Doing Q methodology: theory, method and interpretation

Simon Watts[1] and Paul Stenner[2]

[1]Nottingham Trent University, Burton Street, Nottingham NG1 4BU, UK; [2]University College London, Gower Street, London, WC1E 6BT, UK

This paper has a marked practical aspect. We wish to encourage and facilitate the use of Q methodology amongst psychologists interested in qualitative research. The paper duly answers a number of pertinent 'how to do Q' questions. Yet our primary intention is not to produce an exhaustive 'how to do Q' guide. In discussing issues of theory, method and interpretation in Q methodology, the main aim of the paper is rather to address some of the more common misunderstandings and misrepresentations that constitute obstacles to the use of one of the very first 'alternative' methods to have been developed in the context of psychology. In addressing such obstacles, the paper hopes to bring '*Why* do Q?' questions to the fore. In so doing, Q methodology will also be 'positioned' in relation to a number of other qualitative research methods, each of which currently enjoys a degree of prominence within the psychological discipline. *Qualitative Research in Psychology* 2005; **2:** 67–91

Key words: eigenvalues; factor analysis; factor extraction; factor interpretation; factor rotation; measurement; Q methodology; Q set; Q sorting; subjectivity; William Stephenson

Introduction

This paper has a marked practical aspect. We wish to encourage and facilitate the use of Q methodology amongst psychologists interested in qualitative research. This will inevitably involve our paying attention to the 'nuts and bolts' of the method. Yet our primary intention is not to produce an exhaustive 'how to do Q' guide. In discussing issues of theory, method and interpretation in Q methodology, our main aim is rather to address some of the common misunderstandings and misrepresentations

Correspondence: Simon Watts, Psychology Division, School of Social Sciences, Nottingham Trent University, Burton Street, Nottingham NG1 4BU, UK.
E-mail: simon.watts@ntu.ac.uk

© 2005 Edward Arnold (Publishers) Ltd 10.1191/1478088705qp022oa

that constitute obstacles to the use of one of the very first 'alternative' methods to have been developed in the context of psychology (Stephenson, 1935). Whilst, therefore, the paper will provide a deal of practical information about the conduct of Q methodological research (such that a number of pertinent 'how to do Q' questions will be answered), it also aims to clarify how Q methodology can be understood and positioned in relation to some of the other qualitative research methods that are currently available to the psychologist. In so doing, we hope to bring '*why* do Q?' questions to the fore.

To illustrate some typical misunderstandings of Q methodology, we turn first to a recent volume on psychological research methods by Haslam and McGarty (2003). We wish to stress that these misunderstandings are exceedingly common, and hence no special criticism should be directed at these particular authors. Indeed, it is to their credit that they discuss Q methodology at all, given that many texts on qualitative psychology do not (for example, Smith, 2003). Nonetheless, their discussion illustrates two of the most commonly committed errors. The first involves a misattribution of origins: 'In its original form participants placed 100 cards with statements about personal characteristics into piles ranging from ''not characteristic of me'' to ''very characteristic of me''' (Smith, 2003: 364). But this is not Q methodology in its 'original' form. In fact, Haslam and McGarty are instead referring to an unorthodox and somewhat problematic application of Q methodology, which is typically associated with Carl Rogers (1954), and which emerged nearly 20 years *after* William Stephenson had originally devised the methodology.

The second common error (which was also committed by Rogers) involves a separation of the two fundamental aspects of Q methodology: the Q sorting procedure (which is an original means of collecting data) and the Q pattern analysis (which is effected by means of a by-person factor analysis). As we shall see, Stephenson designed the former precisely in order to enable the legitimate application of the latter. Indeed, it was the effective *combination* of the two aspects – the Q sort and the Q (as opposed to R) technique factor analysis – that allowed Stephenson to make *subjectivity* his principle research focus (the distinction between Q and R techniques will be clarified below). Rogers gathered data in the form of Q sorts, but did not subject those sorts to a by-person factor analysis. Cyril Burt conducted by-person factor analyses on alternative forms of data (see Note 1 and Burt, 1937). Yet neither was doing Q methodology, because each employed only *one* aspect of Stephenson's procedure.

These misunderstandings alone are enough to 'muddy the waters' of Q methodology. Sadly, however, its data analytic and data collection procedures are also frequently *misrepresented*. There is, for example, a tendency for Q methodological factor analysis to be erroneously (mis)identified with its more familiar R methodological incarnation, and hence to be viewed as a 'statistical method of data reduction that identifies and combines sets of dependent variables that are measuring similar things' (McGarty and Haslam, 2003: 387). As we shall see, however, Q methodology makes no such psychometric claims. The method employs a by-person factor analysis in order to identify *groups of participants* who make sense of (and who hence Q 'sort') a pool of items in comparable ways. Nothing more complicated is at issue.

The Q sort itself also comes to be misrepresented as a passive 'response dimension' (McGarty and Haslam, 2003: 364) upon which 'statements' are simply ranked into piles, rather than as a dynamic medium through which subjectivity can be actively expressed (Stephenson, 1953). This limited and overly passive conceptualization of the Q sort once again emerges from Carl Rogers' application. Rogers used a Q sort composed of statements about personal characteristics as a means of providing a standard *measure* of self, of ideal self, and of the distance between them. Yet, for the dedicated Q methodologist, a completed Q sort indicates only that a set of items have been *differentially valued* by a specific participant according to some face valid and subjective criterion. This is not a measurement (for reasons which will become apparent). Rogers' approach may have its own merits, therefore, but the image it creates is nonetheless unhelpful for Q methodology, inasmuch as it thoroughly confuses the R methodological ambition for researcher-led objective measurements (or psychometrics) with the Q methodological pursuit of participant-led *subjective expressions* and *viewpoints*.

Such confusions could easily be avoided (and we hope this paper will make a significant contribution in this regard). We admit, however, that they result in part from the fact that Q methodology's quantitative features render it a highly unusual qualitative research method (Curt, 1994; Watts and Stenner, 2003a). The method is actually *qualiquantological* (see Stenner and Stainton Rogers, 2004). This fact alone may discourage some qualitative researchers from engaging with Q methodology, as many will have turned to qualitative methods precisely in order to escape the particular strand of quantitative logic and the associated hypothetico-deductive methods that have long since claimed the name 'science' in psychology (and in many other related disciplines). It is also easy to avoid Q methodology in the belief that it offers nothing that one of a number of forms of textual analysis cannot do better and in a more straightforwardly 'qualitative' fashion (Willig, 2001). Discourse analysis (Potter and Wetherell, 1987), narrative analysis (Crossley, 2000) and interpretative phenomenological analysis (Smith, 1996), for example, may seem to offer more palatable alternatives in this regard.

It would nonetheless be unfortunate were any qualitative researcher to reject Q methodology for either of these reasons. It makes little sense, for example, as Harré (2004) has recently suggested within these pages, to avoid all forms of mathematical and quantitative representation (or to develop a knee-jerk aversion to science) simply because they have often been poorly employed by psychologists. Qualitative methodologists have rightly challenged the currently extant 'scientific' or hypothetico-deductive approach in psychology, but (and despite its quantitative content) it is important to recognize that Stephenson's Q methodology was actually performing a similar function long before any significant qualitative tradition had been established in the discipline (Stephenson, 1935).

In fact, Q methodology was *designed* for the very purpose of challenging the dated, Newtonian logic of 'testing' that continues to predominate in psychology. It also offered an early critique of the cognitive assertion that people can properly be divided into a series of psychological 'parts'. This same critique has, of course, subsequently become a typical feature of 'constructionist' approaches in psychology (see, for example, Harré, 1999). In these

ways at least, Q methodology is a typically qualitative and a very *critical* method. The surprise, perhaps, is that it achieves its critical stance through the embrace (rather than the rejection) of many natural scientific assumptions. Indeed, having gained doctorates in both physics and psychology, Stephenson actually based his 'challenge' to the Newtonian tradition in psychology on epistemological and ontological presumptions (as well as mathematics) that were more familiarly associated with the quantum mechanics of physics (see Watts and Stenner, 2003a; Stephenson, 1936a,b, 1988/89 for a full discussion of these issues). Q methodology duly lends support to Harré's (2004: 4) recent assertion that it is 'the qualitative techniques and the metaphysical presumptions which back them that come much closer to meeting the ideals of the natural sciences'.

Neither should one mistakenly believe that Q methodology does the same job as *any* of the other textual methods available to the psychologist. Of course, one must immediately acknowledge that this is true of all the textual methods, which having different emphases necessarily perform subtly different tasks. It is apparent, for example, that some of these methods prefer a 'paradigmatic' (or thematic) form of analysis, whilst others take a more holistic or 'narrative' approach (see Polkinghorne, 1995). In the vast majority of cases, however, it is the machinations of everyday talk, the *specific* situation or person, and the generation (and explication) of individual viewpoints which constitute the primary research targets. None of this is ordinarily true of Q methodology.

This is not to deny that Q methodology can be, and has been, effectively employed as a powerful technique for single case studies of various kinds (Smith, 2001; Stephenson, 1953). Yet many qualitative methods are equally capable of such analyses and repertory grid studies provide further serious competition in the domain of the personal construct (Smith, 1995). The case-study approach is, however, not a typical application of Q methodology, especially as it has come to be employed within (primarily British and social) psychology (see Stainton Rogers *et al*., 1995). In this guise, Q methodology ordinarily adopts a multiple-participant format and is most often deployed in order to explore (and to make sense of) highly complex and socially contested *concepts* and *subject matters* from the point of view of the group of participants involved (Stainton Rogers, 1995; Watts and Stenner, 2003a). It does not do this in a thematic fashion, nor does it focus on the viewpoints of specific individuals. It should be no surprise, therefore, to find that this typical form of Q methodology disappoints when themes and/or individuals are the primary research targets.

To properly appreciate Q methodology, we need instead to recognize that it is essentially a *gestalt* procedure (Good, 2000). This gestalt emphasis means it can never 'break-up' its subject matter into a series of constituent themes (which immediately distinguishes Q from various forms of discursive or interpretative phenomenological analyses). What it can do, however, is show us the primary ways in which these themes are being interconnected or otherwise *related* by a group of participants. In other words, it can show us the particular *combinations* or *configurations* of themes which are preferred by the participant group. This openly holistic approach suggests that Q methodology is most closely related to Crossley's (2000) 'narrative analysis'.

Nevertheless, it differs from narrative analysis in at least three ways. First, and most obviously, it does not deal with participants' own discourse as such, but invites participants to engage in the unusual task of relating (in a complex and in-depth way) with a set of prepared items. This 'unusual task' evidently violates the cherished principle of 'naturalism'. One can counter this argument, however, inasmuch as the very idea of 'naturally occurring discourse' is highly problematic (Potter, 1997). There is also no *a priori* reason to assume that unfamiliar activities cannot yield useful insights. Of course, qualitative researchers should never underestimate the impact of research context on findings. Neither should they 'leap' from findings to unwarranted knowledge claims. Yet it would be equally wrong were they to misrecognize 'naturalism' (or any method or methodological rule) as some kind of absolute standard (Watts, 2000).

Secondly, Q methodology is not well suited to dealing with the unfolding temporality of narratives. Narrative analysis actively pursues this temporality and then examines the resultant stories in terms of their temporal structure (e.g., beginning, middle, end) and function. Q methodology forsakes this important form of analysis in order to pursue a 'snap shot' or temporally frozen image of a connected series of subject *positions* (or 'view-points'). It then examines these (methodologically frozen) positions in terms of their overall structure, function and implications. And thirdly, whereas a narrative analysis focuses on the narrative of specific *individuals*, Q methodology typically focuses on the range of viewpoints that are favoured (or which are otherwise 'shared') by specific *groups* of participants. In other words, the typical Q methodological study very deliberately pursues constructions and representations of a *social* kind (Moscovici, 1981).

These differences allow Q methodology to offer the psychologist a quite unique form of qualitative analysis. Indeed, in accenting the group and their shared viewpoints, this form of analysis provides an ideal (and noticeably more *macroscopic*) complement to qualitative approaches which highlight the 'theme' and/or 'the individual' in psychology. In the remainder of this paper, we want to demonstrate this complementarity at work. In so doing, we hope also to allow the interested reader to carry out effective Q methodological studies.

History, theory and statistics: 'inverting' the tradition

Q methodology was originally established via a simple adaptation of the quantitative technique known as factor analysis. This adaptation was first introduced by William Stephenson in 1935. In this introduction and subsequent papers, Stephenson indicated that the conventional factor analytic procedure could also be 'inverted':

> Factor analysis . . . is concerned with a selected population of n individuals each of whom has been measured in m tests. The $(m)(m-1)/2$ intercorrelations for these m variables are subjected to . . . factor analysis. The technique, however, can also be inverted. We begin with a population of n different tests (or essays, pictures, traits or other measurable material), each of which is . . . scaled by m individuals. The $(m)(m-1)/2$ intercorrelations are then factorised in the usual way (Stephenson, 1936b: 344–45).

Such an inversion may appear mundane, but it actually initiates a series of quite dramatic *methodological* departures from the psychological tradition (Stephenson, 1936b). First, it is the 'n different tests or measurable materials', not the participant

group, that become the study sample. Secondly, the 'variables' are no longer tests or hypothesized traits, but the various *persons* who take part in the study. In other words, persons become the variables of interest in an inverted (or 'Q') study. Such studies actively explore 'correlations between persons or whole aspects of persons' (Stephenson, 1936b: 345). As a consequence of these changes, it is also persons (not tests, traits or other types of variables) that load onto the emergent factors of an inverted factor analytic study.

These methodological changes are highly significant, then, in the sense that they distinguish Q methodology from almost every other popular psychometric technique used in psychology (these are known collectively as 'R' methodologies). They are also indicative of possible *theoretical* departures from the psychological tradition.[1] Stephenson was indeed very concerned about the unremitting dominance of hypothetico-deductive methods within psychology and the concomitant neglect of what he considered to be its proper subject matter: subjectivity. The following quote is one of a number that might have been used for illustrative purposes:

> The hallmark of sound scientific procedure nowadays, it seems, is to assert hypotheses and to confirm predictions . . . There is need, however, for care and discernment in these matters . . . Psychology . . . has by no means achieved a sophisticated theoretical status, with ideal constructs such as physics has fashioned for itself. The situations in psychology, therefore, call for an attitude of curiosity, as well as one of hypothetico-deductive logic . . . We should be making discoveries rather than testing our reasoning (Stephenson, 1953: 151).

Sadly, Stephenson's warnings went largely unheeded. Despite initial optimism and his belief that the inverted factor technique would allow the development of a new psychology based upon 'individual differences *in type*' (Stephenson, 1936a: 205), Stephenson never truly managed to establish his Q technique within British psychology. Meeting strong resistance to theoretical change and having failed to claim a Chair at Oxford, Stephenson departed for the USA (and posts outside psychology) in 1947.

In recent years, however, similar concerns have again been expressed within psychology by a growing band of qualitative and critical researchers (see, for example, Harré and Gillett, 1994). Like Stephenson, proponents of a qualitative psychology tend to value curiosity and to promote discovery and understanding in preference to the logic of testing. They have also questioned the very existence of the unobservable entities that psychological tests invariably set out to measure and have duly rejected the subsequent employment of these entities as explanatory devices (see Potter and Wetherell, 1987). The ideal constructs of psychology are, indeed, less than ideal. Couple these conclusions with the centrality that Stephenson's inverted technique affords to persons (in their 'whole aspect'), as well as its stated intention to pursue 'empirical discoveries of a qualitative kind' (Stephenson, 1936a: 205), and one is once again reminded of the close relationship that connects Stephenson's method with the modern qualitative tradition in psychology (Harré, 1999; Watts and Stenner, 2003a,b).

It should be no surprise, therefore, to find that Q methodology has somewhat belatedly found a 'psychological' home for itself within that qualitative tradition. It has informed social constructionist studies (see, for example, Kitzinger, 1987; Stenner and Marshall, 1995; 1999), been adopted by feminist researchers (Kitzinger, 1986; Senn, 1996), and has even spawned its own theoretical or 'critical polytextualist' take

on social psychology (Curt, 1994; Stenner and Eccleston, 1994). Q methodology has now been successfully applied to many diverse psychological topics like, for example, child abuse (Stainton Rogers and Stainton Rogers, 1992), jealousy (Stenner and Stainton Rogers, 1998), I.B.S. (Stenner *et al*., 2000), attitudes to environmental issues (Capdevila and Stainton Rogers, 2000) and love (Stenner and Watts, 1998). In so doing, it has more than demonstrated its 'sense-making' capacity and ability to find qualitative 'order' even in domains where variability and disparity seem initially to have prevailed (see, for example, Watts and Stenner, forthcoming).

Of course, Stephenson had no such track record or qualitative tradition on which to rely as he developed Q methodology. Neither would the discipline allow him the luxury – now afforded to many qualitative psychologists – of turning to various critical philosophers as a means of securing theoretical support (Brown and Stenner, 2001; Stenner, 1993; 1996; 1998). It is perhaps understandable in such circumstances, to find that Stephenson instead turned to Freud for theoretical sustenance, especially given that psychoanalytic notions still permeate and influence a deal of qualitative work in psychology (for example, Billig, 1997; Parker, 1997a,b). Indeed, the basic Freudian premise of pleasure/unpleasure lies at the very heart of the Q methodological procedure (see Stephenson, 1983) – a procedure which, perhaps equally understandably, actually seems to resemble many quantitative methods more closely than it does its qualitative counterparts.

This procedure involves a heterogeneous set or sample of items (ordinarily a set of *statements* about a particular subject matter, although pictures, objects, and so on, might also be employed) being ranked or scaled (along a standardized ranking distribution or continuum) by a group of participants. They were to do this according to their own likes and dislikes and hence as a function of the *personal value* they assigned to each item. The items would thus be ranked according to (what Stephenson called) their 'psychological significance' for any given participant (see Burt and Stephenson, 1939). Those items of *great* psychological significance would evidently be 'ranked or scored highly, whilst those of little relative significance . . . [would be] ranked or scored lowly' (Stephenson, 1936a: 347). The general idea was to render a formerly heterogeneous set of items 'homogeneous with respect to . . . [a particular] individual' (p. 346), such that a single and gestalt *configuration* of items emerged.

As we have already implied, this sort of ranking procedure has some drawbacks for a method which now claims itself as a qualitative 'alternative', insofar as it can make Q methodology look too much like the tests, scales and questionnaires it purports to challenge. Yet the role of the participants in Q methodology, the nature of the gathered data, and the way in which those data are subsequently interpreted, could not be more different from contexts in which testing and measurement are prioritized. It is well documented, for example, that test procedures necessitate a certain *a priori* imposition of meaning. Such meaning must effectively be 'built into' the measurement instrument itself (see Maraun, 1998). Each separate item in a conventional attitude scale is assigned an exact predefined meaning and is henceforth considered to represent a partial measure of the otherwise imperceptible psychological construct that the scale sets out to measure. To complete such a test is to be *subjected to measurement*. The resultant data will be

treated as an independent series of absolute measurements *of* you, and will be interpreted only *in relation to* the operational definition (of the psychological construct) that the researcher has chosen to employ. The possibility that differences of kind, of meaning, interpretation or 'quality' may (be at issue or) have informed responses to the test is rarely considered. Indeed, qualitative researchers rightly chastise these methods for this lack of consideration (Potter and Wetherell, 1987).

Q methodology, on the other hand, neither tests its participants nor imposes meanings *a priori*. Instead, it asks its participants to decide what is 'meaningful' and hence what does (and what does not) have value and significance *from their perspective*. A series of absolute measurements cannot result from this process. Instead, a single set of essentially *relative evaluations* (and hence a gestalt configuration of items) is produced. These gestalt configurations have been made *by* the participants, on the basis of criteria which are *personal to them* (i.e., that which they consider to be 'psychologically significant'), and it is these gestalt configurations which constitute the *research target* of Q methodology.

It would, given the rationale presented above, be quite pointless (and a violation of the procedure) for the Q methodologist to strictly predefine the meaning of the items presented to a participant, for any given item can evidently take on its 'psychological significance' (and hence its meaning for a participant) *only* in the context of an overall configuration. Being the intended research target, it is these overall configurations (not test results or measures) which are then intercorrelated and factor analysed in a Q study (see Stephenson, 1936a). This produces a set of factors (onto which participants load on the basis of the configurations that they produce) which are exemplified and represented, not by different *subsets* of the presented items (as would be the case in conventional factor analysis), but by *all* the presented items configured in different but characteristic ways. The meaning and significance of these configurations must then be 'attributed *a posteriori* through interpretation rather than through *a priori* postulation' (Brown, 1980: 54). This is the basis of the Q 'method'. In the sections which follow, we shall discuss the mechanics of this procedure.

The generation of a Q set or 'item sampling'

The Q set is the collection of 'heterogeneous items' which the participants will sort. There are many possibilities in this context. Stephenson (1936a), for example, performed early studies looking at people's *predilection for vases* (using a Q set of vases) and the *hedonic value of certain odours* (using a Q set of bottled fragrances). He also studied issues of *personality* by asking participants to value a population of moods (e.g., cheerful, elated, affectionate, etc.) as descriptors of their own personality. Such single word stimuli work very well (see Watts, 2002). It is more usual, however, in a qualitative and psychological context, for a Q set to be constituted of *statements*, each making a different (but nonetheless recognizable) assertion about the appropriate subject matter.

In keeping with its rejection of hypothetico-deductive logic, it is not surprising to find that the formulation of specific hypotheses is typically inappropriate in Q methodology. As is true of many other qualitative methods (and in response to its factor analytic heritage), Q methodology is primar-

ily an *exploratory* technique. It cannot prove hypotheses. It can, however, bring a sense of coherence to research *questions* that have many, potentially complex and socially contested answers (Stainton Rogers, 1995).

The research question plays a very important part in any Q methodological study. This is because it dictates the nature and structure of the Q set to be generated. It will also act as a 'condition of instruction' for the participants, and will hence guide the actual sorting process. It is important, therefore, that the research question (or a close approximation to it) is clearly defined *before* the study proper commences. The research question must be straightforward and clearly stated (containing a single proposition only), and the Q set should enable participants to respond to the question in an effective fashion. If, for example, the research question is 'What is partnership love?', all the statements of the Q set *must* represent possible responses to that question. If participants are asked to describe their current love relationship, then all the statements must represent possible descriptions of love relationships, and so on.

Whatever the research question, the Q set must always be *broadly representative* of the opinion domain at issue. The generation of potential statements need not be theory-driven, however, as would be customary in the design of a test or questionnaire. Instead, it is carried out as a *sampling* task. As Curt (1994: 128–9) suggests, this is 'one place where Q-method is noticeably a craft'. In common with other crafts, the Q methodologist must carry out this task skilfully, patiently and with an appropriate application of rigour. As a result, the generation of the final Q set can often 'take up the bulk of the time and the effort involved (often several months of work in contrast to the few hours or at most weeks involved in administering the Q-sort)' (Curt, 1994: 120).

In practice, the final Q set can be elicited from any number of sources: by extensive reference to the academic literature (which generally identifies and breaks down a subject matter into its key 'themes'), from both literary and popular texts (magazines, television programmes, etc.) from formal interviews, informal discussions and often via pilot studies. A complete set of test or scale items can even be used to create a ready-made Q set (see, for example, Watts, 2002). In the end, the exact nature of the sampling task is of little consequence provided that the final Q set can justifiably claim to be 'broadly representative' of the relevant opinion domain, and this aim might clearly be satisfied in a number of different ways.

The exact size of the final Q set will, to a great extent be dictated by the subject matter itself. Generally speaking, however, a Q set of somewhere between 40 and 80 statements is considered satisfactory (*cf.* Curt, 1994; Stainton Rogers, 1995). Any less than this and issues of adequate coverage may be a problem. Any more and the sorting process can become unnecessarily unwieldy. It is always best, however, to initially generate an overly large number of statements, which can then be refined and reduced through processes of piloting.[2]

Even with effective piloting, however, there is a sense in which a Q set can never really be complete (as there is always 'something else' that might potentially be said). Yet this is actually of little import, for the procedural detail of Q methodology ensures that a Q set only needs to contain a representative *condensation of information*. This is because the main concern in a Q methodological context is not the Q set itself (which is, in any event, not consid-

ered to possess any specific meaning prior to the sorting process), but the relative likes and dislikes, meanings, interpretations and overall understandings which inform the participants' *engagement* with the Q set. In practice, the qualitative detail of a Q methodological study actually gets *filled out* as the study proceeds (rather than being 'cancelled out' as can often happen in the context of a test), with the subjective viewpoints of the participant group being central to this process.

An example of this principle may be useful. Watts and Stenner (forthcoming) used a Q set statement which suggested that '*love is an intense commitment to a partner*'. The accuracy of this 'hypothesis' is confirmed by the academic literature (see, for example, Fehr, 1988; Sternberg, 1986). Yet such confirmation says little about the *meaning* of commitment, nor does it tell us how (or indeed 'where') the notion of commitment *fits into* people's wider expectations and understandings of love.

In the study which followed, participants generally considered this 'commitment' statement to be a very good descriptor of partnership love (such that it was frequently 'ranked highly'). It was possible to observe, however, that it was ranked highly for a variety of *quite different reasons*: (a) because some participants viewed partnership love as a long-term, practical style of commitment, such that 'intensity' implied a thoroughgoing and enduring provision of support; (b) because partnership love was viewed as a short-term sexual and emotional commitment, such that 'intensity' came to imply a total immersion in passion; and even (c) because the commitment of love was feared, such that 'intensity' was experienced in terms of the placing of one's self, emotions and trust in the hands of another person, and so on.

It becomes very obvious, therefore, not only that the statement '*love is an intense commitment*' contains a condensation of *potential* information, but also that this condensation was unfolded, unpacked and otherwise expressed in very different ways as the participant group *engaged* with the presented items (and notice that we have completely ignored the participants who gave this statement a more lowly ranking). This unfolding of meaning can be observed across *every* statement of the Q set. One can appreciate, therefore, how qualitative detail proliferates and 'fills out' as a Q study proceeds.

In taking advantage of these processes of participant engagement, Q methodology exploits what Harvey (1997: 146–47) has called 'one of psychology's most basic and well established principles', namely, our desire to structure and to ascribe meaning to all 'impinging stimuli and events'. Indeed, it is these very desires which ensure the robustness of Q methodology, as a group of participants will ultimately make vigorous attempts to impose their viewpoints onto *any set of statements they are given.* In other words, even a 'less than ideal . . . [Q set], because it invites active configuration by participants ('effort after meaning'), may still produce useful results' (Stainton Rogers, 1995: 183). If a Q set is at least 'broadly representative' of its subject matter, therefore, the engagement of the participant group with that Q set (and the resultant configurations) *will* afford a general overview of relevant viewpoints 'on the subject' (which is all that is required for the purposes of Q methodology). We will provide a practical demonstration of this process later in the paper using the example Q set (about 'young offenders and their punishments') outlined in Table 1.

The ranking procedure or 'Q sorting'

The Q methodological ranking procedure – Q sorting – is what Brown (1980: 17) calls 'the technical means whereby data are obtained for factoring'. In practice, it is simply a convenient means of facilitating the (evaluations and) rankings of the participants. For reasons of simplicity and pragmatism, participants are *not* ordinarily required to carry out a complete (1 to *n*) rank ordering of the Q set items. Instead, they simply assign each item a ranking position in a fixed quasi-normal distribution (illustrated in Figure 1 below) and 'along a simple, face-valid dimension, for example most agree to most disagree, most characteristic to most uncharacteristic, most attractive to most unattractive' (Stainton Rogers, 1995: 180). An 11 or a 13 point scale is generally employed, possible ranking values ranging from +6 or +5 for items that are, say, 'most agreeable' in the view of a particular participant, through 'zero', to −5 or −6 for items that are considered 'most disagreeable'.

Figure 1 demonstrates that the distribution also dictates the *number* of items that can be assigned to each ranking position (in the example, two items at the +6 position, three at +5, and so on). For this reason, it is also known as a 'forced' distribution. This apparent forcing (or restriction) of participants may alarm some qualitative researchers. Yet such concerns are largely misplaced. It is, in fact, quite possible to employ different forms of distribution in the context of a Q methodological study, including completely 'free' distributions, which (as the name suggests) allow participants to assign *any* number of items to *any* of the available ranking positions.[3]

Yet Brown (1980: 288–89) has presented an array of statistical comparisons which demonstrate that 'distribution effects are virtually nil'. This means that the chosen distribution actually makes *no noticeable contribution* to the factors which emerge from a particular study (and this is the main reason why a complete rank ordering of the items is also unnecessary). Contradictory as it may seem, therefore, a forced distribution is actually no more restrictive than a 'free' distribution. If Q methodologists generally prefer the forced distribution, therefore, it is because it delimits unnecessary work and because it is convenient for their participants.

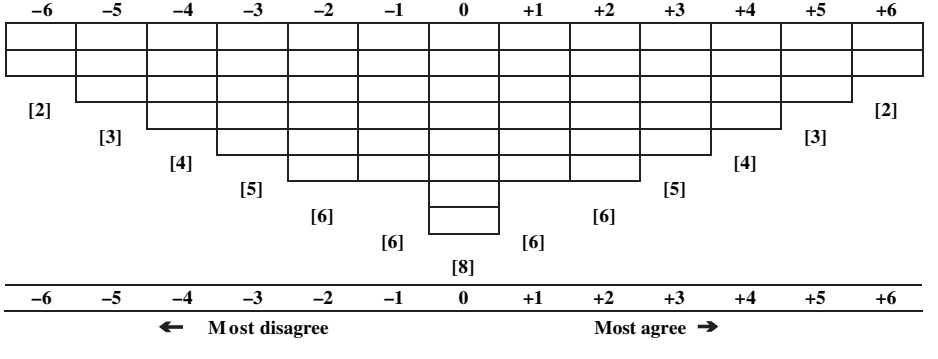

Figure 1 Example Fixed Quasi-Normal Distribution. Ranking values range from +6, through 'zero', to −6. Numbers in [brackets] indicate the number of items that can be assigned to any particular rank. A total of 60 items can be accommodated in the distribution illustrated.

We appreciate, however, that such arguments will not persuade those who consider *any* type of ranking procedure or distribution to be overly restrictive. Concerns are also frequently expressed about further restrictions imposed by the Q methodological procedure, particularly the provision of a *pre-designed* Q set. There is, after all, always a finite number of items in the Q set and it is the researcher who has provided those items (albeit on the basis of careful sampling processes). These and associated arguments often seem to culminate in the belief that 'you can only ever get back what you put in' in a Q methodological context.

Once again, however, we would argue that such objections are misplaced, for they necessarily overlook the basic aims and premises of the method. First, they (at least partially) rely on the premise that a statement like '*love is an intense commitment*' must have a single, predefined meaning and hence that it could only ever be interpreted in a single way. Yet we have already shown that this (essentially R methodological) assumption is not relevant to Q methodology. The singular 'putting in' of this statement does not prevent our 'getting back' multiple and qualitatively diverse responses. The logic of this objection is faulty. The same interview question need not automatically elicit the same response and the same is true of items in a Q set.

Secondly, these objections fail to notice that it is overall item *configurations* which represent the research target of Q methodology. With this in mind, Brown (1980) presents a series of analyses which demonstrate (because of the factorial nature of the sorting procedure) that a study which employed a Q set of only 33 statements and a quite limited (+4 to −4) ranking distribution would make available to its participants 'roughly 11 000 times as many [sorting] options as there are people in the world'. In other words, Q methodology pursues overall item configurations and its procedure renders a *hyperastronomical* number of such configurations available to its participants. It is difficult to see how this might sensibly be construed as restrictive. As Brown (1980: 267) rightly concludes, Q methodology leaves more than 'sufficient room for individuality [to be expressed]'.

In expressing their individuality via the Q sorting procedure, a participant will ultimately be required to allocate *all* the Q set items an appropriate ranking position in the distribution provided. A clear and 'gestalt' configuration of items will duly emerge. If the participant is happy with this configuration, the various item numbers (and hence the 'form' of the overall configuration) should be recorded (as they are in Figure 2).

The next task for the researcher involves the gathering of supporting information from the participant. Open-ended comments are duly requested. This can be done via a brief post-sorting interview (which can then be transcribed and subjected to analysis), or simply via some form of 'response booklet' or post-sorting questionnaire. Such *post hoc* analyses ordinarily investigate: (a) how the participant has interpreted the items given especially high or low rankings in their Q sort, and what implications those items have in the context of their overall viewpoint; (b) if there are any additional items they might have included in their own Q set (what they are, why they are important, and so on); and (c) if there are any further items about which the participant would like to pass comment, which they have not understood, or which they simply found confusing. Such open-ended comments are a vital part of the Q methodological procedure, for they will aid

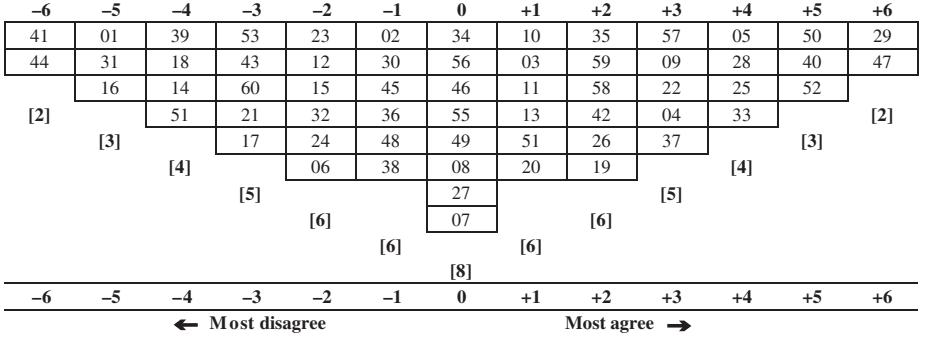

Figure 2 Example Completed 'Q Sort' Configuration. Items 29 and 47 have been allocated rankings of +6 which suggests that they represent the 'most agreeable' items in the view of this participant. Moving from right to left in the configuration, the items are becoming less and less agreeable until we arrive at items 41 and 44 which (having been ranked at −6) are evidently considered to be the 'most disagreeable' items in the Q set.

the later interpretation of the sorting configurations (and viewpoints) captured by each of the emergent factors.

The participant group

Large numbers of participants are not required for a Q methodological study. Q methodology aims to reveal (and to explicate) some of the main viewpoints that are favoured by a particular group of participants. It probably does this most effectively when the participant group contains between 40 and 60 individuals (Stainton Rogers, 1995). This is only a 'rule-of-thumb', however, for highly effective Q studies can be carried out with far fewer participants. A salient viewpoint might, of course, be revealed by reference to a single participant. If, on the other hand, we want to demonstrate that this viewpoint is shared by several persons in 'the group', and hence to make sense of our *subject matter* (rather than a specific individual) on the basis of such a consistency, we must evidently move beyond the single case.

Yet the employment of *very* large numbers of participants in a Q methodological context can itself be problematic. Indeed, such an approach can easily negate many of the subtle nuances, complexities, and hence many of the essential *qualities* contained in the data. This is obviously counterproductive in the context of a qualitative technique. Those seeking to publish in mainstream psychological journals will also find that statistical arguments stand in the way of studies which try to use 'too many' participants.[4] In keeping to smaller numbers, therefore, an emphasis on quality is maintained, pattern and consistency can still be detected within the data, and the prospects of publication are also increased.

As is so often true, however, size is not all important. The exact constitution of the participant group must also be considered. In some contexts it may be sensible to *strategically sample* participants, especially if they seem likely to express a particularly interesting or pivotal viewpoint. One can, for example, carry out Q methodological studies which investigate particular *events* – a specific football match perhaps – in

which case one ought to ensure that Q sorts are gathered from as many of the obviously pertinent demographic groups as possible. Both sets of fans, for example, both sets of players, the ball boys perhaps, neutral spectators, the referee and his assistants, members of the press, and so on.

Where Q studies aim to investigate particular *concepts*, however, participants may not divide so obviously along lines prescribed by demographic characteristics. In such cases, it is better to avoid too many assumptions *a priori*, particularly where these assumptions are based on preconceived demographic notions. The whole point of Q methodology is to allow individuals to categorize *themselves* on the basis of the item configurations they produce (and hence via the viewpoints they express). Its function is exploratory. In such circumstances, it is probably sensible to rely on more opportunistic sampling techniques, at least in the first instance, or until a series of Q methodological explorations (and their emergent factors) provide empirical justification for the belief that certain viewpoints 'belong' exclusively to specific demographic groups.

Statistical analysis (factor extraction, rotation and estimation)

Q methodology employs a *by-person* correlation and factor analytic procedure. Hence, it is the *overall configurations* produced by the participants that are intercorrelated and factor analysed. The initial correlation matrix duly reflects the relationship of each (Q sort) configuration with every other (Q sort) configuration (*not* the relationship of each item with every other item). To subject this matrix to factor analysis is to produce a set of factors onto which the participants load on the basis of the item configurations they have created. Hence, two participants that load onto the same factor will have created very *similar* item configurations. Each factor duly captures a *different* item configuration which is nonetheless shared by (and which is characteristic of) the participants who load onto that factor.

There are now several dedicated Q methodological packages available which allow the appropriate analyses to be conducted. PCQ for Windows is probably the best commercial product available (Stricklin and Almeida, 2001). PQ Method (Schmolck, 2002), on the other hand, does the job effectively and is available as a free download from the internet.[5] Such dedicated packages are to be recommended. They facilitate data input, automatically generate the initial by-person correlation matrix, and make processes of factor extraction, rotation and estimation very straightforward.

Despite this automation, however, the researcher still has to make some important decisions as the analysis proceeds. Different types of factor analysis (and hence factor extraction) exist, as do different methods of factor rotation, and there often seems little reason for preferring one system over another. To further complicate matters, traditional factor techniques do not automatically resolve themselves into a single, correct (by which we imply 'mathematically superior') solution, as would the more recently developed principal components (PCA) or cluster analytic techniques. This is particularly pertinent in a Q methodological context, as the very oldest of the factor techniques (known as the centroid or simple summation method) is generally preferred.

For this reason, dedicated Q methodological packages tend to offer the centroid method as their default choice (and, indeed,

PCQ for Windows offers no alternative). This method offers a potentially *infinite* number of rotated solutions. Indeed, it is exactly this openness and indeterminacy which appeals to the Q methodologist, as it leaves them free to consider any data set from a variety of perspectives, before selecting the rotated solution which they consider to be the most appropriate and *theoretically* informative. Having said that, both authors have also analysed Q data using PCA and have found the results to be equally satisfying.

The accent on theoretical discretion discussed above has nonetheless led many Q methodologists to be critical of modern factor rotation techniques (such as varimax) which are perceived to reveal only the most *mathematically* (not necessarily the most theoretically) informative solution. Such practitioners retain Stephenson's preference for theoretical (judgemental or 'by-hand') rotation techniques (see, for example, Brown and Robyn, 2003). Why, argue the enthusiasts for hand-rotation, let a computer decide which point of view to adopt on one's data when an infinite number are possible? Whilst this argument – which has been known to cause vertigo in those who hope to find certainty in statistics – is undoubtedly valid, in practice many Q methodologists (like many more conventional factor analysts) prefer the simplicity and reliability of the varimax procedure.

The varimax procedure is also consonant with one of our typical aims in using Q methodology: namely, to reveal the range of viewpoints that are favoured by our participant *group.* Given this aim, it makes theoretical sense for us to pursue a rotated solution which maximizes the *amount of variance* explained by the extracted factors and as the varimax procedure automatically seeks this mathematically superior solution, it also makes sense for us to prefer this technique. As an additional advantage from a qualitative perspective, the varimax procedure is also highly reliant 'on the topographical features of the correlation matrix' (Brown, 1980: 238), which affords a welcome priority to the input of the participant group on the emergent factor structure. This prioritization of participant input is of course a shared feature of many qualitative research techniques in psychology (Willig, 2001). In the end, however, the technique of rotation employed is always going to be dependent 'on the *nature of the data* [gathered] and upon the *aims of the investigator*' (Brown, 1980: 238) and there is ultimately no substitute for careful consideration in the context of a particular study.

The next step is then to decide which factors should be selected for interpretation. A standard requirement is to select only those factors with an eigenvalue in excess of 1.00.[6] This is undoubtedly a somewhat arbitrary criterion in the context of Q methodology. It is well known, for example, that several factors with eigenvalues in excess of 1.00 might be extracted even from random data, as random patterns will always arise and be detected. On the other hand, it is a generally accepted means of safeguarding factor reliabilities and factors which go below this minimum will ultimately serve no data-reductive purpose as they explain less of the overall study variance than would any single Q sort.[7] A second standard requirement is that an interpretable Q methodological factor must ordinarily have at least two Q sorts that load significantly upon it alone.[8] Such significantly loading Q sorts are called 'factor exemplars' as they exemplify the shared item pattern or configuration that is characteristic of that factor. There might, in certain circumstances, be a theoretical justification for

accepting and interpreting factors with only one exemplar, but the interest in *shared* orientations nonetheless leads most Q researchers to adopt the more stringent criterion outlined above.

When, as is obviously typical, a factor is loaded by more than one exemplar, a factor estimate is generated through a procedure of weighted averaging (this occurs automatically in programmes such as PCQ and PQ Method). In effect, the Q sorts of all participants that load significantly on a given factor are merged together to yield a single (factor exemplifying) Q sort which serves as an interpretable 'best-estimate' of the pattern or item configuration which characterizes that factor. Confounded Q sorts (which load significantly on two or more factors) are excluded from this weighted averaging procedure.[9] The endpoint of the statistical analysis is duly reached when each of the selected factors is represented by its own best-estimate Q sort or 'factor array' (naturally, factors with a single exemplar will be forced to use that individual's *own* Q sort as a best-estimate). These factor arrays or best-estimate Q sorts can then be subjected to interpretation.

Factor interpretation

The study files provided by dedicated Q methodological packages can be confusing for an uninitiated reader. They ordinarily provide a table which illustrates the loadings of all participant Q sorts on all extracted and rotated factors. Significant loadings will be marked with a star (*).[10] The eigenvalues and the percentage of the study variance explained by each factor will also be provided, as will a table (normally entitled 'Factor Correlations') which indicates the intercorrelations of the various factor arrays (giving a basic indication of the relationships between the factors). A wealth of other information may also be included. The most important aspect of the study file will nonetheless be the factor arrays themselves. These will be found in a table, 'Item or Factor Scores'. Table 1 is presented in a typical format.

The data summarized in Table 1 are taken from a Q study about young offenders and their punishment. It was carried out by an undergraduate student at University College Northampton and is reproduced here with blemishes intact. It nonetheless provides a clear example of Q methodology at work and is illustrative of the method's efficacy and effectiveness even in the hands of an inexperienced user. We shall demonstrate this effectiveness by producing a brief interpretation of the study's first factor (A).

The interpretative task in Q methodology involves the production of a series of summarizing accounts, each of which explicates the viewpoint being expressed by a particular factor. These accounts are constructed by careful reference to the positioning and overall configuration of the items in the relevant 'best-estimate' factor arrays. The obvious place to begin such deliberations is at the two 'poles' of the Factor A configuration. Here one finds the items Factor A (or, to be exact, those participants whose Q sorts exemplify factor A) feels particularly strongly about. There is a strong belief, for example, that 'young offenders have to be punished' (as item 10 is ranked highly at the +4 position) and that they will not 'learn to behave properly' unless they are (item 22 ranked at +3). Yet such punishment must ultimately be tailored to reflect both the seriousness of the crime (item 18 is ranked at +4) and the age of the offender (item 06: +3). In keeping with this stand on age, Factor A also rejects the idea that young

Table 1 Example item scores table for a study about the punishment of young offenders

Statements factors	A	B	C	D	E
01 Current punishments are not severe enough to motivate change	+2	+3	0	−2	+2
02 Young offenders attend counselling for rewards not for rehabilitation	−1	+1	−2	−1	−1
03 The death penalty is fair for young offenders who commit murder	−4	0	−2	−4	−4
04 It is legitimate for under 18's to be tried as adults	−3	−2	−2	0	0
05 Young offenders should be punished in other ways not prison	+2	0	+2	+2	0
06 Age is an important factor in the choice of a punishment type	+3	−1	0	−1	0
07 It is legitimate to punish parents for young offenders crimes	0	−4	−4	−3	−3
08 10 years old is the right age for criminal responsibility to begin	0	−1	−1	+3	−2
09 The age of criminal responsibility should be lower than 10 years	−2	−3	−3	−2	0
10 Young offenders have to be punished	+4	+4	+2	+4	+4
11 Young offenders should be treated less severely than adults	0	−2	0	0	−2
12 Young offenders should be imprisoned with adults	−4	−1	−4	−2	−4
13 Young offenders should receive attention and support	+2	−3	+3	−2	0
14 Community service is a good way of punishing young offenders	+1	0	+1	−1	+1
15 Young offenders who commit murder should get life imprisonment	−1	+4	+4	+1	−2
16 Electronic tagging is an effective way of punishing young offenders	+1	+3	+1	+1	+1
17 Young offenders should repay their victim with manual work	−3	0	+1	+1	0
18 Punishments should reflect the seriousness of the crime	+4	+1	+2	+4	+2
19 Convicted young offenders should be made to pay fines	−1	−4	−1	−3	−3
20 Cautioning has no impact on the behaviour of young offenders	+1	+1	0	0	+3
21 Persistent young offenders should be taught a trade	0	−2	+3	+2	+1
22 If young offenders are not punished they won't learn to behave	+3	+2	+4	+3	+3
23 Punishments are not severe enough to deter young offenders	0	2	0	0	+4
24 Parents should be punished for minor crimes that their child commits	−2	0	−3	−4	−1
25 Young offenders should not have their IDs protected when they commit serious crimes	−2	+2	−1	+2	+2

This table indicates the rankings assigned to each item within each of the factor exemplifying Q sort (or item) configurations. Reading the table *by column* reveals the configuration (or comparative ranking) of items which characterizes a particular factor. In column A, for example, we can see that Factor A ranked item 01 at +2, item 02 at −1, and so on. Reading the table *by row* reveals the comparative ranking of a particular item across all the factors. In row 02, for example, we can see that item 01 was ranked at +2 by Factor A, at +3 by Factor B, and so on.

offenders should be tried *as* adults or imprisoned *with* adults (items 04 and 12 are ranked at −3 and −4 respectively). Young offenders should not have to carry out manual work to repay their victims, nor should the death penalty be considered for underage murderers (item 17: −3, item 03: −4).

One can observe that a relatively clear and unitary viewpoint is *already* beginning to emerge from these item rankings. It is a mistake to assume, however, that all 'the action' is taking place at the poles of the distribution. Consider item 07, for example, which suggests that the parents of young offenders should be punished for their child's crimes. This item is ranked at 'zero' by Factor A. Such a display of indifference and seeming neutrality surely indicates that this item is of little importance to the viewpoint being expressed? Yet when consideration is given to the ranking of this item by the *other factors* in the study (−4, −4, −3 and −3 respectively), the zero awarded by Factor A suddenly becomes 'information rich'. Factor A does not think parents should be punished where minor crimes have been committed (we know this because item 24 is ranked at −2), but the ranking of item 07 suggests that 'parental'

punishment has not been completely dismissed by this factor.

This example demonstrates that much of significance can occur in the supposedly 'neutral' area of the configuration. Indeed, any interpretation which disregards the item rankings in this area will almost certainly fail to capture the subtleties of the viewpoint being expressed. A concentration on too few of the items in the array will also prevent the holistic or 'gestalt' nature of the viewpoint being communicated. This represents a serious interpretative failure given the stated research target of Q methodology. One can safeguard against these problems, however, by checking the qualitative comments gathered from participants who have loaded significantly on the factor being interpreted. These comments can help to verify initial interpretations of specific item rankings. Participant 17, for example, tells us that without suitable punishments young offenders '*will not learn what is and isn't acceptable behaviour*'. This reinforces the ranking of item 22 at +3. Perhaps more importantly, however, participant 26 confirms the salience of this factor's relatively positive ranking of item 07 when he suggests that '*parents are accountable . . . (for they have been) . . . in the best position to exert influence*' over their child.

If this process is carried to its conclusion, such that various item rankings and participant comments are effectively combined, a clear interpretation of the factor will emerge. An example interpretation of Factor A is shown. Notice that this interpretative account has been given a title and that some pertinent demographic information has been provided. The rankings which informed the interpretation at each stage are also included in the text (as a point of reference for the reader). The citing of

Factor A: A crime-related and age sensitive approach to the punishment of young offenders

Demographic information: Factor A has six significantly loading participants and it explains 17% of the study variance. It has an eigenvalue of 5.40. Four of the loading participants are men, two are women. They have an average age of 23.84 years.

Factor interpretation: Young offenders must definitely be punished for their crimes (10: +4). Cautioning has little impact and punishments handed out are generally not severe enough (20: +1; 01: +2). As participant 17 confirms, without suitable punishments young offenders '*will not learn what is and [what] isn't acceptable behaviour*' (22: +3). It is nonetheless important that the chosen punishment reflects both the age of the offender and the type of crime they have committed (06: +3; 18: +4). Under 18's should certainly not be tried as adults, nor should they be imprisoned with adults (04: −3; 12: −4). They should have their identities protected in the context of serious crimes and the death penalty should not be considered – even in cases of murder (25: −2; 03: −4). After all, young offenders may well have committed their crime without having a full '*understanding of the law*' (participant 26 comment). Life imprisonment is questionable for the same reason (15: −1). Given this situation, it would be quite wrong to further lower the age of criminal responsibility (09: −2). Prison sentences are probably not the right way to punish young offenders (05: +2), nor should they be made to do manual work as a means of repaying their victims (17: −3). We need to take a more constructive approach to punishment, offering attention and support where appropriate (13: +2). Electronic tagging devices and community service open up a range of possibilities in this regard (16: +1; 14: +1). Neither should we completely dismiss the idea of punishing the *parents* of young offenders if a serious crime has been committed (07: 0), although this would probably be inappropriate in the context of more minor crimes (24: −2). Indeed, there is a real sense in which the '*parents are accountable. . . [for they have been]. . .in the best position to exert [an] influence*' over their errant child (participant 26 comment).

Factor A: example factor interpretation
This interpretation uses the relative item rankings and participant comments to arrive at a clear, 'gestalt' account of the viewpoint being expressed by Factor A.

'14: +1', for example, indicates that item 14 was ranked in the +1 position by Factor A and that this item is relevant to the current section of the interpretation:

In principle, of course, the process of interpretation is potentially never-ending, there always being different shades of meaning and emphases that could be drawn from the data, and hence different readings applied to it. Yet Q methodological interpretation is necessarily constrained by (and can always be checked against) the subjective input of the participant group, as, in contrast to many other qualitative methods, that input is actually reflected and frozen in the relevant 'factor exemplifying' item configurations. In Q methodology, subjective input produces *objective structures.* No doubt the procedure which makes this possible also necessitates a certain loss of ecological validity in relation to other qualitative methods. In partial compensation, however, the presence of these objective structures allows Q methodology to *demand* a great deal of the interpreting researcher, for any failure to respond fully and faithfully to the subjective input of the participant group can be easily detected.

The accuracy and efficacy of the produced interpretations can also be verified via a second source: simply by asking relevant (significantly loading) participants to comment upon them. This can be a useful source of further insight, although one should bear in mind that all such interpretations are designed to communicate a 'shared' viewpoint, and hence that they need not provide a veridical representation of a participant's *own* opinion. Whilst, therefore, alternative readings of our example factor are a definite possibility, these various checks and balances – thorough attention to the item rankings, to participant comment and to the verifying remarks of loading participants – make us very confident that the main thrust of all competent readings would be very similar to our own. Put another way, in common with most hermeneutic endeavours, Q data make it relatively simple to reject incompetent readings, whilst allowing scope for numerous subtly different competent readings to coexist.

We are thus confident that this interpretative process has produced a coherent expression of the Factor A viewpoint – captured in the item rankings of our participants – about the ways in which young offenders should be punished. This viewpoint is 'shared' by members of the participant group. In contrast to many other qualitative methods, it has also been captured in a genuinely gestalt fashion. Add this viewpoint to the others revealed in the same study (see the item rankings for Factors B, C, D and E in Table 1) and Q methodology has allowed us to gain a comprehensive snapshot of the major viewpoints being expressed by our participant group.

Conclusion

We close with a caveat. Q methodology makes no claim to have identified viewpoints that are *consistent within individuals* across time. Understanding this is crucial to understanding the role and place of this methodology. Factor A need not, and we think should not, be thought of as (nor reduced to) the expression of stable intrapsychic features of the factor exemplars (such as 'attitudes' or 'personality traits'). To think otherwise is to impose *a priori* the essentialist and somewhat counterintuitive assumption that a given participant is

capable of expressing only one coherent viewpoint on an issue. All that we can say is that these participants *did* express these viewpoints via these item configurations. Whilst this leaves individual factor exemplars free to 'change their minds', we might nonetheless expect the emergent *manifold of shared viewpoints* to show a degree of consistency over time. In short, what is decisive from a Q methodological perspective is less '*who* said what about X?' than 'what is currently being *said* about X?'

It is in this respect that Q methodology provides results that are consistent with other broadly constructionist endeavours, such as discourse analysis, social representations theory, autopoietic systems theory and ethogenics. In enabling a cartography of social semantics, Q methodology is in a relation of complementarity with more thematic or 'individual' approaches in qualitative psychology, and also with methods, like conversation analysis, that focus on the micro processes of situated interaction. Using Q methodology, to conclude, does not entail a denial of microlevel communicative pragmatics, nor of macrolevel social structures. Rather it entails a focus on subjectively expressed, socially organized semantic patterns, whether these are subsequently grasped as 'discourses', 'representations' or as 'fields of the sayable and the seeable' (Curt, 1994).

Acknowledgements

The authors would like to thank Michelle Botley for her hard work and for allowing them to take advantage of the illustrative data used in this paper. We are also grateful to Rom Harré and the journal editors for their helpful comments.

Notes

1. Yet, in contrast to the methodological changes mentioned, this theoretical departure is not *guaranteed* by the act of inversion. One can, for example, choose to conduct a Q technique factor analysis on a collection of R methodological 'test' results. In other words, the analysis is simply carried out on 'the transpose of the R matrix – i.e., as the correlation and factorization by rows of the *same matrix of data* that in R is factored by columns' (Brown, 1980: 12–13, emphasis added). This method evidently continues to prioritize psychological traits (hypothesized cognitive 'parts' of persons) and their measurement. No theoretical change occurs here, merely a *statistical* transposition of a set of data. This approach to Q technique factor analysis was employed and promoted by Cyril Burt. Indeed, it is Burt's 'transposed matrix model' which has invariably found its way into mainstream textbooks under the heading 'Q technique' factor analysis (see, for example, Maxwell, 1977). The main problem with this model (if one ignores possible theoretical objections) is that a single matrix of data can properly be transposed (and thus factor analysed by both row and column) only where a single measuring unit is employed throughout the matrix, 'i.e., when the same measuring unit (e.g. the inch) is common to both rows and columns' (Brown, 1980: 13). This condition is rarely satisfied, however, simply because the measurement of differing traits (be they psychological or physical) is hardly ever undertaken using the same measuring unit. As a result, the intercorrelation (and hence the direct comparison) of such traits *is* highly problematic. Stephenson recognized these problems very quickly (for a summary, see Burt and Stephenson, 1939). Indeed, he criticised Burt's transposition procedure quite vociferously, for both methodological and theoretical reasons (see, for example, Stephenson, 1936a). If the Q technique was to invert the theoretical tradition, he realized, a distinct change of *method* and a distinct change in the nature of the gathered *data* would be required.

2. Piloting can be done in a number of different ways. As we implied in the main text, however, asking pilot participants to physically *sort* a very large number of items is probably unwise. It may be more sensible to ask a group of pilot participants (or even one or two participants with subject-specific expertise) to provide general comments on the construction of the Q set, remembering that the primary functions of the piloting process are to ensure: (a) that semantic duplication within the statements is avoided; (b) that statements are clearly expressed (usually in everyday rather than technical terminology); (c) that statements express a single proposition only (as multiple propositions can make sorting *and* interpretation highly problematic); (d) that the Q set is properly 'balanced' (in the context of politically charged or particularly contentious research questions, it may be very important to ensure that relatively equal numbers of 'pro' and 'anti' responses are available to your participants); and finally, and perhaps most importantly, (e) that the Q set provides adequate coverage of relevant issues and that it is, therefore, broadly representative of the appropriate opinion domain.

3. The only exception to this rule is that at least *one* item must be assigned to *every* ranking position. If this is not done, the linear assumptions of correlation would effectively be broken, rendering the intercorrelation of the resultant Q sorts highly problematic (Kline, 1994).

4. The rules of conventional (R methodological) factor analysis demand that your study be designed such that it has at least twice as many participants as it does variables (Kline, 1994). Once the matrix is inverted, however, as it is in a Q methodological context, this same assumption suggests that you should have at least twice as many items in the Q set as you do participants. In truth, however, the two situations are hardly equivalent and objections on these grounds are unusual. This may be because the initial assumption is itself somewhat arbitrary (even in an R methodological context). Nonetheless, if mainstream journal publications are a goal, it is probably sensible to maintain at least a 1:1 ratio of items to participants (and certainly to avoid studies which include many more participants than items).

5. The relevant internet address is: www.rz.unibw-muenchen.de/ ~p41bsmk/qmethod

6. A factor's eigenvalue (the word just means 'characteristic value') is simply the sum of squared factor loadings for that factor. Divide the eigenvalue by n (where n equals the number of participants in your study) and multiply the result by 100, and you will have calculated the variance accounted for by that factor. This gives us the following equation (Eq. 1): $V = 100(\mathrm{EV}/n)$. It follows, therefore, that one can also calculate a factor's eigenvalue by multiplying the variance it accounts for by n and then dividing the result by 100. The following equation (Eq. 2) results: $\mathrm{EV} = V \times n/100$. More detailed (and sometimes contradictory) observations about eigenvalues, their calculation, and their use in a Q methodological context can be found in Brown (1980). This publication is available at the following web address by kind permission of Professor S.R. Brown: http://reserves.library.kent.edu/coursepage. asp?cid = 203&page = 01.

7. Let us assume, for example, that our hypothetical factor has an eigenvalue of 0.86 and that our study includes 50 participants. Insert these figures into Eq. 1 of Note 6 and we see that $V = 100(0.86/50)$. In other words, our factor accounts for 1.72% of the study variance. A single Q sort in the same study will, however, account for 2% of the study variance (as 100% of the study variance divided amongst 50 participant Q sorts = 2%). In other words, the single Q sort has more explanatory power than the factor.

8. Arguably, decisions about the 'significance' of a loading should properly be made on theoretical rather than statistical grounds. On the other hand, one will undoubtedly have to justify such decisions and statistical arguments are likely to hold considerably more weight in a psychological context. A significant factor loading (at $P < 0.05$) can be calculated using the following equation: $1.96(1/\sqrt{n})$ where n equals the number of items in your Q set. If, therefore, our Q set contains 60 items, a statistically significant loading will be equal to or greater

than $1.96(1/\sqrt{60}) = \pm 0.26$. Of course, this will still mean that 1/5 of your significant factor loadings are likely to have occurred by chance. It may be better, therefore, to employ the more stringent ($P < 0.01$) significance level. This can be calculated using the equation: $2.58(1/\sqrt{n})$. Using the same example, this would mean that a statistically significant loading would have to be equal to or greater than $2.58(1/\sqrt{60}) = \pm 0.34$. Again, Brown (1980) provides a more detailed commentary.

9. As we shall see in the section on 'Factor Interpretation', the confounding of a participant Q sort also tends to mean that their qualitative comments are lost from the final analysis (as one generally only uses the comments of significantly loading participants to support a particular factor interpretation). This is a shame. We tend, therefore, to employ a strategy which minimises confounding and which duly maximises the number of significantly loading participants in a particular study. Let us suppose you have followed the instructions laid out in Note 8 above and have calculated that a significant loading in your study (at $P < 0.01$) must reach ± 0.34 or above. You then find that (in using this significance level) 35 of 50 participant Q sorts load on a single factor, 4 have no significant loading and the remaining 11 are confounded. This means that 15 participant Q sorts (30% of the data) are not being used to construct the various factor exemplifying item configurations. In such circumstances, it may be sensible to consider *raising* the level at which a loading is said to be significant (which will only make your criterion *more* stringent from a statistical viewpoint). In the example, we might look again at the solution with a significance level of ± 0.35 or above, then ± 0.36 or above, and so on. At ± 0.40 we find that 42 of the 50 participants load significantly on a single factor, five now have no significant loading and only three are confounded. In other words, only eight participant Q sorts (16% of the data) are not now being used. We also find that raising the level still further (to ± 0.41 or above) only serves to increase the number of Q sorts with no significant loading and hence to *reduce* the overall numbers of participants with significant (and usable) loadings. At ± 0.40, therefore, we have *maximized* the number of participants with significant loadings and have duly achieved our most satisfactory solution.

10. It is perfectly possible for a single factor to have both positive and negative significant loadings. This is known as a 'bipolar' factor. This bipolarity implies that two diametrically opposed viewpoints are being expressed by the participants who load on such a factor, each viewpoint having a factor exemplifying item configuration that it is the 'mirror-image' of the other. In other words, the positive loaders will *agree* with the item rankings and overall item configuration for that factor (as they are contained in the Item or Factor Score table in your printout – see Table 1 in the main text), whereas the negative loaders are agreeing with an entirely *reversed* configuration (and hence they are advocating an *opposed* viewpoint). If the positive loaders have ranked an item at $+4$, therefore, the negative loaders have effectively ranked the same item at -4. If the positive loaders have ranked an item at $+2$, the negative loaders have ranked it at -2, and so on. Such factors necessarily require you to produce *two* distinct factor interpretations, the first interpreting the factor exemplifying configuration as it appears in your Item Score Table (which will reveal the positive viewpoint), the second interpreting the same configuration with all the item rankings reversed (which will reveal the negative or diametrically opposed viewpoint).

References

Billig, M. 1997: The dialogic unconscious: psychoanalysis, discursive psychology and the nature of repression. *British Journal of Social Psychology* **36**, 139–59.

Brown, S. and Stenner, P. 2001: Being affected: Spinoza and the psychology of emotion. *International Journal of Group Tensions* **30**, 81–105.

Brown, S.R. 1980: *Political subjectivity: applications of Q methodology in political science*. New Haven, CT: Yale University Press.

Brown, S.R. and Robyn, R. 2003: *Reserving a key place for reality: philosophical foundations of theoretical rotation.* Paper presented at the 19th annual conference of the International Society for the Scientific Study of Subjectivity, Kent State University, OH, 2–4 Oct.

Burt, C. 1937: Correlations between persons. *British Journal of Psychology* **28**, 59–96.

Burt, C. and Stephenson, W. 1939: Alternative views of correlations between persons. *Psychometrika* **4**, 269–81.

Capdevila, R. and Stainton Rogers, R. 2000: If you go down to the woods today: narratives of Newbury. In Addams, H. and Proops, J., editors, *Social discourse and environmental policy.* Cheltenham: Edward Elgar.

Crossley, M.L. 2000: *Introducing narrative psychology: self, trauma, and the construction of meaning*. Buckingham: Open University Press.

Curt, B. 1994: *Textuality and tectonics: troubling social and psychological science*. Buckingham: Open University Press.

Fehr, B. 1988: Prototype analysis of the concepts of love and commitment. *Journal of Personality and Social Psychology* **55**, 557–79.

Good, J.M.M. 2000: *William Stephenson, Cyril Burt and the demise of the London school.* Paper presented at the History and Philosophy of Psychology section of the British Psychological Society, Annual conference, University of Ripon and York St John, UK April.

Harré, R. 1999: *The singular self: an introduction to the psychology of personhood*. London: Sage.

_____ 2004: Staking our claim for qualitative psychology as science. *Qualitative Research in Psychology* **1**, 3–14.

Harré, R. and Gillett, G. 1994: *The discursive mind.* London: Sage.

Harvey, O.J. 1997: Beliefs, knowledge, and meaning from the perspective of the receiver. In McGarty, C. and Haslam, S.A., editors, *The message of social psychology.* Oxford: Blackwell.

Haslam, S.A. and McGarty, C. 2003: *Research methods and statistics in psychology.* London: Sage.

Kitzinger, C. 1986: Introducing and developing Q as a feminist methodology: a study of accounts of lesbianism. In Wilkinson, S., editor, *Feminist social psychology: developing theory and practice*. Milton Keynes: Open University Press.

_____ 1987: *The social construction of lesbianism*. London: Sage.

Kline, P. 1994: *An easy guide to factor analysis.* London: Routledge.

Maraun, M.D. 1998: Measurement as a normative practice: implications of Wittgenstein's philosophy for measurement in psychology. *Theory and Psychology* **8**, 435–61.

Maxwell, A.E. 1977: *Multivariate analysis in behavioural research*. London: Chapman & Hall.

Moscovici, S. 1981: On social representations. In Forgas, J.P., editor, *Social cognition: perspectives on everyday life.* London: Academic Press.

Parker, I. 1997a: *Psychoanalytic culture: psychoanalytic discourse in Western society.* London: Sage.

_____ 1997b: Discourse analysis and psychoanalysis. *British Journal of Social Psychology* **36**, 479–95.

Polkinghorne, D.E. 1995: Narrative configuration in qualitative analysis. *International Journal of Qualitative Studies in Education* **8**, 8–25.

Potter, J. 1997: Discourse analysis as a way of analysing naturally occurring talk. In Silverman, D., editor, *Qualitative analysis: issues of theory and method*. London: Sage.

Potter, J. and Wetherell, M. 1987: *Discourse and social psychology: beyond attitudes and behaviour*. London: Sage.

Rogers, C.R. 1954: *Psychotherapy and personality change*. Chicago, IL: University of Chicago Press.

Schmolck, P. 2002: PQMethod (version 2.11). [Computer software]. Available at www.rz.unibw-muenchen.de/ ~p41bsmk/qmethod [last accessed: 28 October 2004].

Senn, C.Y. 1996: Q methodology as feminist methodology: women's views and experiences of pornography. In Wilkinson, S., editor, *Feminist social psychologies: international perspectives.* Buckingham: Open University Press.

_____ 1995: Repertory grids: an interactive, case-study perspective. In Smith, J.A., Harré, R. and Van Langenhove, L., editors, *Rethinking methods in psychology*. London: Sage.

_____ 1996: Beyond the divide between cognition and discourse: using interpretative phenomenological analysis in health psychology. *Psychology and Health* **11**, 261–71.

_____ editor, 2003: *Qualitative psychology: a practical guide to research methods*. London: Sage.

Smith, N.W. 2001: *Current systems in psychology: history, theory, research, and applications*. London: Wadsworth.

Stainton Rogers, R. 1995: Q methodology. In Smith, J.A., Harré, R. and Van Langenhove, L., editors, *Rethinking methods in psychology*. London: Sage.

Stainton Rogers, R. and Stainton Rogers, W. 1992: *Stories of childhood: shifting agendas of child concern*. Hemel Hempstead: Harvester-Wheatsheaf.

Stainton Rogers, R., Stenner, P., Gleeson, K. and Stainton Rogers, W. 1995: *Social psychology: a critical agenda.* Cambridge: Polity Press.

Stenner, P. 1993: Wittgenstein and the textuality of emotion. *Practice* **9**, 29–35.

_____ 1996: Shared pathways of Heidegger and Foucault. *Manifold* **3**, 12–17.

_____ 1998: Heidegger and the subject: questioning concerning psychology. *Theory and Psychology* **8**, 59–77.

Stenner, P., Dancey, C.P. and Watts, S. 2000: The understanding of their illness amongst people with Irritable Bowel Syndrome: a Q methodological study. *Social Science and Medicine* **51**, 439–52.

Stenner, P. and Eccleston, C. 1994: On the textuality of being: towards an invigorated social constructionism. *Theory and Psychology* **4**, 85–103.

Stenner, P. and Marshall, H. 1995: A Q methodological study of rebelliousness. *European Journal of Social Psychology* **25**, 621–36.

_____ 1999: On developmentality: researching the varied meanings of 'independence' and 'maturity' extant amongst a sample of young people in East London. *Journal of Youth Studies* **2**, 297–316.

Stenner, P. and Stainton Rogers, R. 1998: Jealousy as a manifold of divergent understandings: a Q methodological investigation. *European Journal of Social Psychology* **28**, 71–94.

_____ 2004: Q methodology and qualiquantology: the example of discriminating between emotions. In Tod, Z., Nerlich, B., Mckeown, S. and Clark, D., editors, *Mixing methods in psychology*. London: Routledge.

Stenner, P. and Watts, S. 1998: [Re]searching for love: subjectivity and the ontology of the Q factor. *Operant Subjectivity* **21**, 27–48.

Stephenson, W. 1935: Correlating persons instead of tests. *Character and Personality* **4**, 17–24.

_____ 1936a: The foundations of psychometry: four factor systems. *Psychometrika* **1**, 195–209.

_____ 1936b: The inverted factor technique. *British Journal of Psychology* **26**, 344–61.

_____ 1953: *The study of behaviour: Q technique and its methodology.* Chicago, IL: University of Chicago Press.

_____ 1983: Quantum theory and Q methodology: fictionalistic and probabilistic theories

conjoined. *The Psychological Record* **33**, 213–30.

_____ 1988/89: The quantumization of psychological events. *Operant Subjectivity* **12**, 1–23.

Sternberg, R.J. 1986: A triangular theory of love. *Psychological Review* **93**, 119–35.

Stricklin, M. and Almeida, J. 2001: PCQ: analysis software for Q-technique (Academic Edition, revised) [Computer software]. Available at http://www.pcqsoft.com/ [last accessed: 28 October 2004].

Watts, S. 2000: Critical psychology in the new millenium. *International Journal of Critical Psychology* **11**, 169–72.

_____ 2002: Stories of partnership love: Q methodological investigations. Unpublished Ph.D Thesis, University of East London.

Watts, S. and Stenner, P. 2003a: Q methodology, quantum theory and psychology. *Operant Subjectivity* **26**, 155–73.

_____ 2003b: On normativity, meaninglessness, and the centrality of the self: a reply to the comments of Good & Brown. *Operant Subjectivity* **26**, 182–89.

_____ 2005: The subjective experience of partnership love: a Q methodological study. *British Journal of Social Psychology* **44**.

Willig, C. 2001: *Introducing qualitative research in psychology: adventures in theory and method.* Buckingham: Open University Press.

About the authors

Dr Simon Watts is a Senior Lecturer in the Division of Psychology at Nottingham Trent University. Simon teaches qualitative research methods and personal relationships and publishes in the same areas. He completed a PhD in 2002 which examined the theory and potential of Q methodology as a qualitative method, with particular reference to its application in the context of partnership love. Simon has also taught in the Department of Human Relations at the University of East London, in the Departments of Psychology at University College Northampton and University College London, and he currently supervises Masters level qualitative dissertations for the Open University.

Dr Paul Stenner is a Lecturer in Psychology at University College London. Paul has been publishing Q methodological research for over a decade and in relation to a variety of subject matters, including jealousy, rebelliousness, love and I.B.S. Paul is a familiar name in qualitative psychology and is the coauthor of two related books. The first of these – 'Textuality and Tectonics' (written with Rex and Wendy Stainton Rogers and others, under the collective pseudonym 'Beryl Curt') – was acclaimed by reviewers as an innovative and groundbreaking text. The second – 'Social Psychology: A Critical Agenda' (coauthored with Rex and Wendy Stainton Rogers and Kate Gleeson) – was described by Kenneth Gergen as 'historically speaking, an enormously important publication'. Paul is currently researching in the area of ethics and the emotions. He has previously held posts at the Universities of East London and Bath.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.