CHAPTER 2

From Theoretical Concept to Survey Question

Joop J. Hox
University of Amsterdam

2.1 INTRODUCTION

Survey methodology has given much attention to the problem of formulating the questions that go into the survey questionnaire. Problems of question wording, questionnaire flow, question context, and choice of response categories have been the focus of much research. Furthermore, experienced survey researchers have written guidelines for writing questions and constructing questionnaires, and cognitive laboratory methods are developing means to test the quality of survey questions.

In comparison, much less effort has been directed at clarifying the problems that occur *before* the first survey question is committed to paper. Before questions can be formulated, researchers must decide upon the concepts that they wish to measure. They have to define what it is that they intend to measure by naming the concept, describing its properties and its scope, and describing how it differs from other related concepts.

Most (if not all) methodologists and philosophers will place the development of concepts for social research in the context of discovery, and not in the context of verification. Consequently, there are no fixed rules that limit the researcher's imagination, and concepts are not rigorously verified or falsified, but they are judged by their fruitfulness for the research process.

The process of concept-formation involves elaborating the concept and defining important sub-domains of its meaning. The next stage is finding empirical indicators for each concept or each subdomain. Empirical indicators

Survey Measurement and Process Quality, Edited by Lyberg, Biemer, Collins, de Leeuw, Dippo, Schwarz, Trewin.
ISBN 0-471-16559-X

could be viewed as "not-quite-variables," because at this stage we have not yet written specific questions or proposed a measurement (scale) model. The last stage is the actual question writing. This means that the conceptual content of each indicator must be translated fully and accurately into actual survey questions.

Although this chapter starts with a discussion of some philosophy of science issues, its main focus is the presentation of general strategies and specific techniques that can be used to structure the process of going from theoretical concept to prototype survey question. It is only after this stage that the usual concerns of intended population, question wording, *etcetera* begin to play a role in designing the questionnaire. These problems are not discussed here.

This chapter starts with a discussion of the theoretical substance of scientific concepts. Recent philosophy of science has come to the conclusion that scientific concepts are all laden with theory; this chapter discusses why and the subsequent consequences for conceptualization and operationalization. The next section discusses several approaches that have been proposed for question development, distinguishing between theory driven and data driven approaches. The last section provides a short summary and discussion.

2.2 SCIENTIFIC THEORY AND THE FUZZINESS OF SCIENTIFIC CONCEPTS

Two major aims of scientific research are to enlarge our knowledge of a specific field, and to solve a practical problem. In social science, survey research may be used for both purposes: the information that surveys provide may be used to explore or test theories, but also to shape or decide on policy. In both cases scientific theory plays an important role. Even in the most practical case we are generally not interested in the responses as such, but in their implications at some more abstract level. For instance, we may ask which candidate the respondent is going to vote for. This is a straightforward question, but the research aim is not just producing a frequency distribution, but rather inferring something about the general "voting intention" of the public, for instance, to predict the outcome of an election. In that respect, the question is not straightforward at all; we might attempt to improve our understanding of "voting intention" by asking the respondents how sure they are of their votes, and what other candidates they are also considering. Another example is the conventional question about the respondent's occupation. The objective of that question is usually not to find out what that person does during working hours, but to assign the respondent to a position on a Social-Economic Status (SES) ladder. Again, although the question is straightforward, the concept "SES" is not, and different conceptualizations of "SES" may lead us to different formulations of the occupation question.

The reason for shifting to a more abstract level in designing or interpreting

a survey is that scientific concepts are embedded in a more general theory that presents a systematic view of the phenomena we are studying. Instead of tabulating the popularity of candidates for our specific sample of individuals, we try to explain (predict) "voting intention" by linking it to various other concepts that indicate political attitudes. Instead of assigning the value "carpenter" to the variable "occupation," we assign "skilled labor" to the variable "SES."

The terms *concept*, *construct*, and *variable* are often used interchangeably. To understand the differences between these, I refer to a definition given by Kerlinger (1986, p. 9). "A theory is a set of interrelated concepts/constructs, definitions, and propositions that present a systematic view of phenomena by specifying relations among variables, with the purpose of explaining and predicting the phenomena." Thus, a theory is a system of propositions about the relations between concepts and constructs. A variable is a term or symbol to which values are assigned based on empirical observations, according to indisputable rules. The difference between a concept and a construct is small. A concept is an abstraction formed by generalization from similar phenomena or similar attributes. A construct is a concept that is systematically defined to be used in scientific theory (Fiske, 1971; Kerlinger, 1986). For example, "delinquency" is a concept that refers to an individual's tendency to exhibit behaviors that are classified as delinquent. "Deviance" is a construct that refers to an individual's tendency to exhibit behaviors that are classified as deviant. The difference is that "delinquency" can be defined by referring to those acts that our society classifies as delinquent, while "deviance" can only be defined in the framework of some theory (containing other constructs such as "norms" and "social reference group") that defines which behaviors are to be called deviant.

Since constructs and concepts are both theoretical abstractions, I use these terms loosely, with concept referring to abstract terms used in applied research and in early stages of theorizing, and construct referring to a more formal scientific concept used in theoretical research. This terminology emphasizes that the concepts used in social research, even if their name is a commonly used word (e.g., social class, disability, unemployment), are adopted or created by researchers for the purpose of their research. They have a theoretical "surplus meaning" of which the boundaries are not sharply drawn. Theoretical constructs have a rich surplus meaning. If we are working from an applied perspective, we are likely to work with concepts that have little surplus meaning. But there will be *some* surplus meaning, unless we are discussing the measurement of concepts like "sex" or "age" which are very close to their empirical measurement. Consequently, both concepts and constructs must be linked to observed variables by an operational definition that specifies to which variables they are linked and how values are assigned to these variables. This process is often viewed as a translation process; theoretical constructs are translated into observable variables that appear to be a suitable representation of that construct.

as one of the possible consequences of negative results that a construct may be abandoned, subdivided, or merged with other constructs. De Groot (1969) explicitly refers decisions about both theoretical constructs and operational definitions to "the scientific forum," that is, to ongoing discussion in the scientific community.

Kuhn (1970) and Lakatos (1978) both deal with the problem of the theory-dependence of observation statements in a slightly different way, by taking into account the way scientific theories evolve. As Chalmers (1976, pp. 63, 71) shows, if Popper's falsification principle had been strictly followed by scientists some of the most powerful theories, including Newton's gravitational theory, would have been rejected in their infancy. Kuhn (1970) maintains that most of the time scientific research is realized in a state of normal science, within a commonly accepted paradigm that contains the commonly accepted assumptions, laws and techniques. Apparent falsifications are regarded as anomalous results, to be resolved by further research and improvements in measurement and experimental techniques. If anomalies proliferate, a scientific revolution takes place, in which the anomaly-ridden paradigm is replaced by a new one. Lakatos (1978) views science as a world of competing research programs. A research program possesses a hard core: the main theoretical assumptions that are not to be rejected. This theoretical core is protected by a belt of auxiliary theory (including unstated assumptions about the accepted research methodology). Faced with a falsification, the theoretical core remains unchanged, and only the auxiliary theory is adjusted. Research programs that consistently lead to interesting new results are said to be *progressive*; programs that fail to do so are said to be *degenerated*. When a research program shows severe signs of degeneration, scientists leave it for a competing program (old theories never die, they just fade away). An example of a research program in Lakatos's sense in survey research is the cognitive approach to studying response behavior. The theoretical core is a set of cognitive theories and accepted information processing processes. The protective belt consists of the various adjustments needed to apply these theories in survey research. A falsification will generally not lead to abandoning the cognitive approach altogether, instead there will be adjustments in the auxiliary theory or research method. As long as the cognitive research program leads to interesting results, researchers will cling to it; when it fails, they will be attracted by another program (if available).

In both Kuhn's and Lakatos's views most research will be conducted in a state that can be described as normal or within a progressive research program. Thus, discussions about theoretical constructs and observations tend to be technical. A critical examination of a theoretical construct will probably focus on its dimensionality (should it be subdivided or merged with another construct) or its nomological network (how it is related to other constructs). New theoretical constructs will fit within a commonly accepted framework, and there are accepted guidelines for translating these into observable measures. Only in a state of theoretical crisis are we in the position described by Feyerabend as "anything goes."

It is interesting to note that operationism and logical positivism regard the fuzziness of theoretical constructs and the theory-dependence of observations as problems that must be eliminated, while Popper, Kuhn and Lakatos regard these problems as the heart of scientific progress. Especially Lakatos, whose views are usually denoted as "sophisticated falsificationism" (Chalmers, 1976), argues strongly that the critical discussion of constructs, their theoretical meaning, and relevant observations of their occurrence, is what brings about scientific progress in the form of new theories or new research questions.

2.3 CONCEPTUALIZATION AND OPERATIONALIZATION

As the previous section illustrates, both philosophers and methodologists have come to the conclusion that there are no pure observations. In the context of measurement in surveys, this means that there are no purely objective survey questions. Instead, there is a continuum that stretches from the theoretical to the observational. Some questions are so close to the observational end that for most purposes we may consider them as perfect measures for the construct. Examples would be questions about age or sex. But even these can be strongly theory-laden. Consider the case of "sex" in a gender study, or "age" in a study about calendar versus biological age, or in life cycle research.

In practice, faced with the problem of elaborating theoretical constructs and constructing questionnaire items, methodologists act as if there were a clear distinction between the theoretical language and the observational language. This leads to the sharp distinction between conceptualization and operationalization. Conceptualization involves concept formation, which establishes the meaning of a construct by elaborating the nomological network and defining important subdomains of its meaning. Operationalization involves the translation of a theoretical construct into observable variables by specifying empirical indicators for the concept and its subdomains. To bridge this gap between theory and measurement, two distinct research strategies are advocated: a theory driven or "top down" strategy, which starts with constructs and works towards observable variables, and a data driven or "bottom up" strategy, which starts with observations and works towards theoretical constructs (cf. Hox and De Jong-Gierveld, 1990).

2.3.1 From Conceptualization To Operationalization: Theory Driven Approaches

The theory driven strategies described in this section are dimension/indicator analysis, semantic analysis, and facet design.

2.3.1.1 Dimension/Indicator Analysis

Lazarsfeld (1958, 1972) outlines an approach that starts with a general and often vague concept. This is followed by a process of *concept specification*; the

2.2.1 Operationism

The scientific/philosophical approach known as *operationism* (Bridgman, 1927) requires that every construct completely coincide with one single observed variable. The operations needed to assign a value to a variable completely define the associated construct. In sociology, Bridgman's ideas were advocated by Lundberg (1939), while in psychology they were influential in the development of behaviorism (Watson, 1924). The essential principle in operationism is that the operational definition *is* the construct, and not a process to observe a construct after it has been defined (Lundberg, 1939). The consequence of the operationist approach is that if we change any aspect of the operational definition we have a new theoretical construct. Thus, if we measure heat with a mercury thermometer and with an alcohol thermometer we are measuring two different things. Similarly, if we change one word in our question, we are measuring a different construct.

The legacy of operationism to social science is a healthy emphasis on explicit definitions and measurement procedures. But the extreme position taken in operationism leads to hopeless problems if we attempt to generalize our results. A related problem in operationism is that we cannot compare two operational definitions to decide which one is better; they are merely different.

2.2.2 Logical Positivism

The various problems and inconsistencies in the operationalist approach have led to the conclusion that it is useful to distinguish between theoretical constructs and observed variables. This is a central notion in the philosophical approach known as *logical positivism*, which originated in the group of philosophers known as the Vienna Circle (cf. Hempel and Oppenheim, 1948). Logical positivism has been highly influential in shaping the ways social scientists think about the relationship between theoretical constructs and observed variables. A lucid exposition of the logical positivist position is given by Carnap (1956). Carnap distinguishes between a *theoretical language* and an *observation language*. The theoretical language contains the assumptions about the theoretical constructs and their interrelations. The observation language contains only concepts that are operationally defined or that can be linked to operationally defined concepts by formal logical or mathematical operations. The observation language should also be completely understandable by all scientists involved; there should be no discussion about the legitimacy of the observations. The connection between the theoretical and the observation language is made by correspondence rules, which are assumptions and inference rules. In Carnap's view the theoretical language is much richer than the observation language. Thus, a scientific theory may contain constructs that are defined solely by their relations with other constructs; in other words, constructs that have no operational definition. The view that the theoretical language is richer than the observation language is summarized by stating that the

theoretical language has surplus meaning with respect to the observation language. In the course of research, we may hope to reduce the surplus meaning by extending our knowledge, but there will always be some surplus meaning left.

In sociology, the influence of logical positivism is strong in the work of Northrop (1947), Lazarsfeld (1972), and Blalock (1982). Northrop distinguishes between concepts-by-intuition, which can be immediately understood upon observation, and concepts-by-postulation, which derive their meaning at least in part from theory. The two kinds of concepts are linked by epistemic correlations, which are analogous to Carnap's correspondence rules. Blalock identifies Northrop's concepts-by-intuition as terms in the observational language, and concepts-by-postulation as terms in the theoretical language. Both kinds of terms must be linked by "auxiliary measurement theory," which can be interpreted as the theory that describes the assumptions and inferences that together form Carnap's correspondence rules. In psychology, the influence of logical positivism is reflected in the distinction between theoretical constructs that are embedded in a nomological network, and operational definitions that specify how some theoretical constructs are measured. A key concept in psychometric theory is the *construct validity* of a measure. Construct validity indicates how well a specific measure reflects the theoretical construct it is assumed to measure (Cronbach and Meehl, 1955; Campbell and Fiske, 1959). Just like the epistemic correlation, the construct validity is not a parameter that can be estimated as some specific value; it refers to the general quality of the correspondence rules.

2.2.3 Sophisticated Falsificationism

The views of logical positivism have been strongly attacked by philosophers like Popper, Kuhn, and Feyerabend. Popper is probably best known for introducing the falsification principle; science grows by formulating theoretical conjectures and exposing these theories to the risk of falsification (Popper, 1934). However, Popper (and other philosophers) also criticized logical positivism for its belief that there are such things as direct, unprejudiced observations. As Popper makes clear in many examples (Popper, 1934; see also Chalmers, 1976), observation statements are always made in the language of some theory. Precise observations require explicit and clearly formulated theories. And, if theories are fallible, the observation statements made in the context of those theories are also subject to falsification.

If observations are theory-laden, straightforward falsification is impossible, because both theories and observations may be false. Popper solves this problem by subjecting not only theories but also observation statements to criticism and testing. Observations that have survived critical examination are accepted by the scientific community—for the time being (Popper, 1934). This position has been adopted by most social science methodologists. For instance, Blalock (1968), referring to Northrop, argues that concepts-by-intuition are accepted by consensus among the scientists in the field. Cronbach (1971) lists

concept is divided into separate components or dimensions. The subdimensions form a further, more detailed specification of the meaning of the theoretical construct. The subdimensions of a theoretical construct may be derived logically, or they may be based on the empirical results from previous research. Next, a set of empirical indicators is defined for each subdimension. For practical research, a subset of indicators is chosen, and usually summated to a total score for each subdimension. Since in Lazarsfeld's view the indicators for a specific subdimension are highly interchangeable, the precise choice of indicators is not very important.

Lazarsfeld does not explain how we should actually achieve concept specification. Fiske (1971) describes a procedure that is analogous to Lazarsfeld's approach. In Fiske's view, an explicit conceptual specification must lead to a well-defined construct that poses a clear target for operationalization and permits multi-component operationalization. Just as Lazarsfeld, Fiske assumes that we start with a general and probably vague concept. The first step is to identify the theoretical context of this concept. Often, concepts are taken from an existing general theory, which provides the theoretical background. Fiske (1971, p. 97) favors this approach because it builds on what has already been worked out conceptually and on previous empirical research. The researcher's task here is to clarify the theoretical background and the way the concept is extended to fit a specific research problem. Researchers may also formulate their own concept, in which case they must elaborate the theoretical context and demonstrate the fruitfulness of the concept and its advantages over similar concepts.

The second step in Fiske's approach is to delineate the core of the construct, the unique quality to which the construct refers. The theoretical core is the essence of the concept, it must cover all phenomena that the concept refers to. This requirement often leads to a theoretical construct that is much too broad. Such a broad theoretical construct can better be considered as a general label that refers to a conglomerate of subconstructs. In practice, it will usually be preferable to examine the theoretical meaning of the subconstructs, and construct separate measures for each of these. Fiske describes several strategies for defining the core meaning, which can be subsumed under the heading *differentiation*. Thus, it is important to state explicitly what is *not* included in the construct, and the differences with other related constructs. Also, we may attempt to describe the opposite pole of the construct, and decide whether this defines the low end of the scale, or whether it introduces a new quality. For instance, if we want to measure "life satisfaction," is the low end of the scale defined by absence of satisfaction, or by presence of much dissatisfaction?

The third step in Fiske's approach is to construct a measurement instrument. For this, it is helpful to make a number of distinctions before the construction process starts. One is the modalities across which the measure is assumed to generalize: may we expect consistency across self-description, description by others (proxy-measurement), and observation? Guttman's facet design (described in more detail later) can be used to make systematic distinctions within modes,

for instance by distinguishing different situations or behavior types. Fiske also recommends explicit decisions about measurement scales and measurement designs.

Evidently Fiske's approach has much in common with Lazarsfeld's. Lazarsfeld's subdimensions resemble Fiske's subconstructs. Lazarsfeld does not make clear why we would want to distinguish subdimensions. Fiske relates this to the research problem. In applied research, we may have a highly specific purpose, for which constructs with limited meaning are adequate. In basic research, we may want to use more general constructs, especially in the early stages of research. Later, we may want to specify subdimensions to achieve higher precision. The choice depends on the research problem and the state of knowledge in a given field, and may also be made on pragmatic grounds.

Although the dimensional/indicator analysis approach appears complicated, it is not difficult to find examples in survey research. An insightful discussion of the specification of several key concepts in social research can be found in Burgess (1986). A good example is the study by Andrews and Withey (1976, pp. 11–14) on subjective well-being. In a general sort of way, we know what well-being is. For some research purposes, we might measure it by asking respondents a single question about their well-being. In general, we would prefer to be more precise. Andrews and Withey proceed by discussing a conceptual model for well-being that distinguishes between different domains of and different criteria for well-being.

Andrews and Withey's domains are based on pilot research about people's concerns, aimed at finding life-domains that are distinguishable and cover most of people's concerns. The result is a list of domains, such as housing, job, family life, and neighborhood. The criteria are the values, standards, or goals by which one judges how one feels about the various domains. Their criteria are further differentiated into absolute and relative criteria, and general and specific. Examples of criteria are standard of living, independence, fun, and safety. Figure 2.1 illustrates a specific conceptualization (cf. Andrews and Withey, 1976, p. 234). The final instrument design also includes considerations about the type of response scales to be used. Given all these distinctions, Andrews and Withey were able to construct a large number of survey questions that tap different aspects of subjective well-being, and to relate these both theoretically and empirically to the theoretical constructs. An example of a question generated by this design is "How do you feel about the beauty and attractiveness of your house?" (Andrews and Withey, 1976, p. 13).

2.3.1.2 Semantic Analysis

Sartori (1984) notes that conceptualization is mediated by language, which makes it important to study the semantic structure of our statements. Verbal theories take the linguistic system through which they are expressed for granted. An important aspect of the linguistic system is the vocabulary. Theories use an existing vocabulary that reflects a global perception of the world. This is important, because what is not named is often not noticed, and the choice of

Dictionary). In social science the meaning is more specific: evolution of human systems to more complex, mature, or higher forms, or in political science: transformations of political systems, notably at the macro level. An analysis of earlier definitions of "development" in political science (23 different definitions) leads to the conclusion that in political science the construct "development" is so overloaded with overlapping meanings that it is virtually useless. In a series of definitions "development" is changed into such constructs as "change," "growth," "modernization," and many more. In some instances this improves the original definition by making it less ambiguous.

Semantic analysis does not directly lead to survey questions. However, it does help to disentangle different meanings and to recognize ambiguity in our constructs. Since semantic analysis works with language, it is language dependent. This can be an advantage, for instance in examining the meaning of translated measures in cross-cultural research. For example, a semantic analysis of "happiness" in Dutch leads to an examination of the concept of "fortunate," because the Dutch word *geluk* combines the meanings of the English words "happy" and "lucky." Thus, a Dutch questionnaire about "geluk" could well contain questions about beliefs on being "lucky," while an American questionnaire about "happiness" would be unlikely to contain such questions.

2.3.1.3 Facet Design

A useful device for the systematic analysis of a construct is Guttman's facet design (Guttman, 1954). Facet design defines a universe of observations by classifying them with a scheme of facets with elements subsumed within facets. Facets are different ways of classifying observations, the elements are distinct classes within each facet. The universe of observations is classified using three kinds of criteria: (1) the population facets that classify the population, (2) the content facets that classify the variables, and (3) the common range of response categories for the variables.

For our present goal, we concentrate on the facet structure of the variables. The various content facets can be viewed as a cross-classification, analogous to an analysis of variance design, that specifies the similarities and dissimilarities among questionnaire items. Each *facet* represents a particular conceptual classification scheme that consists of a set of elements that define possible observations. The common *response range* specifies the kind of information required. For instance, attitude questions typically have a response range formulated as: very negative . . . very positive, and value questions typically have a response range formulated as: very unimportant . . . very important.

The content facets must be appropriate for the construct that they define. In selecting the most appropriate facets and elements, the objective is to describe all important aspects of the content domain explicitly and unequivocally. For example, for many constructs it may be useful to distinguish a behavior facet that defines the relevant behaviors, and a situation facet that defines the situations in which the behaviors occur. In facet design, the facet structure is often verbalized by a mapping sentence, which describes the observations in

Table 2.1 Example of Mapping Sentence

"To what extent does person (X) feel that

Source		Reason
(her own experience)	led her to	(feel healthier)
(her husband)	believe that	(feel fitter)
(her doctor)	she would	(be more physically attractive)
(the media)		(have fewer clothing problems)
		(suffer less social stigma)
		(be less anxious in social situations)
		(feel less depressed)

	Response
if she lost weight, as rated	(not really at all)
	(not very much)
	(to a slight degree)
	(to a fair degree)
	(quite a lot)
	(very much)
	(very much indeed)

where (X) are married women attending slimming groups"

(Source: Gough, 1985, p. 247)

one or more ordinary sentences. An example of a facet design with mapping sentence is Gough's (1985, p. 247) mapping sentence for reasons for attending weight reduction classes, which is given in Table 2.1.

In this facet design the first facet (source) refers to the source of the belief, and the second facet (reason) refers to a specific consequence of losing weight. A facet design as given above can be used to generate questionnaire items. The (X) facet points to a target population of individuals. The source facet has four elements, the reason facet has seven, which defines $4 \times 7 = 28$ questions. For example, combining the first elements of the source and reason facets leads to the survey question "Did your own experience lead you to believe that your health would improve if you lost weight?" (Gough, 1985, p. 257).

Facet design is part of a more general approach called facet theory, which uses the facet structure to generate hypotheses about similarities between items, and for the analysis and interpretation leans heavily on geometric representations (Shye *et al.*, 1994; Borg and Shye, 1995). Facet theory contains no general guidelines to determine the need for introducing specific facets. Instead, it

emphasizes that a facet design is a definition, which means that it should not be judged in terms of right or wrong, but whether it leads to productive research. Also, facet designs must be constructed with the research problem firmly in mind. Thus, as a device to specify the meaning of a construct, facet design assumes that we already have a good notion of the empirical domain we want to investigate. For instance, Gough's facet definition given above is derived from the literature and pilot research.

Usually, facet design is employed to specify the denotative meaning of a construct; the design specifies the set of empirical referents, and the actual survey questions are derived from the mapping sentence. However, the distinctions made in a dimension/indicator or semantic analysis can be introduced as distinct facets in a facet design, in which case the facet design also refers to the connotative meaning. In both cases, using a facet design encourages a systematic and exhaustive conceptual analysis. Facet designs can also be used to classify the questions in an existing instrument.

2.3.2 From Operationalization To Conceptualization: Data Driven Approaches

The data driven strategies described in this chapter are content sampling, symbolic interactionism, and concept mapping.

2.3.2.1 Content Sampling

An alternative approach to bridge the gap between theoretical construct and actual survey questions is to assemble a large set of questions which appear relevant to the research problem, collect data, and determine which questions co-vary and thus seem to measure the same construct. This is a thoroughly inductive approach. We assume that we know enough of the subject to be able to assemble questions that share one or more conceptual cores, and use factor analysis or comparable techniques to determine what these common cores are, and which questions measure them. Starting from data, we infer theoretical constructs, which may be altered in subsequent empirical research.

The data driven approach to measurement and conceptualization may seem rather superficial and *ad hoc*. Nevertheless, it can lead to useful standard scales, as exemplified by Schuessler's analysis of well-being scales (Schuessler 1982). Schuessler analyzed over 140 scales used in American sociology between 1920 and 1970, which contained over 9500 questions. He found that many of these scales were *ad hoc*, in the sense that they were used only once or twice. There were many overlapping items. A disturbing finding was that conceptually different scales could have several items in common. Altogether there were about 500 distinct questions. In an effort to construct a set of standard scales to measure well-being, Schuessler selected 237 items. A factor analysis on the responses of a sample of 1522 respondents revealed 17 factors, which in the end led to 12 scales based on 95 items.

As Fiske (1971) points out, if we use the results of a factor analysis to infer constructs and relationships between constructs, including factoring of subsets

of questions to search for subfactors, we are working along lines very similar to the dimensional/indicator approach, only the starting point is different. The danger to avoid in this type of data driven approach is *reification* of the factors that are found. That is, factors should not be interpreted as fixed causal entities, but rather as convenient descriptive categories that summarize response consistencies (Anastasi, 1985). From a theoretical viewpoint, the interesting question is why these patterns arise. Verbal labels and interpretations of the factors depend on our explanation for the process that leads to these response contingencies.

In content sampling, it is essential not to miss any important aspects of the research domain, and consequently the usual approach is to assemble a *large* set of items. For instance, Cattell (1973) started the development of a mood scale by taking from the dictionary all words that can be used to describe a mood. A random sample of these words was used to construct questionnaire items, and subsequent research led to a mood scale with several subscales, such as "evaluation" and "arousal." Cattell *et al.* (1970) followed a similar approach in the construction of the Sixteen Personality Factors test (16 PFT). Because one starts with a large set of questions, factor analysis is often used to search for underlying dimensions. The goal of such an analysis is to identify sets of questions that measure the same construct. Both Fiske (1982) and Marradi (1981) have pointed out that if a set of questions measures the same construct, a factor analysis of this set should produce one single factor. So, the recommended analysis approach proceeds in two steps. In a first step, potential sets of questions are identified by exploratory factor analysis or similar means. In the second step, these sets are analyzed separately, to test whether a single factor is indeed sufficient.

In the previous paragraphs, "factor analysis" has been used as a general term for any technique designed to examine the similarities between questionnaire items in terms of dimensionality or scalability, as a first stage in data driven construct formation. As a technique, factor analysis has been criticized severely (e.g., Duncan, 1984) because it does not test dimensionality very strictly, and because it does not actually produce a measurement. Instead, item response models such as the Rasch model are preferred. Still, this does not affect the warnings given above against reification of latent variables and misinterpretation. Searching for sets of items that scale well according to some scaling model such as the Mokken or Rasch model merely identifies observed response consistencies. Theoretical interpretations of the scales depend on our explanation of why these response contingencies arise.

2.3.2.2 Symbolic Interactionism

The social sciences often use concepts that are taken from common language. These concepts often have a manifest function in our social life. As researchers, we may decide to attach a more precise meaning to such a concept to make it more useful in scientific theory (Fiske, 1971; Kerlinger, 1986). Still, if we aim to describe or explain the experiences and behaviors of individuals in society,

we should understand how these concepts are used and understood in social life. Thus, one approach to construct formation is to start with common language concepts as they are used by ordinary people in daily life, and bring these into the scientific discourse.

This approach is congruent with the social research paradigm known as symbolic interactionism. Symbolic interactionism views the social world as a world of interacting individuals, who are constantly negotiating a shared definition of the interaction situation (cf. Blumer, 1969). Symbolic interactionist research attempts to discover individuals' interpretations of social interactions, by developing concepts that describe these interpretations. This leads to a qualitative research strategy that focuses on the processes of interaction and interpretation. The starting points for conceptualization are *sensitizing concepts*; vague notions that do no more than point the research process in some promising directions. Qualitative field research gives these sensitizing concepts a more detailed empirical reference and clears up the connections between the various concepts.

This comes close to Sartori's aim to clarify the denotative and connotative meaning of theoretical constructs (Sartori, 1984). The difference is that Sartori mainly directs his attention to theoretical clarification by semantic analysis, and symbolic interactionists prefer to remain as close as possible to the individuals and their social worlds. This preference is well epitomized by the term "grounded theory" coined by Glaser and Strauss (1967).

Symbolic interactionist research often uses unstructured interviews and naturalistic observation. The data are the recorded material: field notes, interview texts, or observation protocols. Starting from the sensitizing concepts, the development and empirical specification of the concepts proceeds by examining the data and investigating which concepts and basic processes seem to be involved. During the analysis, new data may be collected to amass evidence about the value of new concepts. The output of this investigation is a file of "emerging concepts" that is updated constantly when new insights arise or new data come in. The process is stopped when the concepts appear *saturated*, that is, no new aspects of concepts can be found. The analysis is concluded by an effort to reduce the set of concepts to a smaller set of key concepts, which are then integrated in an overall theory by establishing their relationships.

Symbolic interactionism and other qualitative approaches have not evolved into a unified method of social research. However, they have led to methods that are also useful in developing concepts with a view of further operationalizing these into survey questions. The general procedure is to use open interviews to collect information from respondents, followed by an analysis as outlined above. Methods are available that bring rigor and reliability to the coding process (Miles and Huberman, 1994), and computer programs are available to support the analysis. The use of symbolic interactionist methods to construct survey questions is also discussed by Foddy (1993).

An example of the symbolic interactionist approach is Schaeffer and Thomson's (1992) description of grounded uncertainty. They start with the observation that respondents in surveys express uncertainty in different ways, such as "don't know," "doesn't apply to me," or "it depends." It appears that such phrases point to two different types of uncertainty: state uncertainty (respondents are uncertain about their true state) and task uncertainty (respondents are uncertain how to communicate their true state). Schaeffer and Thomson use semi-structured telephone interviews with standardized open questions in a study of fertility motivation. Their analysis roughly follows the symbolic interactionist methods outlined above. They arrive at core concepts such as: neutrality, lack of clarity, ambivalence, indecision, and truly mixed expressions. In the end, they recommend to expand the Don't Know category in the closed questionnaire, and to measure the different types of state uncertainty directly.

Another example from a different field is Sternberg's research into the meaning of "intelligence" (Sternberg, 1985). Sternberg asked respondents to list typical "intelligent" and "unintelligent" behaviors. From these lists, 250 behaviors were selected. A factor analysis on a second sample produced three factors: "practical problem solving ability, verbal ability, and social competence" (Sternberg, 1985, p. 60). An interesting observation was that the factor structures of lay people and experts were very similar, which leads to the conclusion that it is possible to pose questions about intelligence to ordinary respondents.

2.3.2.3 Concept Mapping

A disadvantage of the symbolic interactionist approach is the cost in time and effort. Another weakness is its lack of structure, which makes it difficult to separate the respondents' contribution from the researcher's. The basic method, to have respondents formulate their own thoughts on a topic, can be applied in a more direct way. With a view to constructing questionnaire items, the symbolic interactionist research paradigm can be reduced to two leading principles: generation of statements and structuring of statements. The generation stage is explorative, and aims to produce a comprehensive list of statements that ideally should span the entire domain of interest. The structuring stage seeks to reduce the statement list to a smaller list of key concepts subsuming the former.

Focus groups have often been used to find out what concepts respondents use to describe their experiences or reactions. In such focus groups, carefully selected groups of respondents discuss a central topic, and examine it from several different points of view. Typically, focus groups are composed of individuals from the target population. The sampling criterion is not so much *representativeness* as *variance*; a wide variety of individuals is included. Sometimes specific groups are used, such as people who have certain situations or experiences in common. The discussion in the focus group is recorded, and analyzed later to extract statements about the domain of interest.

The statement list can be structured by the researchers, or by collecting more data from the respondents. If the number of statements is small, all pairwise

combinations can be presented to respondents who are asked to assign a similarity rating to each pair. The resulting similarity matrix can be subjected to multidimensional scaling or other similarity analyses to determine how the statements cluster into more abstract concepts. Since with k statements, the number of paired comparisons is $(k \times (k-1))/2$, paired comparisons are not attractive with more than about ten statements. If the number of statements is large, a sorting task is often used to examine the empirical structure of the statements. In a sorting task, the respondents must sort the statements in different clusters of mutually similar statements. The sorting task can be completely unconstrained, in the sense that the respondent simply must put the statements in a number of different clusters, but has complete control over the number of clusters and the number of statements per cluster (usually it is not permitted to put all statements in one cluster, or to produce as many clusters as there are statements). The sorting task can also be constrained, as for example, in natural grouping analysis. Natural grouping analysis starts with all statements in one large group. The respondent must split this group in two, and explain how the two new groups differ. Next, one of the available groups must be split further, and again the respondent must explain how the two new groups differ. This process is repeated until the respondent feels unable to make further splits. By tabulating how often statements are grouped together, it is possible to construct a similarity matrix, which can then be subjected to some kind of similarity analysis.

Various approaches have been proposed to systematize concept mapping, for an overview see Jonassen *et al.* (1993). Trochim (1989) has developed a comprehensive system for concept mapping in focus groups, which involves six steps: (1) developing the focus, (2) statement generation, (3) statement structuring, (4) statement representation, (5) concept map interpretation, and (6) utilization of concept maps. In step 1 "developing the focus" one must specify the primary focus or domain of interest, and select the participants for the focus group. Step 2 "statement generation" is a brainstorming session with the focus group to generate statements that describe all relevant aspects of the domain of interest. Trochim (1989) gives as an example a brainstorming session with as focus "... the issues, problems, concerns or needs which the elderly have ..." (Trochim 1989, p. 4). In step 3 "structuring" the individual participants are instructed to sort the printed statements into different piles "the way that makes sense to you." The participants are also asked to rate the statements as to importance. The individual sorts are combined in a group similarity matrix. Step 4 "statement representation" subjects this similarity matrix to a multidimensional scaling, and superimposes a cluster solution on the map to facilitate interpretation. In fact, clusters may be viewed as "emerging concepts." It is also possible to overlay the importance ratings upon the concept map. An example of the concept map from Trochim's (1989) study about the elderly is given in Figure 2.2. Step 5 "concept map interpretation" is again a group activity. The participants discuss possible meanings and acceptable names for each statement cluster. They are also asked to see whether there are plausible groups of clusters

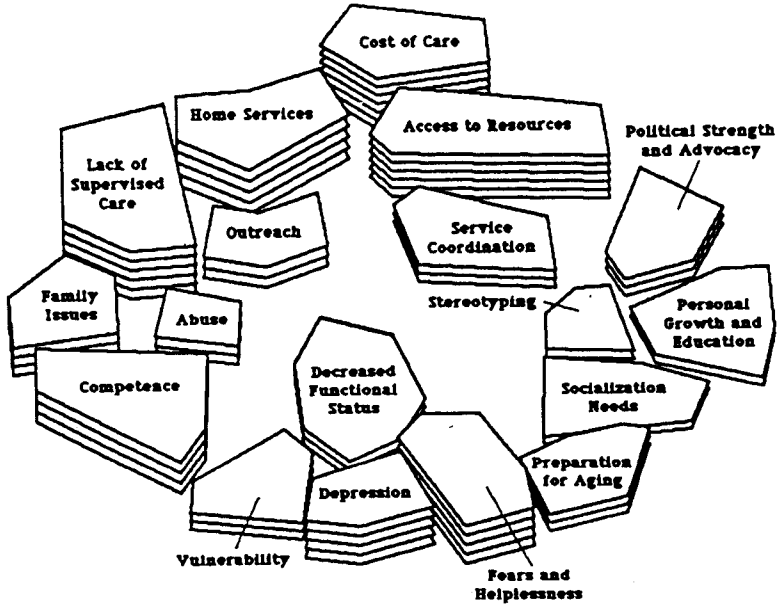

Figure 2.2 Concept mapping of concerns about elderly. (Reprinted from W.M.K. Trochim, An introduction to concept mapping for evaluation and planning, *Evaluation and Program Planning*, 12, pp. 1–16, © (1989), with kind permission from Elsevier Science Ltd., The Boulevard, Langford Lane, Kidlington OX5 1GB, U.K.)

or regions in the cluster space that can be named. Basically, this step attempts to identify relations between concepts in the form of a group-approved concept map. Step 6 "utilization" applies the final concept map to plan evaluations, interventions, *etcetera*. In our case, utilization would mean using the individual statements to translate the concepts into survey questions, and using the concept map to provide a first hypothesis as to the empirical structure of the questionnaire.

2.4 SUMMARY AND DISCUSSION

Theoretical constructs used in social science are often complex, and have an indirect relationship to the empirical observations they are designed to explain. The distance between a theoretical construct and observable phenomena creates the problem that researchers must state how they plan to measure what they are theorizing about. In the context of survey research, the link between theory and data is provided by formulating survey questions that are related to the constructs in the theory at hand.

This chapter presents two global strategies to conceptualization and question construction. The first strategy gives a logical priority to the theory and the constructs embedded in it. The researcher starts with the theoretical constructs, and their attached surplus meanings. To arrive at specific survey questions, it is necessary to clarify the surplus meaning of the construct. Three different approaches are described that guide this process. The first is dimension/indicator analysis. This approach has been advanced by many methodologists, such as Lazarsfeld and Fiske, and is implicit in the writings of many others such as Blalock, Cronbach, and Meehl. Many researchers probably view it as *the* approach. In dimension/indicator analysis we subdivide the central concept into subconcepts and specify empirical indicators for each subconcept. The second approach relies on semantic analysis to specify the various related meanings of a construct, and to find empirical referents for each subconstruct. The third approach is facet design, which lays out all aspects of a construct's content by specifying a facet structure and a common response range.

The second strategy gives logical priority to the empirical observations. Of course, there must be some notion of a central issue, but that can be vague. From the observations, the researcher must extract theoretical constructs. Three different approaches are described that guide this process. The first is content sampling. In this case, there is already a large set of questions that appears to relate to the central issue, or such a set can easily be assembled. Next, factor analysis or related methods are used to condense the large set of questions to a smaller set of theoretical constructs. The two other approaches rely on common language. The first adapts research methods derived from symbolic interactionism to the problem of construct formation. The other uses concept mapping methods, derived from cognitive research.

The merits and demerits of theory driven versus data driven strategies are discussed at length in Hox and De Jong-Gierveld (1990). Note that except for the case when we start with a set of questions to be explored, the various approaches do not directly produce survey questions. In the theory driven strategy, we end with a list of specifications which are (we hope) sufficiently tangible that the translation into survey questions is a comparably routine operation. In the data driven strategy, we end with a list of fairly concrete statements and their relationships to more abstract constructs. Compared to the theory driven approach, we have the advantage that we usually also have raw material that is condensed in these statements. This raw material often contains literal transcripts of respondents' verbalizations, which can often be turned into survey questions. Another advantage of the data driven strategy is that one can start with the respondent's understanding of a particular concept. However, with the data driven approach we run the risk of ending up with theoretical constructs that do not fit well into the theories that we wish to use.

In the end, researchers faced with the job of constructing valid measures may find comfort in Fiske's (1971) position that the actual choice for a specific strategy may not be all important. Constructing valid measures is not a one-time exercise. Rather, it is a continuous process of devising measures, investigating their theoretical and empirical meaning, and devising better measures. For a discussion of the importance of a systematic approach toward constructing measures for survey use see the chapter by Heath and Martin (Chapter 3). At the end of the process, the actual starting point is probably not crucial (cf. Oosterveld and Vorst, 1995). Important is that we must start at some sensible point and let *both* theoretical analysis and empirical tests guide our progress. In doing this, we shuttle back and forth between the theoretical and the empirical ends of a continuum.

The end product of concept formation can be characterized as a *prototype survey question.* The implication is that we still have to address the usual concerns of intended population, question wording, and so on. The issues involved in constructing the questionnaire and testing it in the cognitive laboratory and the field are treated in this section.

ACKNOWLEDGMENTS

I thank Edith de Leeuw, Anthony Heath, Jean Martin, Norbert Schwarz, and Arie van Peet for their comments on earlier versions. I thank the members of the Dutch Research Committee on Conceptualization and Research Design collectively for many stimulating discussions about the topics presented in this chapter.

REFERENCES

Anastasi, A. (1985), "Some Emerging Trends in Psychological Measurement: A Fifty-year Perspective," *Applied Psychological Measurement*, 9, pp. 121–138.

Andrews, F.M., and Withey, S.B. (1976), *Social Indicators of Well-being*, New York: Plenum.

Bridgman, P.W. (1927), *The Logic of Modern Physics*, New York: Macmillan.

Blalock, H.M. Jr. (1968), "The Measurement Problem: A Gap Between the Languages of Theory and Research," in H.M. Blalock, Jr., and A.B. Blalock (eds.), *Methodology in Social Research*, New York: McGraw-Hill, pp. 5–27.

Blalock, H.M. Jr. (1982), *Conceptualization and Measurement in the Social Sciences*, London: Sage.

Blumer, H. (1969), *Symbolic Interactionism. Perspective and Method*, Englewood Cliffs: Prentice-Hall.

Borg, I., and Shye, S. (1995), *Facet Theory: Form and Content*, Newbury Park, CA: Sage.

Burgess, R.G. (1986), *Key Variables in Social Investigation*, London: Routledge.

Campbell, D.T., and Fiske, D.W. (1959), "Convergent and Discriminant Validation by the Multitrait-multimethod Matrix," *Psychological Bulletin*, 56, pp. 81–105.

Carnap, R. (1956), "The Methodological Character of Theoretical Concepts," in H. Feigl, and M. Scriven (eds.), *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, Minneapolis: University of Minneapolis Press.

Cattell, R.B. (1973), *Personality and Mood by Questionnaire*, San Francisco: Jossey-Bass.

Cattell, R.B., Eber, H.W., and Tatsuoka, M. (1970), *Handbook for the Sixteen Personality Factor Questionnaire*, Champaign, IL: Institute for Personality and Ability Testing.

Chalmers, A.F. (1976), *What Is This Thing Called Science?* Milton Keynes: Open University Press.

Cronbach, L. (1971), "Test Validation," in R.L. Thorndyke (ed.), *Educational Measurement*, Washington, D.C.: American Council on Education, pp. 443–507.

Cronbach, L., and Meehl, P.E. (1955), "Construct Validity in Psychological Tests," *Psychological Bulletin*, 52, pp. 281–302.

De Groot, A.D. (1969), *Methodology. Foundations of Inference and Research in the Behavioral Sciences*, The Hague, NL: Mouton.

Duncan, O.D. (1984), *Notes on Social Measurement. Historical and Critical*, New York: Sage.

Fiske, D.W. (1971), *Measuring the Concepts of Personality*, Chicago: Aldine.

Fiske, D.W. (1982), "Convergent-discriminant validation in measurements and research," in D. Brinberg, and L.H. Kidder (eds.), *Forms of Validity in Research*, San Francisco: Jossey-Bass, pp. 77–92.

Foddy, W. (1993), *Constructing Questions for Interviews and Questionnaires*, New York: Cambridge University Press.

Glaser, B.G., and Strauss, A.L. (1967), *The Discovery of Grounded Theory*, Chicago: Aldine.

Gough, G. (1985), "Reasons for Slimming and Weight Loss," in D. Canter (ed.), *Facet Theory*, New York: Springer, pp. 245–259.

Guttman, L. (1954), "An Outline of Some New Methodology for Social Research," *Public Opinion Quarterly*, 18, pp. 395–404.

Hempel, C.G., and Oppenheim, P. (1948), "The Logic of Explanation," *Philosophy of Science*, 15, pp. 135–175.

Hox, J.J., and De Jong-Gierveld, J.J. (eds.) (1990), *Operationalization and Research Strategy*, Lisse, NL: Swets & Zeitlinger.

Jonassen, D.H., Beissner, K., and Yacci, M. (1993), *Structural Knowledge*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Kerlinger, F.N. (1986), *Foundations of Behavioral Research*, New York: Holt, Rinehart & Winston.

Kuhn, T.S. (1970), *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press.

Lakatos, I. (1976), *Proofs and Refutations*, Cambridge, UK: Cambridge University Press.

Lakatos, I. (1978), *The Methodology of Scientific Research Programmes*, Cambridge, UK: Cambridge University Press.

Lazarsfeld, P.F. (1958), "Evidence and Inference in Social Research," *Daedalus*, 87, pp. 99–130.

Lazarsfeld, P.F. (1972), *Qualitative Analysis. Historical and Critical Essays*, Boston: Allyn & Bacon.

Lundberg, G.A. (1939), *Foundations of Sociology*, New York: Macmillan.

Marradi, A. (1981), "Factor Analysis as an Aid in the Formation and Refinement of empirically useful concepts," in D.J. Jackson, and E.F. Borgatta (eds.), *Factor Analysis and Measurement in Sociological Research*, London: Sage, pp. 11–49.

Miles, M.B., and Huberman, A.M. (1994), *Qualitative Data Analysis: An Expanded Sourcebook*, Newbury Park: Sage.

Northrop, F.S.C. (1947), *The Logic of the Sciences and the Humanities*, New York: World Publishing Company.

Oosterveld, P., and Vorst, H. (1995), "Methods for the Construction of Questionnaires: A Comparative Study," *Proceedings of the International Conference on Survey Measurement and Process Quality*. Alexandria, VA: American Statistical Association, pp. 269–274.

Popper, K.R. (1934), *Logic der Forschung*, Wien: Springer. Reprinted as Popper, K.R. (1968), *The Logic of Scientific Discovery*, New York: Harper & Row.

Riggs, F.W. (1984), "Development," in G. Sartori (ed.), *Social Science Concepts*, Beverley Hills, CA: Sage, pp. 125–203.

Sartori, G. (1984), "Guidelines for Concept Analysis," in G. Sartori (ed.), *Social Science Concepts*, Beverley Hills, CA: Sage, pp. 15–85.

Schaeffer, N.C., and Thomson, E. (1992), "The Discovery of Grounded Uncertainty: Developing Standardized Questions About the Strength of Fertility Motivations," in P.V. Marsden (ed.), *Sociological Methodology* 1992, 22, pp. 37–82.

Schuessler, K.F. (1982), *Measuring Social Life Feelings*, San Francisco: Jossey-Bass.

Shye, S., Elizur, D., and Hoffman, M. (1994), *Introduction to Facet Theory*, Newbury Park, CA: Sage.

Sternberg, R.J. (1985), *Beyond IQ*, Cambridge, MA: Cambridge University Press.

Trochim, W.M.K. (1989), "An Introduction to Concept Mapping for Evaluation and Planning," *Evaluation and Program Planning*, 12, pp. 1–16.

Watson, J.B. (1924), *Behaviorism*, Chicago: University of Chicago Press.