George, J. M. (1990). Personality, affect, and behavior in groups. *Journal of Applied Psychology, 75,* 107-116.

Gevers, J. M. P., & Peeters, M. A. G. (2009). A pleasure working together? The effects of dissimilarity in team member conscientiousness on team temporal processes and individual satisfaction. *Journal of Organizational Behavior, 30,* 379-400.

Gini, C. (1912). Variabilitá e mutabilitá: Contributo allo studio delle distribuzioni e relazioni stastistiche. *Studi Economico-Giuridici dell' Universitá di Cagliari, 3*: 1-158.

Glick, W., Miller, C., & Huber, G. (1993). The impact of upper echelon diversity on organizational performance. In G. Huber and W. Glick (Eds.), *Organizational Change and Redesign* (pp. 176-224). New York: Oxford University Press.

González-Romá, V., Peiró, J. M., & Tordera, N. (2002). An examination of the antecedents and moderator influences of climate strength. *Journal of Applied Psychology, 87*: 465-473.

Gorard, S. (2005). Revisiting a 90-year-old debate: the advantages of the mean deviation. *The British Journal of Educational Studies, 53,* 417-430.

Greenberg, J. (1956). The measure of linguistic diversity. *Language, 32,* 109-115.

Greenhaus, J., Parasuruman, S., & Wormley, W. (1990). Effect of race on organizational experiences, job performance evaluations, and career outcomes. *Academy of Management Journal, 33,* 64-86.

Greenleaf, E. (1992). Improving rating scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research, 29,* 176-188.

Hambrick, D., Cho, T., & Chen, M. (1996). The influence of top management team heterogeneity on firms' competitive moves. *Administrative Science Quarterly, 41,* 659-684.

Harrison, D. A., & Klein, K. J. (2007). What's the difference? Diversity constructs as *separation*, *variety*, or *disparity* in organizations. *Academy of Management Review, 32,* 1199-1228.

Harrison, D. A., Price, K. H., & Bell, M. P. (1998). Beyond relational demography: Time and the effects of surface- and deep-level diversity on work group cohesion. *Academy of Management Journal, 41,* 96-107.

Harrison D. A., Price, K. H., Gavin, J. H., & Florey, A. T. (2002). Time, teams, and task performance: Changing effects of surface- and deep-level diversity on group functioning. *Academy of Management Journal, 45*, 1029-45.

Harrison, D. A., & Sin, H.-S. (2005). What is diversity and how should it be measured? In A. M. Konrad, P. Prasad, & J. K. Pringle (Eds.), *Handbook of workplace diversity*: 191-216.

Haward, R., Amir, Z., Borrill, C., Dawson, J. F., Scully, J., West, M. & Sainsbury, R. (2003). Breast cancer teams: the impact of constitution, new cancer workload, and methods of operation on their effectiveness. *British Journal of Cancer, 89,* 15-22.

Helmert, F. R. (1876). Die Berechnung des wahrscheinlichen Beobachtungsfehlers aus den ersten Potenzeni der Differenzen gleichgenauer directer Beobachtungen. *Astronomische Nachrichten, 88,* 127-132.

Hill, M. O. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology, 54, 4*27–431.

Hogg, M., & Abrams, D. (1988). *Social identification*. London: Routledge.

Hogg, M. A., & Terry, D. J. (2000). Social identity and self-categorisation processes in organizational contexts. *Academy of Management Review*, 25: 121-140.

Hooper, D. T., & Martin, R. (2008). Beyond personal Leader–Member Exchange (LMX) quality: The effects of perceived LMX variability on employee reactions. *The Leadership Quarterly, 19,* 20-30.

Horwitz, S. K., & Horwitz, I. B. (2007). The effects of team diversity on team outcomes: A meta-analytic review of team demography. *Journal of Management*, *33,* 987-1015.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P., & V. Gupta (Eds.). (2004). *Leadership, culture, and organizations: The GLOBE study of 62 societies*. Thousand Oaks, CA: Sage Publications.

Hülsheger, U. R., Anderson, N., & Salgado, J. F. (2009). Team-level predictors of innovation at work: A comprehensive meta-analysis spanning three decades of research. *Journal of Applied Psychology, 94,* 1128-1145.

Hurlbert, S. H. (1971). The nonconcept of species diversity: A critique and alternative parameters. *Ecology, 52,* 577-586.

Jackson, S. E. (1992). Team composition in organizational settings: Issues in managing an increasingly diverse work force. In *Group process & productivity*, ed. S. Worchel, W. Wood, J. A. Simpson, pp. 136-80. Newbury Park, CA: Sage

Jackson, S. E., Brett, J. F., Sessa, V. I, Cooper, D. M., Julin, J. A., & Peyronnin, K. (1991). Some differences make a difference: Individual dissimilarity and group heterogeneity as correlates of recruitment, promotions and turnover. *Journal of Applied Psychology, 76,* 675-689.

Jackson, S. E., Joshi, A., & Erhardt, N. L. (2003). Recent research on team and organizational diversity: SWOT analysis and implications. *Journal of Management, 29,* 801-830.

Jackson, S. E., May, K. E., & Whitney, K. (1995). Understanding the dynamics of diversity in decision-making teams. In R. A. Guzzo & E. Salas (ed.), *Team effectiveness and decision making in organizations*. San Francisco: Jossey-Bass.

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology, 69*, 85-98.

James, L. R., Demaree, R. G., & Wolf, G. (1993). r_{wg}: An assessment of within-group interrater agreement. *Journal of Applied Psychology, 78*, 306-309.

Jehn, K. A., & Bezrukova, K. (2004). A field study of group diversity, workgroup context, and performance. *Journal of Organizational Behavior*, *25,* 703-729.

Jehn, K. A., & Mannix, E. A. (2001). The dynamic nature of conflict: A longitudinal study of intra-group conflict and group performance. *Academy of Management Journal*, 44: 238-251.

Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why differences make a difference: A field study of diversity, conflict, and performance in workgroups. *Administrative Science Quarterly, 44*, 741-763.

Joshi, A., Liao, H., & Jackson, S. E. (2006). Cross-level effects of workplace diversity on sales performance and pay. *Academy of Management Journal, 49,* 459-481.

Joshi, A., & Roh, H. (2009). The role of context in work team diversity research: A meta-analytic review. *Academy of Management Journal, 52,* 599-627.

Judge, T., & Ferris, G. (1993). Social context of performance evaluation decisions. *Academy of Management Journal, 36,* 80-105.

Kearney, E., & Gebert, D. (2009). Managing diversity and enhancing team outcomes: The promise of transformational leadership. *Journal of Applied Psychology, 94,* 77-89.

Kearney, E., Gebert, D., & Voelpel, S. C. (2009). When and how diversity benefits teams: The importance of team members' need for cognition. *Academy of Management Journal, 52,* 581-598.

Kendall, M., & Stuart, A. (1977). *The Advanced Theory of Statistics, vol. 1 (4th edition).* New York: Macmillan.

Kent, R., & McGrath, J. (1969).Task and group characteristics as factors affecting group performance. *Journal of Experimental Social Psychology, 5,* 429-440.

Kimura, K. (1994). A micro-macro linkage in the measurement of inequality: Another look at the GINI coefficient. *Quality and Quantity, 28,* 83–97.

Kirchmeyer, C. (1995). Managing the work-non-work boundary : an assessment of organizational responses, *Human Relations , 48,* 515-536.

Knight, D., Pearce, C. Smith, K. G., Olian, J., Sims, H., Smith, K. A. & Flood, P. (1999). Top management team diversity, group process, and strategic consensus. *Strategic Management Journal, 20,* 445-465.

Kooij-De Bode, H. J. M., van Knippenberg, D., & van Ginkel, W. P. (2008). Ethnic diversity and distributed information in group decision making: The importance of information elaboration. *Group Dynamics, 12,* 307-320.

Kotz, S., Balakrishnan, N., & Johnson, N. L. (2000). *Continuous multivariate distributions*. New York: Wiley-Interscience.

Kozlowski, S. W. J., & Hattrup, K. (1992). A disagreement about within-group agreement: Disentangling issues of consistency versus consensus. *Journal of Applied Psychology, 77*, 161-167.

Kramer, R. (1991). Intergroup relations and organizational dilemmas: The role of categorisation processes. in B. Staw and L. Cummings (Eds.), *Research in Organizational Behavior* (Vol. 13, pp. 191-228). Greenwich, CT: JAI Press.

Kvålseth, T. O. (1995). Coefficients of variation for nominal and ordinal categorical data. *Perceptual and Motor Skills, 80,* 843-847.

Lai, D., Huang, J., Risser, J. M., & Kapadia, A. S. (2008). Statistical properties of generalized Gini coefficient with application to health inequality measurement. *Social Indicators Research, 87,* 249-258.

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods, 9,* 202-220.

Lau, D. C., & Murnighan, J. K. (1998). Demographic diversity and faultlines : The compositional dynamics of organizational groups. *Academy of Management Review, 23,* 325–340.

LeBreton, J. M., James, L. R., & Lindell, M. K. (2005). Recent issues regarding r_{WG}, r^*_{WG}, $r_{WG(J)}$, and $r_{*WG(J)}$. *Organizational Research Methods*, *8*, 128-138.

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11,*815-852.

Leik, R. K. (1966). A measure of ordinal consensus. *Pacific Sociological Review, 9,* 85-90.

Leonard, J. S., Levine, D. I., & Joshi, A. (2004). Do birds of a feather shop together? The effects on performance of employees' similarity with one another and with customers. *Journal of Organizational Behavior, 25,* 731-754.

Lieberson, S. (1964). An extension of Greenberg's linguistic diversity measure. *Language, 40,* 526-531.

Lieberson, S. (1969). Measuring population diversity. *American Sociological Review, 34,* 850-862.

Lindell, M. K., Brandt, C. J., & Whitney, D. J. (1999). A revised index of interrater agreement for multi-item ratings of a single target. *Applied Psychological Measurement, 23*, 127-135.

Linn, R. L. (1968). Range restriction problems in the use of self-selected groups for test validation. *Psychological Bulletin, 69,* 69-73.

Linville, P., & Jones, E. (1980). Polarized appraisals of outgroup members. *Journal of Personality and Social Psychology, 38,* 689-703.

Mannix, E., & Neale, M. A. (2005). What differences make a difference? *Psychological Sciences in the Public Interest, 6,* 31-55.

Marcel, J. J. (2009). Why top management team characteristics matter when employing a chief operating officer: A strategic contingency perspective. *Strategic Management Journal, 30,* 647-658.

Mayo, M., Pastor, J. C., & Meindl, J. R. (1996). The effect of group heterogeneity on the self-perceived efficacy of group leaders. *Leadership Quarterly, 7,* 265-284.

McLeod, P., & Lobel, S. (1992). The effects of ethnic diversity on idea generation in small groups. Paper presented at the Annual Academy of Management Meeting, Las Vegas, Nevada.

Milliken, F. J., & Martins, L. L. (1996). Searching for common threads: Understanding the multiple effects of diversity in organizational groups. *Academy of Management Review, 21,* 402-33.

Mills, M. (2009). Globalization and inequality. *European Sociological Review, 25,*1-8.

Mohammed, S., & Angell, L. C. (2004). Surface- and deep-level diversity in workgroups: Examining the moderating effects of team orientation and team process on relationship conflict. *Journal of Organizational Behavior, 25,* 1015-1039.

Moscovici, S., & Nemeth, C. J. (1974). *Social psychology: Classic and contemporary integrations (7th ed.)*. Oxford: Rand McNally.

Mueller, J. H., Schuessler, K. F., & Costner, H. L. (1977). *Statistical reasoning in sociology (3rd ed.)*. Boston: Houghton Mifflin.

Murray, A. I. (1989). Top management group heterogeneity and firm performance. *Strategic Management Journal, 10*: 125-141.

Nemeth, C. J. (1986). Differential contributions of majority and minority influence. *Psychological Review, 93,* 23-32.

Neuman, G. A., Wagner, S. H., & Christiansen, N. D. (1999). The relationship between work-team personality composition and the job performance of teams. *Group & Organization Management, 24,* 28-45.

Newcomb, T. M. (1961). *The acquaintance process*. New York: Holt, Rinehart & Winston.

Newman, D. A., & Sin, H.-S. (2009). How do missing data bias estimates of within-group agreement? Sensitivity of SD_{WG}, CV_{WG}, $r_{WG(J)}$, $r_{WG(J)}$*, and ICC to systematic non-response. *Organizational Research Methods, 12,* 113-147.

Nickell, S., & van Reenen, J. (2001). *Technological innovation and performance in the United Kingdom*. Centre for Economic Performance Discussion Papers. London: London School of Economics.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory (3rd Ed.)*. New YorkL McGraw-Hill.

Office for National Statistics (1991). *UK Census 1991*. London: Office for National Statistics.

Office for National Statistics (2001). *UK Census 2001*. London: Office for National Statistics.

Olson, B. J., Parayitam, S., & Bao, Y. (2007). Strategic decision making: The effects of cognitive diversity, conflict, and trust on decision outcomes. *Journal of Management*, *33,* 196-222.

O'Reilly, C. A., Caldwell, D. F., & Barnett, W. P. (1989). Work group demography, social integration, and turnover. *Administrative Science Quarterly, 34,* 21-37.

O’Reilly, C., Snyder, R., & Boothe, J. (1993). Effects of executive team demography on organizational change. In G. Huber and W. Glick (Eds.), *Organizational Change and Redesign* (pp. 147-175). New York: Oxford University Press.

O’Reilly, C., Williams, K., & Barsade, S. (1998). Group demography and innovation: Does diversity help? In E. Mannix & M. Neale (Eds.), *Research in the management of groups and teams* (Vol. 1). Greenwich, CT: JAI Press.

Pearsall, J. (Ed.) (1998). *The New Oxford Dictionary of English*. Oxford: Oxford University Press.

Pearsall, M. J., Ellis, A. P. J., & Evans, J. M. (2008). Unlocking the effects of gender faultlines on team creativity: Is activation the key? *Journal of Applied Psychology, 93,* 225-234.

Pegels, C. C., Song, Y. I., & Yang, B. (2000). Management heterogeneity, competitive interaction groups, and firm performance. *Strategic Management Journal*, *21,* 911-923.

Pelled, L. H. (1996). Relational demography and perceptions of group conflict and performance: A field investigation. *International Journal of Conflict Management, 7,* 230-246.

Pelled, L., Cummings, T., & Kizilos, M. (2000). The influence of organizational demography on customer-oriented prosocial behavior: An exploratory investigation. *Journal of Business Research, 47,* 209-216.

Pelled, L. H., Eisenhardt, K. M., & Xin, K. R. (1999). Exploring the black box: An analysis of work group diversity, conflict, and performance. *Administrative Science Quarterly, 44*, 1-28.

Pelled, L. H., Ledford, G. E., & Mohrman, S. A. (1999). Demographic dissimilarity and workplace inclusion. *Journal of Management Studies, 36,* 1013-1031.

Pfeffer, J. (1983). Organizational demography. In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior, vol. 5*, 299-357. Greenwich, CT: JAI Press.

Pfeffer, J., & O'Reilly, C. (1987). Hospital demography and turnover among nurses. *Industrial relations, 26,* 158–173.

Phillips, K., Mannix, E., Neale, M., & Gruenfeld, D. (2004). Diverse groups and information sharing: The effect of congruent ties. *Journal of Experimental Social Psychology, 40,* 497–510.

Phillips, D. J., & Zuckerman, E. W. (2001). Middle-status conformity: Theoretical restatement and empirical demonstration in two markets. *American Journal of Sociology, 107,* 379–429.

Pielou, E. C. (1977). *Mathematical Ecology*. New York: Wiley.

Priem, R. L., Lyon, D. W., & Dess, G. G. (1999). Inherent limitations of demographic proxies in top management team heterogeneity research. *Journal of Management, 25,* 935-953.

Pritchard, R. D., Jones, S. D., Roth, P. L., Stuebing, K. K., & Ekeberg, S. E. (1988). Effects of group feedback, goal setting, and incentives on organizational productivity. *Journal of Applied Psychology, 73,* 337-358.

Randel, A. E. (2002). Identity salience: A moderator of the relationship between group gender composition and work group conflict. *Journal of Organizational Behavior, 23,* 749-766.

Reagans, R., & Zuckerman, E. W. (2001). Networks, diversity, and productivity: The social capital of corporate R&D teams. *Organization Science*, *12,* 502-517.

Reagans R. E., Zuckerman E. W., & McEvily, B. (2004). How to make the team: Social networks vs demography as criteria for designing effective teams. *Administrative Science Quarterly, 49,* 101-133.

Rentsch, J. R., & Klimoski, R. J. (2001). Why do "great minds" think alike? Antecedents of team member schema arrangement. *Journal of Organizational Behavior, 22,* 107-120.

Richardson, J., Naylor, P. B., Dawson, J. F., & West, M. A. (2010). *Effective Multi-professional Team Working in Mental Health Care*. Paper presented at the 3rd Qualitative Research in Mental Health conference, Nottingham, August 2010.

Roberson, Q. M., Sturman, M. C., & Simons, T. L. (2007). Does the measure of dispersion matter in multilevel research? A comparison of the relative performance of dispersion indices. *Organizational Research Methods, 10,* 564-588.

Roulston, M. S. (1999). Estimating the errors on measured entropy and mutual information. *Physica D, 125,* 285-294.

Russell, C. J., & Bobko, P. (1992). Moderated regression analysis and Likert scales: Too coarse for comfort. *Journal of Applied Psychology, 77,* 336-342.

Sacco, J. M., & Schmitt, N. (2005). A dynamic multilevel model of demographic diversity and misfit effects. *Journal of Applied Psychology, 90,* 203-231.

Sarin, S., & O'Connor, G. C. (2009). First among equals: The effect of team leader characteristics on the internal dynamics of cross-functional product development teams. *Journal of Product Innovation Management, 26,* 188-205.

Schippers, M. C., den Hartog, D. N., Koopman, P. L., & Wienk, J. A. (2003). Diversity and team outcomes: The moderating effects of outcome interdependence and group longevity and the mediating effect of reflexivity. *Journal of Organizational Behavior*, *24,* 779-882.

Schneider, B. (1987). The people make the place. *Personnel Psychology, 40*: 437-453.

Schneider, B., & Goldstein, H. W. (1995). The ASA framework: An update. *Personnel Psychology*, 48: 747-773.

Schneider, B., Salvaggio, A. N., & Subirats, M. (2002). Climate strength: A new direction for climate research. *Journal of Applied Psychology, 87*: 220-229.

Shannon, C. (1948). A mathematical theory of communications. *Bell System Technical Journal, 27*, 397-423, 623-656.

Shaw, J. B. (2004). The development and analysis of a measure of group faultlines. *Organizational Research Methods, 7,* 66-114.

Shin, S. J., & Zhou, J. (2007). When is educational specialization heterogeneity related to creativity in research and development teams? Transformational leadership as a moderator. *Journal of Applied Psychology, 92,* 1709-1721.

Simpson, E. H. (1949). Measurement of diversity. *Nature*, 163: 688.

Smith, K. G., Smith, K. A., Olian, J. D., Sims, Jr., H. P., O'Bannon, D. P., and Scully, J. (1994). Top management team demography and process: The role of social integration and communication. *Administrative Science Quarterly, 39,* 412-438.

Somech, A. (2006). The effects of leadership style and team process on performance and innovation in functionally heterogeneous teams. *Journal of Management*, *32,* 132-157.

Somech, A. & Drach-Zahavy, A. (2007) Schools as team-based organizations: A structure-process outcomes approach. *Group Dynamics: Theory, Research and Practice, 11,* 305-320.

Sommers, S. R. (2006). On racial diversity and group decision making: Identifying multiple effects of racial composition on jury deliberations. *Journal of Personality and Social Psychology, 90,* 597–612.

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 3,* 271–295.

Stewart, G. L. (2006). A meta-analytic review of the relationship between team design features and team performance. *Journal of Management*, *32,* 29-54.

Swift, T. A., & West, M. A. (1998). *Reflexivity and group processes: Research and practice*. Sheffield: The ESRC Centre for Organization and Innovation.

Taagapera, M., & Ray, J. L. (1977). A generalized index of concentration. *Sociological Methods & Research, 5,* 367-384.

Tajfel, H. (1978). Social categorization, social identity and social comparison. In H. Tajfel (ed.), *Differentiation between social groups: Studies in the social psychology of intergroup relations (pp. 61-76)*. London: Academic Press.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. Austin (eds.), *Psychology of intergroup relations (pp. 7-24)*. Chicago: Nelson-Hall.

Teachman, J. D. (1980). Analysis of population diversity. *Sociological Methods and Research, 8*, 341-362.

Thatcher, S. M. B., Jehn, K. A., & Zanutto, E. (2003). Cracks in diversity research: The effects of diversity faultlines on conflict and performance. *Group Decision and Negotiation, 12,* 217-241.

The Macmillan Directory of Breast Cancer Services in the UK: 2nd Edition (1996). *Raising Standards in Breast Cancer Care*. London: Cancer Relief Macmillan Fund.

Thornburg, T. (1991). Group size and member diversity influence on creative performance. *Journal of Creative Behavior, 25,* 324-333.

Tihanyi, L., Ellstrand, A. E., Daily, C. M., & Dalton, D. R. (2000). Composition of the top management team and firm international diversification. *Journal of Management, 26,* 1157-1177.

Timmerman, T. J. (2000). Racial diversity, age diversity, interdependence, and team performance. *Small Group Research, 31,* 592-606.

Tindale, R. S., Smith, C. M., Thomas, L., S., Filkins, J., & Sheffey, S. (1996). Shared representations and asymmetric social influence processes in small groups. In E. Witte & J. H. Davis (Eds.), *Understanding group behavior: Consensual action by small groups, Volume 1 (pp. 81-103)*. Mahwah, NJ: Lawrence Erlbaum Associates.

Tonidandel, S., Avery, D. R., Bucholtz, B. E., & McKay, P. F. (2008). An Alternative Explanation for the Asymmetrical Effects in Relational Demography Research. *Personnel Psychology, 61,* 617-633.

Triandis, H., Hall, E., & Ewen, R. (1965). Member heterogeneity and dyadic creativity. *Human Relations, 18,* 33-55.

Tsui, A. S., Egan, T. D., & O'Reilly, C. (1992). Being different: Relational demography and organizational attachment. *Administrative Science Quarterly, 37*, 549-579.

Tsui, A. S., Egan, T. D., & Xin, K. R. (1995). Diversity in organizations: Lessons from demography research. In M. M. Chemers, S. Oskamp, & M. A. Costanzo (eds.), *Diversity in Organizations: New Perspectives for a Changing Workplace (pp. 191–219)*. Thousand Oaks, CA: Sage.

Tsui, A. S., & O'Reilly, C. (1989). Beyond simple demographic effects: The importance of relational demography in supervisor-subordinate dyads. *Academy of Management Journal*, *32*, 402-423.

Turner, J. (1985). Social categorization and the self concept: A social cognitive theory of group behavior. *Advances in group processes (Vol. 2, pp. 77–121)*. Greenwich, CT: JAI Press.

Turner, J. C., Hogg, M. A., Oakes, P. J., & Peicher, S. D. (1987). *Rediscovering the Social Group: A Self-Categorisation Theory.* New York: Basil Blackwell.

Tziner, A., & Eden, D. (1985). Effects of crew composition on crew performance; Does the whole equal the sum of its parts? *Journal of Applied Psychology, 70,* 85-93.

Van der Vegt, G., & Bunderson, J. S. (2005). Learning and performance in multidisciplinary teams: The importance of collective team identification. *Academy of Management Journal, 48,* 532-47.

Van der Vegt, G. S., & Janssen, O. (2003). Joint impact of interdependence and group diversity on innovation. *Journal of Management, 29,* 729-51.

Van der Vegt, G. S., Van de Vliert, E., & Oosterhof, A. (2003). Information dissimilarity and organizational citizenship behavior: The role of intrateam interdependence and team identification. *Academy of Management Journal, 46,* 715-27.

van Dijk, H., van Engen, M. L, & van Knippenberg, D. (under review). Work group diversity and group performance: A meta-analysis. *Organizational Behavior and Human Decision Processes*.

Van Knippenberg, D., Dawson, J. F., West, M. A., & Homan, A. (2011). Diversity faultlines, shared objectives, and top management team performance. *Human Relations, 64,* 307-336.

Van Knippenberg, D., De Dreu, C. K. W., & Homan, A. C. (2004). Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology* 89: 1008-22.

Van Knippenberg, D., Haslam, S. A., & Platow, M. J. (2007). Unity through diversity: Value-in-diversity beliefs, work group diversity, and group identification. *Group Dynamics Theory, Research and Practice, 11,* 207-222.

Van Knippenberg, D., & Schippers, M. C. (2007). Work group diversity. *Annual Review of Psychology, 58,* 515-541.

Van Vianen, A. E. M., & De Dreu, C. K. W. (2001). Personality in teams: Its relationship to social cohesion, task cohesion, and team performance. *European Journal of Work and Organizational Psychology, 10,* 97-120.

Wagner, G. W., Pfeffer, J., & O'Reilly, C. A. (1984). Organizational demography and turnover. *Administrative Science Quarterly, 29,* 74-92.

Watson, W. E., Johnson, L., & Zgourides, G. D. (2002). The influence of ethnic diversity on leadership, group process, and performance: An examination of learning teams. *International Journal of Intercultural Relations, 26,* 1-16.

Webber, S. S., & Donahue, L. M. (2001). Impact of highly and less job-related diversity on work group cohesion and performance: A meta-analysis. *Journal of Management*, *27,* 141-162.

Wegge, J., Roth, C., Neubach, B., Schmidt, K., & Kanfer, R. (2008). Age and gender diversity as determinants of performance and health in a public organization: The role of task complexity and group size. *Journal of Applied Psychology, 93,* 1301-1313.

Weick, K. E., & Roberts, K. H. (1993). Collective mind in organizations: Heedful interrelating on flight decks. *Administrative Science Quarterly, 38,* 357-381.

West, M. A., Markiewicz, L., & Dawson, J. F. (2005) *The Aston Team Performance Inventory: User's Guide.* London: ASE.

White, M. J. (1986). Segregation and diversity measures in population distribution. *Population Index, 52,* 198-221.

Wiersema, M., & Bantel, K. (1992). Top management team demography and corporate strategic change. *Academy of Management Journal, 35,* 91-121.

Wiersema, M., & Bantel, K. (1993). Top management team turnover as an adaptation mechanism: The role of the environment. *Strategic Management Journal, 14,* 485-504.

Wiersema, M. F., & Bird, A. (1993). Organizational demography in Japanese firms: Group heterogeneity, individual dissimilarity, and top management team turnover. *Academy of Management Journal*, *36,* 996-1025.

Wilcox, A. R. (1973). Indices of qualitative variation and political measurement. *The Western Political Quarterly, 26,* 325-343.

Williams, H., & Meân, L. J. (2004). Measuring gender composition in work groups: A comparison of existing methods. *Organizational Research Methods, 7,* 456-474.

Williams, K., & O'Reilly, C. (1998). Demography and diversity in organizations: A review of 40 years of research. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior, Vol. 20*, 77-140. Greenwich, CT: JAI Press.

Zajac, E., Golden, B., & Shortell, S. (1991). New organizational forms for enhancing innovation: The case of internal corporate joint ventures. *Management Science, 37,* 170-184.

Zellmer-Bruhn, M. E., Maloney, M. M., Bhappu, A. D., & Salvador, R. B. (2008). When and how do differences matter? An exploration of perceived similarity in teams. *Organizational Behavior and Human Decision Processes, 107,* 41-59.

Zenger, T., & Lawrence, B. (1989). Organizational demography: The differential effects of age and tenure distributions on technical communications. *Academy of Management Journal, 32,* 353-376.

APPENDIX 1

Proof of the equivalence of the two adjusted forms of Blau's Index

Harrison and Klein (2007), and Biemann and Kearney (2010), both give the formula for the adjusted version of Blau's index as:

$$D_{B-adj} = 1 - \sum_{i=1}^{k} \frac{n_i(n_i - 1)}{n(n-1)}$$

It follows that:

$$D_{B-adj} = 1 - \sum_{i=1}^{k} \frac{n_i(n_i - 1)}{n(n-1)}$$

$$= 1 - \frac{1}{n(n-1)} \sum_{i=1}^{k} ({n_i}^2 - n_i)$$

$$= 1 - \frac{1}{n(n-1)} \left(\sum_{i=1}^{k} {n_i}^2 - \sum_{i=1}^{k} n_i \right)$$

$$= 1 - \frac{1}{n(n-1)} \left(\sum_{i=1}^{k} {n_i}^2 - n \right)$$

$$= \frac{n-1}{n-1} - \frac{1}{n(n-1)} \sum_{i=1}^{k} {n_i}^2 + \frac{1}{n-1}$$

$$= \frac{n}{n-1} - \frac{1}{n(n-1)} \sum_{i=1}^{k} {n_i}^2$$

$$= \frac{n}{n-1} \left(1 - \sum_{i=1}^{k} \frac{{n_i}^2}{n^2} \right)$$

$$= \frac{n}{n-1} \left(1 - \sum_{i=1}^{k} {p_i}^2 \right)$$

QED ■

APPENDIX 2

Derivation of computational formula for D_{SR}

$$D_{SR} = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j<i} |x_i - x_j|$$

Let us order x_i such that $x_1 \leq x_2 \leq \cdots \leq x_N$. Then

$$D_{SR} = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j<i} (x_i - x_j)$$

$$= \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j<i} [(x_i - \bar{x}) + (\bar{x} - x_j)]$$

$$= \frac{2}{N(N-1)} \left[\sum_{i=1}^{N} (i-1)(x_i - \bar{x}) + \sum_{i=1}^{N} \sum_{j<i} (\bar{x} - x_j) \right]$$

$$= \frac{2}{N(N-1)} \left[\sum_{i=1}^{N} (ix_i - x_i - i\bar{x} + \bar{x}) + \sum_{j=1}^{N} (N-j)(\bar{x} - x_j) \right]$$

$$= \frac{2}{N(N-1)} \left[\left(\sum_{i=1}^{N} (i-1)x_i \right) - \left(\sum_{j=1}^{N} (N-j)x_j \right) \right]$$

$$= \frac{2}{N(N-1)} \left[2 \sum_{i=1}^{N} ix_i - (N+1) \sum_{i=1}^{N} x_i \right]$$

$$= \frac{4}{N(N-1)} \sum_{i=1}^{N} ix_i - \frac{2(N+1)N\bar{x}}{N(N-1)}$$

$$= \frac{4}{N(N-1)} \sum_{i=1}^{N} ix_i - \frac{2(N+1)\bar{x}}{(N-1)}$$

QED ■

APPENDIX 3

Proof of results from Chapter 5 about Blau's and Teachman's indices

A3.1 Blau's index

Theorem

If group 1, comprising n people, has $n_1, n_2, \dots, n_k$ members in each of k categories, and group 2 (also comprising n people) has the same makeup except for one member being in a $(k + 1)$th category, then group 2 has a Blau's index greater than or equal to that of group 1.

Proof

Let us assume, without loss of generality, that group 2 has $n_1, n_2, \dots, n_{k-1}, n_k - 1, n_{k+1}$ members in each of the k categories, where $n_{k+1} = 1$ as this is the "new" category.

Let us denote Blau's index for the two groups as Blau_1 and Blau_2 respectively.

Then

$$\text{Blau}_1 = 1 - \sum_{i=1}^{k} \left(\frac{n_i}{n}\right)^2$$

$$\text{Blau}_2 = 1 - \sum_{i=1}^{k-1} \left(\frac{n_i}{n}\right)^2 - \left(\frac{n_k - 1}{n}\right)^2 - \left(\frac{1}{n}\right)^2$$

$$= 1 - \left[\sum_{i=1}^{k-1} \frac{{n_i}^2}{n^2} + \frac{{n_k}^2 - 2n_k + 1}{n^2} + \frac{1}{n^2}\right]$$

$$= 1 - \sum_{i=1}^{k} {p_i}^2 + \frac{2(n_k - 1)}{n^2}$$

If $n_k = 1$ then this is the same as Blau_1 (note that in this case group 1 and group 2 have identical relative frequencies). If $n_k > 1$ then $\frac{2(n_k-1)}{n^2} > 0$ and $\text{Blau}_2 > \text{Blau}_1$.

QED ■

A3.2 Teachman's index

Theorem

If group 1, comprising n people, has $n_1, n_2, \dots, n_k$ members in each of k categories, and group 2 (also comprising n people) has the same makeup except for one member being in a $(k + 1)$th category, then group 2 has a Teachman's index greater than or equal to that of group 1.

Proof

Let us assume, without loss of generality, that group 2 has $n_1, n_2, \dots, n_{k-1}, n_k - 1, n_{k+1}$ members in each of the k categories, where $n_{k+1} = 1$ as this is the "new" category.

Let us denote Teachman's index for the two groups as Teachman_1 and Teachman_2 respectively.

Then

$$\text{Teachman}_1 = -\sum_{i=1}^{k} \frac{n_i}{n} \ln\left(\frac{n_i}{n}\right)$$

$$\text{Teachman}_2 = -\sum_{i=1}^{k-1} \frac{n_i}{n} \ln\left(\frac{n_i}{n}\right) - \frac{n_k - 1}{n} \ln\left(\frac{n_k - 1}{n}\right) - \frac{1}{n} \ln\left(\frac{1}{n}\right)$$

Let us assume that $\text{Teachman}_1 > \text{Teachman}_2$. Then:

$$-\sum_{i=1}^{k} \frac{n_i}{n} \ln\left(\frac{n_i}{n}\right) > -\sum_{i=1}^{k-1} \frac{n_i}{n} \ln\left(\frac{n_i}{n}\right) - \frac{n_k - 1}{n} \ln\left(\frac{n_k - 1}{n}\right) - \frac{1}{n} \ln\left(\frac{1}{n}\right)$$

$$-\frac{n_k}{n} \ln\left(\frac{n_k}{n}\right) > -\frac{n_k - 1}{n} \ln\left(\frac{n_k - 1}{n}\right) - \frac{1}{n} \ln\left(\frac{1}{n}\right)$$

$$\frac{n_k}{n} \ln\left(\frac{n_k}{n}\right) < \frac{n_k - 1}{n} \ln\left(\frac{n_k - 1}{n}\right) + \frac{1}{n} \ln\left(\frac{1}{n}\right)$$

$$\frac{n_k}{n} \ln\left(\frac{n_k}{n}\right) - \frac{1}{n} \ln\left(\frac{1}{n}\right) < \frac{n_k - 1}{n} \ln\left(\frac{n_k - 1}{n}\right)$$

$$\frac{n_k}{n}(\ln n_k - \ln n) + \frac{1}{n}\ln n < \frac{n_k - 1}{n}\ln(n_k - 1) - \frac{n_k - 1}{n}\ln n$$

$$\frac{n_k}{n}\ln n_k + \frac{1 - n_k}{n}\ln n < \frac{n_k - 1}{n}\ln(n_k - 1) + \frac{1 - n_k}{n}\ln n$$

$$\frac{n_k}{n}\ln n_k < \frac{n_k - 1}{n}\ln(n_k - 1)$$

$$\frac{n_k}{n}[\ln n_k - \ln(n_k - 1)] < -\frac{1}{n}\ln(n_k - 1)$$

If $n_k = 1$ then this inequality is inadmissible, since $\ln(n_k - 1)$ does not exist. However, in this case groups 1 and 2 have identical relative frequencies and therefore $\text{Teachman}_1 = \text{Teachman}_2$.

If $n_k = 2$ then $\ln(n_k - 1) = 0$, and hence this states that $\frac{n_k}{n}\ln n_k < 0$, which is impossible.

If $n_k > 2$ then $\ln(n_k - 1) > 0$, implying:

$$\frac{n_k}{n}[\ln n_k - \ln(n_k - 1)] < 0$$

$$\ln n_k - \ln(n_k - 1) < 0$$

$$\ln n_k < \ln(n_k - 1)$$

which is also impossible, since the natural logarithm is a monotonically increasing function.

Therefore whatever the value of n_k we have a contradiction, implying that Teachman_1 cannot be greater than Teachman_2. Therefore we have $\text{Teachman}_1 \leq \text{Teachman}_2$.

QED ■

APPENDIX 4

Effect sizes from real data sets

Notes:
Figures shown in tables are standardised regression coefficients (with *p* values in parentheses)
For diversity type columns, "S" = separation, "V" = variety

Table A4.1: Primary Health Care Teams – Continuous Variables

Diversity attribute & Outcome	Diversity type	D_{SR}/D_{SI}	$D_{RR}/D_{RI}/D_{RO}$	D_{VR}/D_{VI}	D_{DR}	SD	AD_M	Variance	Euclidean	Gini	D_{DI}
Age & TCI - participation	S	-.13 (.21)	.00 (.98)	-.11 (.27)	-.10 (.31)	-.09 (.38)	-.15 (.13)	-.11 (.30)	-.10 (.32)	.00 (.99)	
Age & TCI - support for innovation	S	-.08 (.46)	.04 (.69)	-.06 (.55)	-.05 (.65)	-.04 (.70)	-.10 (.31)	-.07 (.49)	-.05 (.62)	.07 (.51)	
Age & TCI - objectives	S	-.08 (.43)	.04 (.72)	-.08 (.47)	-.05 (.64)	-.05 (.63)	-.09 (.39)	-.08 (.46)	-.06 (.58)	.10 (.38)	
Age & TCI - task orientation	S	-.13 (.20)	-.07 (.55)	-.07 (.51)	-.15 (.15)	-.13 (.21)	-.16 (.13)	-.15 (.13)	-.13 (.19)	-.01 (.92)	
Age & Reflexivity	S	-.16 (.11)	.01 (.91)	-.18 (.08)	-.12 (.23)	-.11 (.29)	-.19 (.06)	-.12 (.23)	-.13 (.23)	.00 (.99)	
Age & Self-report effectiveness	S	-.21 (.04)	.00 (.97)	-.18 (.07)	-.13 (.20)	-.15 (.15)	-.21 (.03)	-.17 (.11)	-.16 (.11)	.17 (.12)	
Age & Externally rated effectiveness	S	-.08 (.47)	.01 (.94)	-.12 (.27)	.02 (.83)	-.03 (.82)	-.03 (.77)	-.06 (.62)	-.03 (.80)	.19 (.10)	
Team climate & Self-report effectiveness	S	.04 (.57)	.11 (.10)	-.09 (.15)	.12 (.10)	.10 (.13)	.05 (.40)	.08 (.18)	.09 (.16)	.32 (.00)	.14 (.08)
Team climate & Externally rated effectiveness	S	.08 (.53)	.15 (.26)	.18 (.15)	.11 (.46)	.10 (.45)	.08 (.53)	.13 (.32)	.10 (.46)	.27 (.22)	.11 (.50)
Tenure & Self-report innovation	V	-.25 (.01)	-.12 (.26)	-.22 (.04)	-.19 (.06)	-.22 (.03)	-.24 (.02)	-.22 (.03)	-.23 (.03)	.11 (.32)	
Tenure & Innovation: magnitude	V	-.04 (.71)	.10 (.32)	-.19 (.05)	-.02 (.85)	.03 (.80)	-.01 (.94)	.02 (.83)	.02 (.85)	.11 (.30)	
Tenure & Innovation: radicalness	V	-.06 (.50)	.07 (.47)	-.19 (.04)	-.04 (.68)	.00 (.97)	-.05 (.59)	-.01 (.94)	-.01 (.89)	.05 (.63)	
Tenure & Innovation: novelty	V	.04 (.68)	.21 (.04)	-.06 (.54)	.06 (.50)	.10 (.32)	.03 (.78)	.07 (.43)	.08 (.37)	.14 (.18)	
Tenure & Innovation: impact	V	-.02 (.86)	.21 (.04)	-.18 (.07)	.05 (.59)	.08 (.43)	.00 (.98)	.06 (.56)	.06 (.51)	.16 (.14)	

Table A4.2: Primary Health Care Teams – Nominal Variables

Diversity attribute & Outcome	Diversity type	D_{SN}	D_{RN}	D_{VN}	Blau/IQV
Sex & TCI - participation	S	.10 (.34)	.10 (.35)	.11 (.28)	.11 (.26)
Sex & TCI - support for innovation	S	.01 (.92)	.04 (.73)	.04 (.74)	.04 (.73)
Sex & TCI - objectives	S	.00 (.99)	.06 (.56)	.03 (.77)	.02 (.83)
Sex & TCI - task orientation	S	-.03 (.76)	-.02 (.85)	-.02 (.84)	-.02 (.85)
Sex & Reflexivity	S	.04 (.73)	.10 (.35)	.06 (.55)	.05 (.60)
Sex & Self-report effectiveness	S	-.04 (.72)	.01 (.95)	-.01 (.94)	-.01 (.93)
Sex & Externally rated effectiveness	S	.03 (.81)	.03 (.82)	.03 (.78)	.03 (.77)
Functional background & Self-report innovation	V	-.05 (.61)	.02 (.85)	-.02 (.87)	-.03 (.79)
Functional background & Innovation: magnitude	V	-.22 (.03)	.20 (.07)	.04 (.66)	-.06 (.52)
Functional background & Innovation: radicalness	V	-.09 (.37)	.22 (.04)	.11 (.27)	.02 (.81)
Functional background & Innovation: novelty	V	.14 (.16)	.44 (.00)	.34 (.00)	.24 (.01)
Functional background & Innovation: impact	V	-.03 (.79)	.39 (.00)	.26 (.01)	.14 (.16)

Table A4.3: Community Mental Health Teams – Continuous Variables

Diversity attribute & Outcome	Diversity type	D_{SR}/D_{SI}	$D_{RR}/D_{RI}/D_{RO}$	D_{VR}/D_{VI}	D_{DR}	SD	AD_M	Variance	Euclidean	Gini	D_{DI}
Age & TCI - participation	S	.02 (.81)	.12 (.23)	-.07 (.48)	.09 (.31)	.05 (.56)	.05 (.62)	.04 (.66)	.05 (.60)	.12 (.23)	
Age & TCI - support for innovation	S	.03 (.78)	.15 (.14)	.02 (.84)	.12 (.22)	.06 (.51)	.03 (.73)	.04 (.64)	.05 (.57)	.18 (.07)	
Age & TCI - objectives	S	-.04 (.67)	.05 (.65)	-.07 (.48)	.01 (.95)	-.02 (.85)	-.04 (.66)	-.04 (.69)	-.03 (.79)	.05 (.64)	
Age & TCI - task orientation	S	-.05 (.60)	.07 (.48)	-.05 (.63)	.01 (.92)	-.03 (.77)	-.05 (.57)	-.05 (.62)	-.04 (.70)	.14 (.16)	
Age & Reflexivity	S	-.02 (.86)	.11 (.30)	.00 (.98)	.05 (.64)	.01 (.90)	-.02 (.84)	-.01 (.91)	.00 (.98)	.11 (.27)	
Age & Self-report effectiveness	S	.06 (.54)	.19 (.06)	.09 (.33)	.13 (.17)	.09 (.34)	.05 (.61)	.06 (.51)	.08 (.39)	.14 (.16)	
Team climate & Self-report effectiveness	S	.08 (.18)	.07 (.22)	.02 (.77)	.10 (.14)	.08 (.15)	.09 (.13)	.08 (.14)	.08 (.15)	.22 (.19)	.12 (.12)
Tenure & Self-report innovation	V	-.09 (.34)	.05 (.61)	.01 (.90)	-.02 (.81)	-.07 (.49)	-.10 (.29)	-.07 (.46)	-.08 (.43)	.17 (.10)	
Tenure & Innovation: magnitude	V	-.01 (.92)	.10 (.27)	.16 (.06)	.02 (.80)	.00 (.98)	.00 (.98)	.00 (.98)	.00 (.97)	.18 (.04)	
Tenure & Innovation: radicalness	V	.03 (.70)	.17 (.08)	.16 (.07)	.09 (.31)	.05 (.55)	.03 (.71)	.05 (.55)	.05 (.59)	.21 (.02)	
Tenure & Innovation: novelty	V	.01 (.91)	.13 (.20)	.09 (.35)	.08 (.43)	.03 (.73)	-.02 (.83)	.03 (.78)	.02 (.80)	.22 (.03)	
Tenure & Innovation: impact	V	-.02 (.84)	.09 (.39)	.09 (.33)	.06 (.56)	.00 (1.00)	-.03 (.79)	.00 (.96)	-.01 (.95)	.24 (.02)	

Table A4.4: Community Mental Health Teams – Nominal Variables

Diversity attribute & Outcome	Diversity type	D_{SN}	D_{RN}	D_{VN}	Blau/IQV
Sex & TCI - participation	S	.02 (.80)	-.04 (.64)	.06 (.54)	.06 (.53)
Sex & TCI - support for innovation	S	-.07 (.47)	-.02 (.80)	-.04 (.67)	-.04 (.65)
Sex & TCI - objectives	S	-.17 (.07)	-.09 (.34)	-.16 (.10)	-.16 (.09)
Sex & TCI - task orientation	S	-.06 (.56)	-.05 (.62)	-.03 (.74)	-.03 (.72)
Sex & Reflexivity	S	-.08 (.40)	-.07 (.47)	-.06 (.53)	-.06 (.53)
Sex & Self-report effectiveness	S	-.11 (.24)	.04 (.67)	-.08 (.40)	-.09 (.33)
Functional background & Self-report innovation	V	-.11 (.26)	-.01 (.93)	-.07 (.50)	-.09 (.36)
Functional background & Innovation: magnitude	V	.06 (.48)	.14 (.18)	.11 (.24)	.10 (.28)
Functional background & Innovation: radicalness	V	.03 (.76)	.13 (.21)	.09 (.33)	.07 (.46)
Functional background & Innovation: novelty	V	.11 (.24)	.28 (.01)	.19 (.07)	.14 (.16)
Functional background & Innovation: impact	V	.07 (.45)	.24 (.03)	.17 (.10)	.10 (.29)

Table A4.5: Breast Cancer Care Teams – Continuous Variables

Diversity attribute & Outcome	Diversity type	D_{SR}/D_{SI}	$D_{RR}/D_{RI}/D_{RO}$	D_{VR}/D_{VI}	D_{DR}	SD	AD_M	Variance	Euclidean	Gini	D_{DI}
Age & TCI - participation	S	-.06 (.59)	-.04 (.75)	-.06 (.60)	-.09 (.49)	-.05 (.69)	-.03 (.79)	-.08 (.50)	-.05 (.69)	-.13 (.37)	
Age & TCI - support for innovation	S	-.13 (.29)	-.11 (.44)	-.16 (.18)	-.13 (.32)	-.10 (.43)	-.05 (.69)	-.14 (.28)	-.10 (.45)	-.05 (.74)	
Age & TCI - objectives	S	.03 (.80)	.12 (.36)	.07 (.53)	.04 (.77)	.06 (.66)	.02 (.89)	.06 (.61)	.05 (.70)	.05 (.74)	
Age & TCI - task orientation	S	-.10 (.40)	-.03 (.83)	-.08 (.51)	-.10 (.45)	-.07 (.56)	-.08 (.54)	-.11 (.37)	-.08 (.55)	.07 (.67)	
Age & Reflexivity	S	-.17 (.17)	-.12 (.39)	-.26 (.03)	-.13 (.32)	-.12 (.33)	-.13 (.30)	-.13 (.31)	-.13 (.32)	.10 (.51)	
Age & Self-report effectiveness	S	-.16 (.16)	-.11 (.43)	-.14 (.24)	-.12 (.34)	-.14 (.27)	-.11 (.36)	-.16 (.17)	-.13 (.27)	.20 (.19)	
Team climate & Self-report effectiveness	S	-.01 (.92)	.00 (.97)	-.04 (.63)	-.01 (.91)	-.02 (.80)	.00 (.97)	.00 (1.00)	-.02 (.84)	.38 (.08)	.00 (.96)
Tenure & Self-report innovation	V	-.08 (.48)	-.08 (.57)	-.12 (.29)	-.07 (.58)	-.06 (.60)	-.04 (.75)	-.08 (.48)	-.06 (.60)	.05 (.73)	
Tenure & Innovation: magnitude	V	.02 (.90)	.00 (1.00)	-.04 (.73)	.02 (.88)	.03 (.81)	.08 (.55)	-.01 (.97)	.04 (.75)	.05 (.77)	
Tenure & Innovation: radicalness	V	.11 (.43)	.02 (.91)	.10 (.40)	.07 (.62)	.08 (.54)	.11 (.40)	.04 (.76)	.10 (.49)	-.09 (.56)	
Tenure & Innovation: novelty	V	.08 (.54)	.06 (.70)	-.03 (.84)	.09 (.54)	.10 (.48)	.13 (.33)	.07 (.64)	.11 (.45)	-.07 (.67)	
Tenure & Innovation: impact	V	-.01 (.93)	-.12 (.40)	-.06 (.59)	.02 (.90)	-.01 (.93)	.08 (.55)	-.03 (.82)	.01 (.95)	.17 (.26)	

Table A4.6: Breast Cancer Care Teams – Nominal Variables

Diversity attribute & Outcome	Diversity type	D_{SN}	D_{RN}	D_{VN}	Blau/IQV
Sex & TCI - participation	S	.11 (.34)	.16 (.14)	.12 (.28)	.11 (.35)
Sex & TCI - support for innovation	S	.22 (.07)	.22 (.06)	.24 (.04)	.23 (.05)
Sex & TCI - objectives	S	.03 (.80)	-.12 (.31)	.05 (.69)	.07 (.55)
Sex & TCI - task orientation	S	.21 (.09)	.20 (.07)	.24 (.04)	.23 (.05)
Sex & Reflexivity	S	.15 (.22)	.11 (.34)	.18 (.13)	.18 (.13)
Sex & Self-report effectiveness	S	.05 (.70)	.16 (.14)	.09 (.44)	.07 (.55)
Functional background & Self-report innovation	V	.05 (.78)	-.03 (.87)	.04 (.82)	.16 (.46)
Functional background & Innovation: magnitude	V	-.20 (.27)	-.16 (.35)	-.13 (.50)	-.08 (.69)
Functional background & Innovation: radicalness	V	-.40 (.02)	-.35 (.04)	-.46 (.01)	-.56 (.01)
Functional background & Innovation: novelty	V	-.21 (.27)	-.16 (.36)	-.20 (.29)	-.24 (.27)
Functional background & Innovation: impact	V	-.30 (.08)	.01 (.95)	-.07 (.69)	-.16 (.44)

Table A4.7: Top Management Teams – Continuous and Ordinal Variables

Diversity attribute & Outcome	Diversity type	D_{SR}/D_{SI}	$D_{RR}/D_{RI}/D_{RO}$	D_{VR}/D_{VI}	D_{DR}	SD	AD_M	Variance	Euclidean	Gini
Age & TCI - participation	S	-.05 (.71)	-.03 (.81)	.06 (.63)	-.09 (.47)	-.05 (.66)	-.07 (.57)	-.04 (.75)	-.05 (.66)	-.04 (.74)
Age & TCI - support for innovation	S	-.05 (.66)	.04 (.76)	.07 (.58)	-.05 (.68)	-.04 (.75)	-.09 (.48)	-.03 (.84)	-.05 (.71)	.09 (.49)
Age & TCI - objectives	S	-.04 (.73)	-.06 (.63)	-.03 (.84)	-.11 (.40)	-.06 (.62)	-.05 (.68)	-.06 (.66)	-.06 (.63)	-.11 (.44)
Age & TCI - task orientation	S	-.12 (.36)	-.08 (.55)	-.02 (.85)	-.11 (.38)	-.11 (.38)	-.14 (.28)	-.13 (.31)	-.11 (.37)	.08 (.56)
Age & Company productivity	S	-.28 (.06)	-.21 (.19)	-.21 (.16)	-.30 (.06)	-.26 (.10)	-.31 (.05)	-.25 (.10)	-.26 (.09)	-.20 (.28)
Age & Company profit	S	-.31 (.04)	-.35 (.02)	-.26 (.07)	-.41 (.01)	-.34 (.03)	-.38 (.01)	-.34 (.02)	-.35 (.02)	-.36 (.04)
Team climate & Company productivity	S	-.22 (.24)	-.17 (.42)	-.13 (.46)	-.14 (.51)	-.19 (.32)	-.20 (.26)	-.16 (.41)	-.19 (.31)	.16 (.53)
Team climate & Company profit	S	-.31 (.08)	-.41 (.04)	-.21 (.22)	-.32 (.12)	-.33 (.07)	-.31 (.08)	-.25 (.17)	-.33 (.07)	-.10 (.69)
Tenure & Innovation in products	V	-.10 (.54)	-.01 (.96)	-.12 (.50)	-.02 (.91)	-.06 (.74)	-.09 (.60)	-.08 (.65)	-.07 (.70)	.26 (.19)
Tenure & Innovation in production technology	V	.14 (.40)	.24 (.16)	.12 (.49)	.20 (.24)	.22 (.18)	.22 (.21)	.27 (.11)	.22 (.18)	.23 (.25)
Tenure & Innovation in production techniques/processes	V	.01 (.95)	.15 (.39)	-.05 (.77)	.12 (.48)	.09 (.59)	.06 (.75)	.13 (.45)	.08 (.62)	.35 (.08)
Tenure & Innovation in work design	V	.15 (.36)	.22 (.19)	.12 (.47)	.24 (.14)	.18 (.28)	.19 (.26)	.12 (.49)	.18 (.29)	.50 (.01)
Tenure & Innovation in HRM	V	.23 (.16)	.24 (.16)	.19 (.27)	.30 (.07)	.27 (.10)	.30 (.08)	.30 (.07)	.28 (.10)	.31 (.13)
Educational background & Innovation in products	V	-.07 (.70)	-.12 (.53)		-.11 (.57)	-.07 (.73)	-.03 (.87)	-.12 (.52)	-.06 (.76)	.08 (.74)
Educational background & Innovation in production technology	V	.25 (.17)	.31 (.09)		.34 (.07)	.31 (.10)	.30 (.11)	.29 (.11)	.31 (.10)	.62 (.00)
Educational background & Innovation in production techniques/processes	V	.22 (.23)	.24 (.19)		.26 (.17)	.24 (.20)	.25 (.18)	.24 (.18)	.25 (.20)	.45 (.03)
Educational background & Innovation in work design	V	.01 (.94)	.09 (.61)		.09 (.65)	.03 (.87)	-.02 (.93)	.05 (.79)	.03 (.88)	.18 (.40)
Educational background & Innovation in HRM	V	-.02 (.90)	.07 (.72)		.04 (.85)	.01 (.94)	.02 (.93)	.04 (.81)	.02 (.94)	.08 (.71)

(table continues)

(Table A4.7 continued)

Diversity attribute & Outcome	Diversity type	D_{DI}	a_{wg}	D_{SO}	D_{VO}	l	LOV
Age & TCI - participation	S						
Age & TCI - support for innovation	S						
Age & TCI - objectives	S						
Age & TCI - task orientation	S						
Age & Company productivity	S						
Age & Company profit	S						
Team climate & Company productivity	S	-.12 (.60)	.18 (.34)				
Team climate & Company profit	S	-.31 (.14)	.29 (.11)				
Tenure & Innovation in products	V						
Tenure & Innovation in production technology	V						
Tenure & Innovation in production techniques/processes	V						
Tenure & Innovation in work design	V						
Tenure & Innovation in HRM	V						
Educational background & Innovation in products	V	-.12 (.54)	.06 (.73)	-.22 (.24)	-.04 (.80)	-.22 (.23)	-.19 (.30)
Educational background & Innovation in production technology	V	.34 (.06)	-.26 (.15)	.30 (.09)	.39 (.02)	.30 (.09)	.27 (.12)
Educational background & Innovation in production techniques/processes	V	.26 (.17)	-.22 (.22)	.23 (.20)	.33 (.04)	.23 (.19)	.22 (.21)
Educational background & Innovation in work design	V	.10 (.60)	-.01 (.98)	.12 (.50)	.10 (.54)	.13 (.46)	.15 (.41)
Educational background & Innovation in HRM	V	.04 (.83)	-.02 (.91)	.07 (.68)	.10 (.54)	.07 (.68)	.09 (.63)

Table A4.8: Top Management Teams – Nominal Variables

Diversity attribute & Outcome	Diversity type	D_{SN}	D_{RN}	D_{VN}	Blau/IQV
Sex & TCI - participation	S	.13 (.29)	.16 (.18)	.15 (.22)	.14 (.24)
Sex & TCI - support for innovation	S	.23 (.06)	.25 (.04)	.25 (.04)	.25 (.05)
Sex & TCI - objectives	S	.02 (.85)	.04 (.76)	.04 (.76)	.04 (.76)
Sex & TCI - task orientation	S	.10 (.42)	.14 (.26)	.13 (.30)	.12 (.32)
Sex & Company productivity	S	.01 (.95)	-.04 (.83)	-.01 (.97)	.00 (.99)
Sex & Company profit	S	-.23 (.15)	-.33 (.04)	-.28 (.09)	-.26 (.11)
Functional background & Innovation in products	V	.24 (.27)	.18 (.27)	.22 (.25)	.23 (.28)
Functional background & Innovation in production technology	V	-.05 (.82)	.21 (.22)	.26 (.16)	.26 (.21)
Functional background & Innovation in production techniques/processes	V	-.12 (.60)	.17 (.30)	.09 (.63)	.04 (.85)
Functional background & Innovation in work design	V	-.05 (.80)	.12 (.48)	.05 (.80)	.02 (.92)
Functional background & Innovation in HRM	V	-.11 (.60)	.14 (.40)	.09 (.65)	.03 (.87)

Table A4.9: ATPI Nursing Teams – Continuous and Ordinal Variables

Diversity attribute & Outcome	Diversity type	D_{SR}/D_{SI}	$D_{RR}/D_{RI}/D_{RO}$	D_{VR}/D_{VI}	D_{DR}	SD	AD_M	Variance	Euclidean	Gini
Age & ATPI: objectives	S	.06 (.59)	.06 (.59)		-.02 (.84)	.01 (.95)	.05 (.64)	.06 (.56)	.01 (.92)	-.07 (.48)
Age & ATPI: reflexivity	S	-.04 (.71)	.02 (.87)		-.08 (.43)	-.03 (.76)	-.01 (.95)	-.02 (.86)	-.03 (.76)	-.08 (.40)
Age & ATPI: participation	S	-.03 (.75)	-.05 (.70)		-.08 (.41)	-.07 (.55)	-.02 (.88)	-.02 (.81)	-.06 (.57)	-.11 (.26)
Age & ATPI: task focus	S	.10 (.31)	.06 (.64)		-.01 (.95)	.01 (.95)	.04 (.74)	.04 (.73)	.01 (.94)	-.02 (.82)
Age & ATPI: team conflict	S	-.05 (.57)	-.03 (.78)		-.04 (.67)	-.05 (.68)	-.11 (.33)	-.05 (.62)	-.05 (.65)	-.03 (.79)
Age & ATPI: satisfaction	S	-.01 (.95)	-.08 (.52)		-.11 (.28)	-.09 (.40)	-.03 (.81)	-.05 (.63)	-.09 (.43)	-.11 (.31)
Age & ATPI: attachment	S	-.03 (.74)	-.03 (.83)		-.10 (.32)	-.07 (.50)	-.04 (.74)	-.01 (.91)	-.07 (.52)	-.13 (.18)
Age & ATPI: effectiveness	S	-.06 (.56)	-.17 (.16)		-.06 (.54)	-.17 (.12)	-.12 (.26)	-.08 (.46)	-.17 (.12)	-.08 (.47)
Age & ATPI: inter-team relationships	S	.10 (.34)	.12 (.33)		.01 (.93)	.02 (.84)	.04 (.73)	.08 (.48)	.02 (.85)	-.04 (.68)

(table continues)

(Table A4.9: continued)

Diversity attribute & Outcome	Diversity type	D_{DI}	a_{wg}	D_{SO}	D_{VO}	*l*	*LOV*
Age & ATPI: objectives	S	-.06 (.54)	-.04 (.68)	.03 (.80)	.14 (.23)	.04 (.69)	.04 (.70)
Age & ATPI: reflexivity	S	-.10 (.29)	.03 (.75)	-.03 (.82)	.07 (.57)	-.02 (.85)	-.02 (.85)
Age & ATPI: participation	S	-.11 (.26)	.05 (.66)	-.06 (.62)	.02 (.89)	-.04 (.68)	-.03 (.76)
Age & ATPI: task focus	S	-.06 (.51)	.00 (.98)	.02 (.85)	.12 (.32)	.03 (.78)	.03 (.78)
Age & ATPI: team conflict	S	-.02 (.81)	.05 (.63)	-.06 (.60)	-.08 (.52)	-.06 (.61)	-.10 (.38)
Age & ATPI: satisfaction	S	-.14 (.17)	.10 (.34)	-.08 (.49)	-.01 (.96)	-.06 (.57)	-.04 (.71)
Age & ATPI: attachment	S	-.12 (.22)	.03 (.82)	-.05 (.64)	.05 (.66)	-.04 (.74)	-.03 (.76)
Age & ATPI: effectiveness	S	-.06 (.56)	.06 (.57)	-.16 (.16)	-.08 (.52)	-.13 (.24)	-.12 (.28)
Age & ATPI: inter-team relationships	S	-.04 (.70)	-.06 (.61)	.04 (.75)	.18 (.15)	.05 (.63)	.01 (.92)

Table A4.10: ATPI Nursing Teams – Nominal Variables

Diversity attribute & Outcome	Diversity type	D_{SN}	D_{RN}	D_{VN}	Blau/IQV
Sex & ATPI: objectives	S	-.06 (.56)	-.08 (.45)	-.07 (.48)	-.07 (.51)
Sex & ATPI: reflexivity	S	-.13 (.15)	-.16 (.13)	-.13 (.18)	-.13 (.20)
Sex & ATPI: participation	S	-.10 (.29)	-.12 (.24)	-.12 (.22)	-.12 (.23)
Sex & ATPI: task focus	S	-.14 (.15)	-.11 (.29)	-.12 (.22)	-.12 (.21)
Sex & ATPI: team conflict	S	.10 (.24)	.03 (.79)	.09 (.37)	.11 (.29)
Sex & ATPI: satisfaction	S	-.22 (.02)	-.27 (.01)	-.25 (.01)	-.23 (.02)
Sex & ATPI: attachment	S	.00 (.99)	-.04 (.70)	-.03 (.80)	-.02 (.81)
Sex & ATPI: effectiveness	S	-.01 (.91)	-.02 (.88)	-.03 (.78)	-.03 (.76)
Sex & ATPI: inter-team relationships	S	-.10 (.33)	-.10 (.35)	-.09 (.39)	-.08 (.41)
Ethnic background & ATPI: objectives	S	-.11 (.29)	-.05 (.62)	-.07 (.45)	-.09 (.38)
Ethnic background & ATPI: reflexivity	S	-.10 (.29)	-.01 (.95)	-.04 (.72)	-.05 (.59)
Ethnic background & ATPI: participation	S	-.24 (.01)	-.19 (.06)	-.21 (.03)	-.22 (.02)
Ethnic background & ATPI: task focus	S	-.23 (.02)	-.13 (.18)	-.16 (.10)	-.18 (.06)
Ethnic background & ATPI: team conflict	S	.21 (.02)	.15 (.14)	.19 (.06)	.20 (.05)
Ethnic background & ATPI: satisfaction	S	-.13 (.18)	-.12 (.22)	-.13 (.17)	-.14 (.16)
Ethnic background & ATPI: attachment	S	-.07 (.47)	-.07 (.46)	-.09 (.36)	-.09 (.37)
Ethnic background & ATPI: effectiveness	S	-.07 (.45)	-.07 (.46)	-.10 (.32)	-.10 (.30)
Ethnic background & ATPI: inter-team relationships	S	.00 (.97)	.11 (.29)	.07 (.48)	.05 (.61)

Table A4.11: ATPI Mental Health Teams – Continuous and Ordinal Variables

Diversity attribute & Outcome	Diversity type	D_{SR}/D_{SI}	$D_{RR}/D_{RI}/D_{RO}$	D_{VR}/D_{VI}	D_{DR}	SD	AD_M	Variance	Euclidean	Gini
Age & ATPI: objectives	S	.19 (.03)	.15 (.11)		.26 (.00)	.15 (.10)	.12 (.19)	.15 (.08)	.15 (.10)	.33 (.00)
Age & ATPI: reflexivity	S	.19 (.04)	.18 (.06)		.25 (.00)	.15 (.10)	.16 (.08)	.15 (.09)	.15 (.10)	.30 (.00)
Age & ATPI: participation	S	.23 (.01)	.25 (.01)		.27 (.00)	.21 (.02)	.18 (.04)	.18 (.04)	.21 (.02)	.37 (.00)
Age & ATPI: task focus	S	.17 (.05)	.20 (.03)		.19 (.03)	.16 (.08)	.14 (.13)	.11 (.20)	.15 (.09)	.30 (.00)
Age & ATPI: team conflict	S	-.18 (.04)	-.24 (.01)		-.15 (.09)	-.20 (.03)	-.14 (.11)	-.17 (.05)	-.19 (.03)	-.22 (.01)
Age & ATPI: satisfaction	S	.20 (.02)	.19 (.04)		.25 (.00)	.16 (.08)	.14 (.12)	.16 (.07)	.16 (.08)	.34 (.00)
Age & ATPI: attachment	S	.15 (.09)	.25 (.01)		.18 (.04)	.21 (.02)	.17 (.06)	.17 (.05)	.20 (.02)	.27 (.00)
Age & ATPI: effectiveness	S	.10 (.27)	.20 (.03)		.16 (.07)	.17 (.05)	.14 (.13)	.12 (.18)	.17 (.06)	.26 (.00)
Age & ATPI: inter-team relationships	S	.08 (.36)	.12 (.21)		.18 (.04)	.10 (.27)	.08 (.36)	.10 (.27)	.10 (.27)	.30 (.00)
Age & CMHT effectiveness	S	.09 (.30)	.10 (.28)		.16 (.07)	.07 (.42)	.07 (.47)	.06 (.47)	.08 (.40)	.29 (.00)

(table continues)

(Table A4.11 continued)

Diversity attribute & Outcome	Diversity type	D_{DI}	a_{wg}	D_{SO}	D_{VO}	*l*	*LOV*
Age & ATPI: objectives	S	.28 (.00)	-.17 (.06)	.17 (.07)	.17 (.07)	.17 (.07)	.17 (.06)
Age & ATPI: reflexivity	S	.26 (.00)	-.17 (.06)	.17 (.06)	.20 (.03)	.17 (.06)	.19 (.03)
Age & ATPI: participation	S	.26 (.00)	-.20 (.02)	.24 (.01)	.27 (.00)	.22 (.01)	.24 (.01)
Age & ATPI: task focus	S	.18 (.04)	-.13 (.14)	.18 (.05)	.21 (.02)	.15 (.09)	.17 (.05)
Age & ATPI: team conflict	S	-.14 (.12)	.19 (.03)	-.21 (.02)	-.25 (.01)	-.20 (.02)	-.19 (.04)
Age & ATPI: satisfaction	S	.25 (.00)	-.18 (.05)	.19 (.03)	.22 (.01)	.18 (.04)	.19 (.03)
Age & ATPI: attachment	S	.19 (.03)	-.20 (.02)	.23 (.01)	.25 (.01)	.21 (.02)	.23 (.01)
Age & ATPI: effectiveness	S	.17 (.05)	-.16 (.08)	.19 (.04)	.18 (.04)	.16 (.08)	.19 (.04)
Age & ATPI: inter-team relationships	S	.23 (.01)	-.13 (.14)	.13 (.16)	.15 (.10)	.12 (.18)	.13 (.15)
Age & CMHT effectiveness	S	.19 (.03)	-.09 (.33)	.11 (.23)	.14 (.13)	.09 (.30)	.11 (.20)

Table A4.12: ATPI Mental Health Teams – Nominal Variables

Diversity attribute & Outcome	Diversity type	D_{SN}	D_{RN}	D_{VN}	Blau/IQV
Sex & ATPI: objectives	S	-.01 (.94)	-.02 (.85)	.02 (.79)	.03 (.72)
Sex & ATPI: reflexivity	S	.02 (.78)	-.02 (.85)	.01 (.90)	.02 (.81)
Sex & ATPI: participation	S	-.06 (.53)	.01 (.92)	.01 (.94)	.01 (.95)
Sex & ATPI: task focus	S	-.04 (.63)	.01 (.95)	.03 (.73)	.04 (.69)
Sex & ATPI: team conflict	S	.00 (.99)	-.12 (.20)	-.08 (.39)	-.06 (.47)
Sex & ATPI: satisfaction	S	-.12 (.19)	-.05 (.56)	-.07 (.45)	-.07 (.44)
Sex & ATPI: attachment	S	-.01 (.88)	.09 (.33)	.08 (.34)	.08 (.36)
Sex & ATPI: effectiveness	S	-.03 (.73)	.10 (.29)	.08 (.35)	.08 (.38)
Sex & ATPI: inter-team relationships	S	-.09 (.29)	-.10 (.26)	-.08 (.36)	-.07 (.42)
Sex & CMHT effectiveness	S	-.02 (.85)	.03 (.76)	.01 (.92)	.00 (.98)
Ethnic background & ATPI: objectives	S	-.03 (.75)	-.04 (.63)	-.04 (.64)	-.04 (.66)
Ethnic background & ATPI: reflexivity	S	-.01 (.93)	-.02 (.81)	-.01 (.87)	-.02 (.86)
Ethnic background & ATPI: participation	S	-.16 (.06)	-.14 (.12)	-.16 (.06)	-.16 (.06)
Ethnic background & ATPI: task focus	S	-.03 (.74)	-.04 (.69)	-.04 (.69)	-.04 (.68)
Ethnic background & ATPI: team conflict	S	.14 (.12)	.14 (.12)	.14 (.12)	.13 (.15)
Ethnic background & ATPI: satisfaction	S	-.15 (.09)	-.14 (.12)	-.15 (.08)	-.15 (.08)
Ethnic background & ATPI: attachment	S	-.14 (.11)	-.11 (.21)	-.12 (.18)	-.11 (.21)
Ethnic background & ATPI: effectiveness	S	-.14 (.10)	-.10 (.28)	-.11 (.20)	-.11 (.21)
Ethnic background & ATPI: inter-team relationships	S	-.08 (.37)	-.05 (.58)	-.06 (.50)	-.06 (.48)
Ethnic background & CMHT effectiveness	S	-.15 (.09)	-.14 (.10)	-.15 (.08)	-.15 (.09)
Functional background & ATPI: innovation	V	-.18 (.04)	-.14 (.12)	-.17 (.06)	-.17 (.05)

Table A4.13: MSc Student Teams – Nominal Variables

Diversity attribute & Outcome	Diversity type	D_{SN}	D_{RN}	D_{VN}	Blau/IQV
Sex & Mutual trust (t1)	S	-.28 (.02)	-.42 (.00)	-.32 (.01)	-.29 (.02)
Sex & Mutual trust (t2)	S	-.11 (.37)	-.16 (.21)	-.13 (.31)	-.11 (.37)
Sex & Mutual trust (t3)	S	-.12 (.41)	-.14 (.37)	-.14 (.36)	-.13 (.38)
Sex & Attendance (t1)	S	-.12 (.35)	-.20 (.12)	-.14 (.27)	-.12 (.36)
Sex & Attendance (t2)	S	-.04 (.73)	-.08 (.54)	-.05 (.68)	-.04 (.73)
Sex & Attendance (t3)	S	-.02 (.89)	-.01 (.94)	-.02 (.90)	-.02 (.90)
Sex & Commitment (t1)	S	-.27 (.03)	-.27 (.05)	-.28 (.03)	-.28 (.03)
Sex & Commitment (t2)	S	.03 (.81)	.00 (.99)	.02 (.88)	.02 (.85)
Sex & Group performance (t1)	S	-.05 (.79)	.02 (.93)	-.04 (.85)	-.05 (.79)
Sex & Group performance (t2)	S	-.10 (.45)	-.16 (.24)	-.11 (.38)	-.10 (.45)
Sex & Group performance (t3)	S	-.20 (.16)	-.26 (.09)	-.23 (.13)	-.21 (.16)
Nationality & Mutual trust (t1)	S	-.35 (.00)	-.42 (.00)	-.40 (.00)	-.37 (.00)
Nationality & Mutual trust (t2)	S	-.39 (.00)	-.36 (.01)	-.39 (.00)	-.41 (.00)
Nationality & Mutual trust (t3)	S	-.25 (.22)	-.12 (.50)	-.18 (.35)	-.25 (.24)
Nationality & Attendance (t1)	S	-.36 (.00)	-.29 (.04)	-.34 (.01)	-.37 (.00)
Nationality & Attendance (t2)	S	-.28 (.02)	-.22 (.11)	-.27 (.05)	-.29 (.03)
Nationality & Attendance (t3)	S	.38 (.06)	.33 (.07)	.36 (.06)	.40 (.06)
Nationality & Commitment (t1)	S	-.29 (.02)	-.26 (.05)	-.29 (.02)	-.30 (.02)
Nationality & Commitment (t2)	S	-.16 (.21)	-.13 (.36)	-.15 (.27)	-.16 (.21)
Nationality & Group performance (t1)	S	-.20 (.67)	-.21 (.41)	-.21 (.51)	-.22 (.64)
Nationality & Group performance (t2)	S	-.24 (.06)	-.22 (.12)	-.24 (.08)	-.24 (.06)
Nationality & Group performance (t3)	S	-.23 (.27)	-.06 (.75)	-.14 (.47)	-.23 (.28)
Language & Mutual trust (t1)	S	-.27 (.03)	-.20 (.14)	-.25 (.06)	-.28 (.03)
Language & Mutual trust (t2)	S	-.38 (.00)	-.25 (.06)	-.33 (.01)	-.40 (.00)
Language & Mutual trust (t3)	S	.13 (.51)	.19 (.29)	.18 (.36)	.13 (.52)
Language & Attendance (t1)	S	-.29 (.02)	-.22 (.11)	-.26 (.05)	-.30 (.02)
Language & Attendance (t2)	S	-.36 (.00)	-.29 (.04)	-.33 (.01)	-.37 (.00)
Language & Attendance (t3)	S	-.18 (.37)	-.06 (.75)	-.10 (.59)	-.19 (.38)
Language & Commitment (t1)	S	-.12 (.37)	-.04 (.79)	-.09 (.54)	-.12 (.38)
Language & Commitment (t2)	S	-.22 (.08)	-.13 (.34)	-.19 (.16)	-.23 (.08)
Language & Group performance (t1)	S	.07 (.83)	-.08 (.73)	-.02 (.94)	.06 (.87)
Language & Group performance (t2)	S	-.27 (.04)	-.13 (.34)	-.22 (.11)	-.28 (.03)
Language & Group performance (t3)	S	-.02 (.94)	.22 (.21)	.13 (.49)	-.01 (.95)

Table A4.14: Business Game Teams – Continuous Variables

Diversity attribute & Outcome	Diversity type	D_{SR}/D_{SI}	$D_{RR}/D_{RI}/D_{RO}$	D_{VR}/D_{VI}	D_{DR}	SD	AD_M	Variance	Euclidean	Gini
Age & Mutual trust (t1)	S	.18 (.17)	.19 (.15)	.19 (.16)	.18 (.16)	.19 (.16)	.18 (.17)	.19 (.15)	.18 (.16)	.00 (1.00)
Age & Mutual trust (t2)	S	.12 (.37)	.11 (.40)	.13 (.37)	.12 (.39)	.12 (.38)	.12 (.37)	.11 (.38)	.12 (.38)	-.05 (.75)
Age & Mutual trust (t3)	S	.12 (.39)	.11 (.42)	.13 (.36)	.12 (.39)	.12 (.38)	.14 (.31)	.12 (.35)	.12 (.37)	-.08 (.60)
Age & Group performance	S	.27 (.03)	.23 (.06)	.30 (.02)	.23 (.06)	.24 (.05)	.26 (.04)	.26 (.04)	.25 (.05)	.05 (.70)
Age & Attendance (t1)	S	.05 (.70)	.06 (.68)	.08 (.58)	.04 (.76)	.06 (.67)	.05 (.68)	.12 (.34)	.06 (.67)	-.19 (.15)
Age & Attendance (t2)	S	.00 (.98)	-.02 (.89)	-.03 (.85)	-.03 (.84)	.00 (.99)	.00 (.98)	.11 (.39)	.00 (.99)	-.24 (.12)
Age & Attendance (t3)	S	.08 (.56)	.07 (.60)	.08 (.58)	.07 (.61)	.08 (.56)	.09 (.49)	.11 (.39)	.08 (.55)	-.13 (.37)

Table A4.15: Business Game Teams – Nominal Variables

Diversity attribute & Outcome	Diversity type	D_{SN}	D_{RN}	D_{VN}	Blau/IQV
Sex & Mutual trust (t1)	S	-.10 (.60)	-.05 (.81)	-.10 (.62)	-.10 (.59)
Sex & Mutual trust (t2)	S	.09 (.63)	.12 (.57)	.10 (.62)	.09 (.64)
Sex & Mutual trust (t3)	S	.13 (.59)	.12 (.68)	.14 (.58)	.13 (.58)
Sex & Group performance	S	-.06 (.66)	-.08 (.55)	-.06 (.66)	-.06 (.69)
Sex & Attendance (t1)	S	.22 (.25)	.34 (.09)	.25 (.19)	.22 (.25)
Sex & Attendance (t2)	S	.17 (.38)	.09 (.66)	.16 (.43)	.17 (.37)
Sex & Attendance (t3)	S	.40 (.08)	.13 (.65)	.39 (.11)	.41 (.07)
Nationality & Mutual trust (t1)	S	-.10 (.47)	-.07 (.62)	-.08 (.56)	-.10 (.47)
Nationality & Mutual trust (t2)	S	-.17 (.28)	-.17 (.23)	-.17 (.26)	-.17 (.27)
Nationality & Mutual trust (t3)	S	-.04 (.81)	-.04 (.76)	-.04 (.77)	-.03 (.83)
Nationality & Group performance	S	-.16 (.20)	-.14 (.26)	-.15 (.22)	-.16 (.21)
Nationality & Attendance (t1)	S	.04 (.80)	.05 (.72)	.05 (.73)	.03 (.82)
Nationality & Attendance (t2)	S	-.17 (.26)	-.17 (.23)	-.17 (.25)	-.17 (.26)
Nationality & Attendance (t3)	S	-.02 (.92)	-.02 (.91)	-.02 (.91)	-.01 (.95)
Ethnic background & Mutual trust (t1)	S	-.16 (.27)	-.01 (.97)	-.09 (.52)	-.15 (.28)
Ethnic background & Mutual trust (t2)	S	-.22 (.14)	.02 (.86)	-.11 (.45)	-.21 (.15)
Ethnic background & Mutual trust (t3)	S	-.07 (.64)	.16 (.25)	.03 (.83)	-.07 (.65)
Ethnic background & Group performance	S	-.14 (.25)	-.04 (.73)	-.11 (.40)	-.15 (.24)
Ethnic background & Attendance (t1)	S	-.03 (.84)	.13 (.32)	.05 (.73)	-.02 (.87)
Ethnic background & Attendance (t2)	S	-.17 (.27)	-.02 (.89)	-.10 (.47)	-.16 (.27)
Ethnic background & Attendance (t3)	S	.05 (.71)	.23 (.09)	.13 (.36)	.05 (.70)

APPENDIX 5

Correlations between diversity indices

This appendix shows correlations between the different diversity indices as applied to one attribute of each data type from the Top Management Teams data set: Age, Climate, Educational Background and Functional Background. Figures shown are Pearson correlation coefficients.

Table A5.1: Correlations between Diversity Indices: Age Diversity in Top Management Teams

	1	2	3	4	5	6	7	8
1. D_{SR}								
2. D_{RR} (range)	0.86							
3. D_{VR}	0.87	0.80						
4. D_{DR} (CV)	0.94	0.91	0.79					
5. SD	0.97	0.94	0.83	0.97				
6. AD_M	0.97	0.85	0.79	0.95	0.97			
7. Variance	0.96	0.91	0.81	0.94	0.98	0.96		
8. Euclidean	0.98	0.92	0.83	0.97	1.00	0.98	0.98	
9. Gini	0.19	0.43	0.17	0.45	0.31	0.31	0.23	0.31

Table A5.2: Correlations between Diversity Indices: Climate Strength in Top Management Teams

	1	2	3	4	5	6	7	8	9
1. D_{SI}									
2. D_{RI} (range)	0.90								
3. D_{VI}	0.86	0.79							
4. D_{DI}	0.94	0.92	0.77						
5. SD	0.98	0.96	0.83	0.96					
6. CV	0.96	0.94	0.79	1.00	0.98				
7. AD_M	0.99	0.91	0.81	0.95	0.99	0.97			
8. Variance	0.96	0.93	0.80	0.95	0.98	0.96	0.97		
9. Euclidean	0.99	0.95	0.83	0.96	1.00	0.98	0.99	0.98	
10. Gini	0.39	0.63	0.29	0.63	0.51	0.60	0.45	0.48	0.50

Table A5.3: Correlations between Diversity Indices: Educational Background Diversity in Top Management Teams

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. D_{SO}													
2. D_{RO} (range)	0.83												
3. D_{VO}	0.77	0.93											
4. *l*	1.00	0.82	0.77										
5. *LOV*	0.99	0.77	0.73	0.99									
6. D_{SI}	0.81	0.89	0.80	0.80	0.79								
7. D_{DI}	0.94	0.93	0.85	0.94	0.92	0.94							
8. SD	0.85	0.95	0.84	0.84	0.81	0.97	0.96						
9. CV	0.93	0.94	0.86	0.93	0.90	0.95	1.00	0.98					
10. AD_M	0.81	0.88	0.77	0.81	0.80	0.96	0.93	0.98	0.95				
11. Variance	0.87	0.91	0.81	0.87	0.83	0.94	0.94	0.97	0.95	0.95			
12. a_{wg}	-0.72	-0.86	-0.74	-0.71	-0.67	-0.87	-0.83	-0.92	-0.85	-0.92	-0.94		
13. Euclidean	0.84	0.94	0.83	0.84	0.81	0.97	0.96	1.00	0.97	0.98	0.97	-0.92	
14. Gini	0.67	0.76	0.83	0.67	0.67	0.70	0.76	0.70	0.76	0.65	0.58	-0.47	0.70

Table A5.4: Correlations between Diversity Indices: Functional Background Diversity in Top Management Teams

	1	2	3
1. D_{SN}			
2. D_{RN} (count)	0.48		
3. D_{VN} (Teachman)	0.70	0.94	
4. Blau	0.84	0.82	0.96