

## Getting from Neuron to Checkmark: Models and Methods in Cognitive Survey Research

BREGJE C. HOLLEMAN<sup>1\*</sup> and JAAP M. J. MURRE<sup>2</sup>

<sup>1</sup>*Utrecht Institute of Linguistics, Utrecht University, Utrecht, The Netherlands*

<sup>2</sup>*Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands*

### SUMMARY

Since the 1980s much work has been done in the field of Cognitive Survey Research. In an interdisciplinary endeavour, survey methodologists and cognitive psychologists (as well as social psychologists and linguists) have worked to unravel the cognitive processes underlying survey responses: to improve survey measurement, but also to obtain fundamental insight into the process of question answering, the nature of attitudes, the functioning of human memory, etc. Yet, despite the amount of work that has been done, less progress has been made than was deemed possible. This paper suggests ways in which to develop cognitive theories that may help to obtain an integrated understanding of question answering in surveys. One way to build better theories is to focus on model construction. Different types of cognitive models are discussed that can be used to model question answering. Second, a larger variety of methods from the cognitive toolbox can be used to further develop and test these models. An overview is given of the various tools available to researchers of cognitive aspects of survey research, with a special focus on newer methods from cognitive neuroscience. To connect different tools and methods into an integrative theory of question answering, the idea of hierarchical modelling is proposed. Copyright © 2007 John Wiley & Sons, Ltd.

Jabine, Straf, Tanur, and Tourangeau (1984) started to build a bridge between the disciplines of survey methodology and cognitive psychology. In their book ‘Cognitive Aspects of Survey Methodology’, they reported on two seminars (Maryland 1983 and Baltimore 1984) on Cognitive Aspects of Survey Methodology (CASM), thus starting the so-called CASM-movement. The movement aims to strengthen the collaboration between cognitive scientists and survey researchers for the advancement of both fields. In this paper, we will discuss the current status of the CASM route and review ways in which its construction can be facilitated.

Standardized surveys are widely used in the social and political sciences (including psychology and communication studies), politics and marketing. Yet it turns out that many factors ‘distort’ the answers to these questions in seemingly inexplicable ways. For many decades, research has shown time and again that answers are influenced by many, apparently superficial characteristics of questions (see overviews such as Ayidiya & McClendon, 1990; Hippler & Schwarz, 1987; Molenaar, 1982, 1991; Sudman & Bradburn, 1974). Slight differences in question wording cause huge differences in the answers: Schuman and Presser (1981) demonstrated, for example, that subtle differences in the tone of

\*Correspondence to: Bregje C. Holleman, Utrecht Institute for Linguistics OTS Utrecht University, Trans 10, 3512 JK Utrecht, The Netherlands. E-mail: Bregje.Holleman@let.uu.nl

wording cause large differences in the answers obtained: a question worded negatively elicits many more ‘no’-answers than the number of ‘yes’-answers obtained by an equivalent question phrased positively. The observed scores are also affected by the order of the questions (e.g. Tourangeau & Rasinski, 1988), the number or order of the response options (e.g. Schwarz, Knauper, Hippler, Noelle-Neumann, & Clark, 1991), the visual design of the survey (e.g. Tourangeau, Couper, & Conrad, 2004), and the behaviour of the interviewer (e.g. Van der Zouwen, Dijkstra, & Smit, 1991). The example below, taken from Sniderman and Theriault (2004), shows how the context in which an issue is presented (e.g. a context of quality of life and egalitarianism vs. one of economic growth and individual costs) greatly influences the reported attitudes towards welfare: a difference of about 30% (Figure 1).

A great deal of research shows these types of effects, often due to even more subtle variations in the questionnaire. And needless to say, the differences in observed answers lead to differences in the conclusions one would draw from these surveys. So then the issue is: which question is a more valid measure of respondents’ ‘true attitudes’ or ‘true opinions’? Cognitive psychology (as well as related disciplines, such as social psychology or linguistics) can describe the causes of these types of instability across slightly different measures of the same underlying attitudes, focusing on issues such as text processing, interpretation and communication and the nature of attitudes.

In addition to this cross-sectional approach, other research (e.g. Van der Veld and Saris, 2004) shows that reported attitudes are subject to change across measurements at different points in time—in panels, for example. In order to obtain more insight into the causes of this drift, methods are needed that can distinguish between measurement error and stable aspects of the observed scores. Also, we need theories to describe what causes attitude instability and what causes measurement error. Cognitive psychology can help to expose whether perhaps some people hold no opinion at all, and answer randomly—as Converse (1964) would say. Or whether it is wrong to presume that attitudes are stable constructs, and are merely associative networks, constructed on the spot (as Anderson, 1993, would suggest. See also Schwarz & Bohner, 2001).

Applying cognitive and communicative theory to survey research can help to reduce response error or can facilitate a valid interpretation of the responses obtained. Also, the application of cognitive, communicative and memory principles to problems in survey

<p><b>Two question versions with different frames</b></p> <p><b>Version a.</b> (Egalitarianism) Are you in favor of or opposed to a big increase in government spending to increase opportunities for poor people, so they can have a better chance of getting ahead in life? Yes/No</p> <p><b>Version b.</b> (Economic Growth) Are you in favor of or opposed to a big increase in government spending to increase opportunities for poor people, even if it means higher taxes? Yes/No</p> <p><b>Effect:</b> This results in up to 30% more ‘Yes’ answers for version a, due to the framing of the question.</p>
--

Figure 1. Example of a framing effect on question answering. Adapted from Sniderman and Theriault (2004, pp. 144–145)

methodology can advance cognitive theory, because the techniques used to examine how respondents understand and answer survey questions offer efficient and controlled ways to study real-life processes of cognition and communication, to test existing hypotheses or to formulate new ones. These are the two basic ideas behind the CASM movement.

Since 1984, much interesting work has been done on combining cognitive theory with survey research (see various books and theme issues of journals, such as Hippler, Schwarz, & Sudman, 1987; Jobe & Loftus, 1991; Tanur, 1992; Schwarz & Sudman, 1995; Sudman, Bradburn, & Schwarz, 1996), culminating in the most recent integrative work of Tourangeau, Rips, and Rasinski, (2000). In spite of these advances, however, a lot of the construction of the bridge between cognitive science and survey research remains unfinished.

As Belli (2005) pointed out in a call for a special issue on CASM, survey methodologists have not been as responsive as they could have been with respect to the potential role that applied cognitive psychology could have in their work. At the same time, applied cognitive psychologists have not been as knowledgeable as they should be concerning the interesting and relevant work that survey methodologists are conducting (Belli, 2005, p. 245). The cognitive survey movement is not guided sufficiently by cognitive theory. There *is* a dialogue, but survey researchers are not receptive enough to the potential role of applied cognitive psychology, while the psychologists are not listening enough to the practical needs of survey researchers.

Yet, there is a persistent and undoubtedly growing interest in this endeavour, as is evident from a constant stream of publications and well-attended sessions and seminars on this topic (such as sessions at the conferences of the American Association of Public Opinion Research, of the Conference on Logic and Methodology, RC33, of the European Association for Survey Research and the theme group working on cognition and survey research at NIAS, the Netherlands Institute for the Advanced Study in the Humanities and Social Sciences, in 2003/2004).

In 2000, Tourangeau, Rips, and Rasinski formulated the problem as follows: the cognitive approach has provided a new paradigm for understanding survey errors, as well as many new methods for detecting survey errors—but this deeper understanding did not yet bring much improvement in survey design, due to four factors. First of all, according to Tourangeau et al., the cognitive movement has focused on a limited set of issues (mainly retrieval of attitudes or memories) at the expense of others, which may have directed research away from the largest sources of errors, such as non-response, or the interaction between interviewer and respondent. Second, the CASM movement may have suffered from a problem encountered often in interdisciplinary research: they breathed new life into already discredited theories or outmoded methods in survey research long after (cognitive) psychology disposed of them for good reasons. Third, cognitive psychology may be better at diagnosing the cause of response problems than at curing them. And finally, remedies for response problems proposed by cognitive psychologists may work in the lab but not necessarily in the real-world (Tourangeau et al., 2000, pp. 337–340).

Tourangeau et al.'s analysis of the pitfalls of cognitive survey research still applies today. Many of the flaws and problems they mention are general and apply to all interdisciplinary research, so it would be rather optimistic to think that CASM would have overcome these problems in the past 5 years. We think the problems they outline will have to be a constant focal point in any disciplinary research endeavour, hence also in CASM research. In this paper, we will sketch several research approaches that may help to overcome the second problem Tourangeau et al. sketch: the use of 'Wrong Theories'.

With 'the use of wrong theories', Tourangeau, Rips, and Rasinski refer to the fact that interdisciplinary researchers tend to use discredited theories as well as outmoded methods. Psychoanalysis, for example, is a theory that took on a second life in literary theory long after psychology abandoned it. The application of schemata theory to survey phenomena would be an example of the use of discredited theories, or at least too vague a theory, in the field of survey research. As an example of outmoded methods they say that CASM researchers accept the results of protocol analysis (cognitive interviewing) rather uncritically, based on 'looser data collection methods and less formal coding procedures than most psychologists would allow', whereas in cognitive and social psychology 'there is a keen appreciation of the limits of verbal protocols' (*ibid.*: 338).

We will propose ways to build better theories and make more interdisciplinary progress. In our view, this can be achieved by focusing more on the construction of new cognitive models and relating experimental results explicitly to these or existing models. In Section Modelling Question Answering, we will give an overview of different types of models and theories that are being used in (cognitive) psychology and discuss ways to use these in CASM research to help advance insight into the process of question answering as a whole.

Another way to advance in the field of CASM is to make better use of the cognitive toolkit of research methods, to triangulate and get a full view of the problem at hand. In Section The Toolbox of Cognitive Survey Research, we will outline a number of research methods available to the CASM researcher. The inevitable advantages and disadvantages of each method suggest that it is often sensible to triangulate and use different methods to investigate a single phenomenon. Furthermore, we will give various examples of new tools available from cognitive neuroscience that are not yet used frequently in CASM research but that do offer promising possibilities.

Different types of models are used to describe different kinds of problems, and all too often the relation between the problems and models is largely under-specified. Thus, we obtain small islands of knowledge, of problems that are being solved, without a view on question answering in a real-life survey context as a whole. To improve this, researchers should aim for coherent (i.e. internally consistent) as well as coordinated (externally consistent) cognitive models. This will be discussed in more detail in the section 'Towards an Integrated Model of Question Answering'.

These proposals will work best if collaborative structures are set up in order to maintain true interdisciplinary cooperation, so that the survey researchers keep the research focused on the real-world problems and the cognitivists can come up with up-to-date theories and methods. We will elaborate somewhat on this point in the Discussion.

All in all, the approaches proposed here aim to gain a deeper understanding into the processes that go on during survey question answering. Such an understanding is not only interesting in itself, it can also provide a basis to develop guidelines for survey design and survey testing.

## MODELLING QUESTION ANSWERING

A model is a schematization and formalization of some domain of knowledge. We will briefly discuss four types of models here that are established in the cognitive sciences: box-and-arrow models, connectionist models, rule-based models and mathematical models.

### Box-and-arrow models

The most prevalent models are the so called box-and-arrow models. These are simple boxes that each denote some type of process, such as 'Interpretation' or 'Decision'. The different processing stages are connected by arrows, typically denoting which stages are dependent on others. Though these models impose some kind of schema on the theorizing, they take only minimal steps towards formalization. One could call them proto-models. A danger of this approach is that it is often unclear what exactly is supposed to go on inside the boxes; the formalization is still incomplete. Nonetheless, they serve a very useful purpose in a field where little prior modelling has been done.

Tourangeau et al. (2000), for example, propose a four-stage model of question answering, which is still largely in the box-and-arrow stage:<sup>1</sup>

- Comprehension of the question;
- Retrieval of memories or attitudinal information;
- Deriving a judgment;
- Map judgment onto a response option.

On the basis of this model, many predictions can be derived that may guide research. For example, a question about some major news event would tend to have a shorter retrieval time immediately after the event than a year later. The other stages, however, should not be affected by the event latency. Reaction time distributions could be used to verify such hypotheses (e.g. Bassili, 1996; Chessa & Holleman, 2007; Draisma & Dijkstra, 2004).

Box-and-arrow models in themselves may fuel a good deal of empirical research before they become the basis for more sophisticated models. But formalization is necessary to specify what goes on inside the boxes. In the case of the four-stage model above, it is unclear, for example, whether the stages are strictly consecutive or whether they may overlap; one stage starting before the other has been completed. The precise dynamics of the timeline are not made apparent either. Such aspects could be captured in a mathematical version of this model.

In order to come to a more formalized model from box-and-arrows models, such as the Tourangeau, Rips, and Rasinski model of question answering, the model should not just be used as a source of inspiration or as a framework to position research. Instead, research should be conducted to formulate restrictions on the model, or to refine it. For example, recently, Van Berkum, Holleman, Murre, and Nieuwland (2006) tested whether the first stage of question comprehension and the second stage of attitude retrieval are really separate stages. Results indicate that these stages are more intertwined than the model seems to suggest at present.

Of the more formalized models three types dominate in cognitive psychology: connectionist models (also known as neural network models), rule-based models and mathematical models. The behaviour of connectionist and rule-based models can rarely be derived by means of formal analysis (i.e. deductive reasoning and formal proofs as is custom in mathematics), which is why they are evaluated by computer simulations. Mathematical models typically—but not always—allow a purely formal analysis of the

<sup>1</sup>Other examples of box-and-arrow models of survey answering include the question-answering model by Cannell, Miller, and Oksenberg (1981), the model of information processing in survey situations, by Strack and Martin (1987). See Callegaro (2005) for a (historical) overview of cognitive models in the survey domain.

main aspects of the model behaviour. This has the advantage of knowing with absolute certainty what the characteristics and limitations of a model are. A disadvantage is that the analysis techniques can become very complex and advanced training in mathematics is often required to understand the derivations. This may hamper both their development and their dissemination.

### Connectionist models

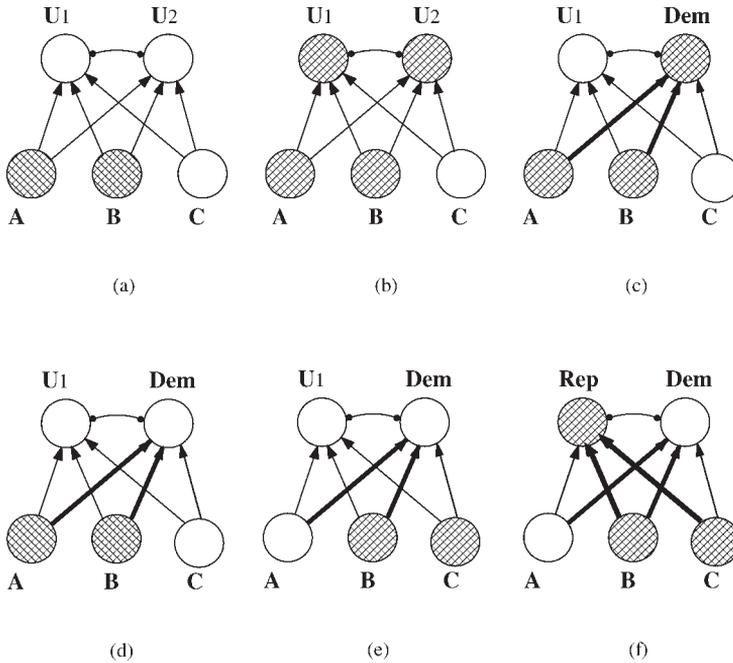
Connectionist models use a brain metaphor: they consist of large numbers of artificial neurons (nerve cells) connected in dense networks. The neurons exchange pulses over the connections and will fire when sufficiently strong pulses arrive at their input side. 'Firing' means: distributing a pulse to neurons connected to the output side of a neuron. The connections have different efficiencies (weights), which may be inhibitory (negative weights). Since the early 1980s, an extensive body of knowledge has emerged regarding the characteristics of such neural networks. A principal finding is that the *function* or behaviour of a neural network is determined by its *structure*, that is, the network of connections with different weights. Neural networks are self-programming in the sense that they can develop suitable connection weights by merely being exposed to the domain in which they need to function. An example of this is presented in Box 1, where two political categories are formed. Category formation proceeds without a teacher telling the system that 'this is a left-wing candidate'; they are formed on the basis of regularities in the input. Though this is a very simple example, very complex and quite realistic systems have been built in this manner, for example, for the categorization of speech sounds or visual patterns.

**BOX 1.** A highly simplified example of learning in a neural network. The network has two uncommitted artificial neurons,  $U_1$  and  $U_2$ , which initially do not represent anything. Each neuron receives weak inputs from three neurons A, B and C. The activation level of these neurons represents the presence of specific political statements, as follows: A: 'The Government should socialize medicine'; B: 'The Government should improve foreign relations'; C: 'The Government should increase defence spending'.

The network represents a member of the public who is politically naive. In (a) he reads about a political candidate Bill who argues for socialized medicine and improved foreign relations. This activates input neurons A and B. In (b), we see how two uncommitted neurons are both activated. The neurons strongly inhibit each other (shown with connections represented by filled black circles). In (c), due to small, random variations in connection strength neuron  $U_2$  has won the competition. The strong inhibition allows it to suppress the activation of neuron  $U_1$ . After the categorization on node  $U_2$  has been achieved, the network's connections are adjusted by increasing connections between activated neurons (so called Hebbian learning). It also weakens the connection between the not activated neuron C and neuron  $U_2$  (shown by a dashed line). We can now characterize node  $U_2$  as representing a 'democratic' viewpoint. In (d), he would read about another candidate with similar viewpoints, this would immediately activate the Democrat neuron.

In (e), he reads about another candidate George who also wants to improve foreign relations and also increase defence spending. This activates input neurons B and C. In (f), we see how George's viewpoints have activated neuron  $U_1$ . Hebbian learning has adjusted the connections and node  $U_1$  can now be characterized as representing certain Republican viewpoints.

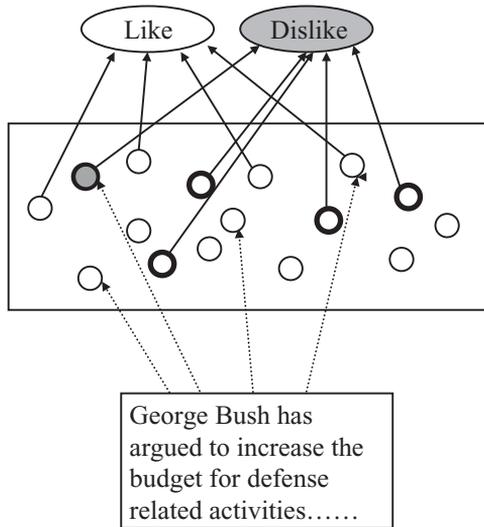
The network has thus formed two political categories. A more realistic example would include many more statements, candidates, and output neurons. It would be interesting to see how such a system would acquire political categories, especially if the input somehow resembled real input such as reaches the public.



As we shall see in the next section, with rule-based models, if one tries to model the *process* of question answering, it is nearly always necessary to make certain assumptions about the underlying knowledge structures. With broad box-and-arrow models or highly abstracted mathematical models, these assumptions remain implicit, but as soon as one tries to incorporate details of some real-world domain, like answering questions about political facts, the way in which this knowledge domain is represented will have a strong bearing on the types of processes that are possible. A classic example of this is the model of letter recognition by McClelland and Rumelhart (1981). Many of the details of the recognition process emerge from the interactions between massively parallel structures of the neural network model. Such knowledge structures may be entered 'by hand', but this often invites a suspicion of tinkering. It is more elegant to train or construct a model on the basis of patterns (e.g. political facts) as they may be encountered in real life. In this way, one can create 'artificial subjects', neural networks that have slightly different initial parameters due to their different learning histories, mimicking how real-life subjects might come into a survey interview. In the next section, we will describe a recent rule-based model by Kim, Lodge, and Taber (2004, 2007), called *John Q. Public*, which follows this exact approach. They remark that we know surprisingly little about how individual citizens actually form and revise their political beliefs and attitudes. Once artificial subjects have been created, we can investigate the process of question answering, much like we can investigate the process of letter perception (e.g. a subset of all English words). A consequence of such more detailed modelling is, thus, that we must direct more research efforts to the study of development and change of knowledge structures.

Connectionist models have, as far as we know, not yet been applied to modelling large-scale question-answering processes. This, however, is perfectly feasible as neural networks can learn on the basis of training exemplars from the task domain and their behaviour generalizes well to unseen exemplars. One could, for example, train a model on a certain knowledge domain, like political candidates. Networks could be made ‘Democrat’ or ‘Republican’, among others, by spending more time studying candidates of their political affiliation. During the training phase, the networks could be prompted on different aspects of knowledge and their answers could then be compared with those of human subjects. Different aspects of attitude formation could be built in as is illustrated in the model in Box 2, which gives a Like/Dislike response when presented with political statements (also see the recent connectionist model of attitude formation and change by Van Overwalle & Siebler, 2005).

**BOX 2.** Here we see a network that has been trained with various political statements and has been given explicit ‘Like’ or ‘Dislike’ signals. This particular network has been trained with a Democrat orientation. Different political facts are extracted from the input text. Each fact activates a corresponding artificial neuron, as in Box 1. Given an input statement as shown here, one or more neurons will be activated. For the sake of clarity ‘Dislike’ statement neurons have been made thicker. Depending on the political history of the model, as reflected in the connections that have been learned, the Like or Dislike response neuron will be more highly activated. If the difference in activation between the two neurons exceeds some preset threshold, the network will respond with either Like or Dislike. If the input statement contains more positively evaluated political facts and fewer negative ones it will respond faster. A model like this could be developed into a full-fledged model of political attitudes that is able to produce attitudes as a function of input text and political history of the model. It could also produce reaction time distributions.



Though the implicit promise of neural networks is that they reflect the functioning brain, this is often not the case. There is a spectrum from very biologically oriented models—so called ‘single neuron models’ that possess many biological details—to highly abstracted artificial neurons that have little to do with real neurons. Most neural networks in cognitive science lie somewhere in between, capturing some aspects of brain-like processes and structures while being biologically implausible in other respects. Thus, a political attitude may be represented by one or two ‘neurons’, whereas it is clear that in reality many groups of neurons will be involved. Such representations must, therefore, always be considered preliminary, though they may suffice when we are more interested in the interaction of attitudes or their formation rather than in the nature of the representations.

Smith (1999) discusses some implications of connectionist models for modelling question-answering processes. Searching memory to answer an attitude question cannot be regarded as the processes of looking for a single desired representation and retrieving it. Instead, retrieval is a reconstructive process that seeks to satisfy multiple constraints simultaneously. So if a survey question is asked about how you feel about Bush’s presidency so far, many different stored representations contribute to reconstructed memories of events during his presidency, but also related general knowledge, memory traces of related events, etc. The flow of activations through all the connections contributes to the final output. How this would work exactly, and how the output would differ depending on different question inputs, could be modelled and refined in a connectionist network. Such networks need to be simulated on a computer because the exchange of signals occurs in a massively parallel process, causing it to be generally impossible to predict their behaviour exactly. It is, however, possible to repeat the simulations under many slightly different conditions so that a thorough understanding of the model’s behaviour can be developed, much like in human subjects, except of course that in those we cannot manipulate the structure of the brain.

### Rule-based models

In contrast to neural network models, rule-based models do not use a brain metaphor. Rather, they look at the brain as one large processing system with a short-term memory, which—like in humans—has a limited capacity. The system consists of a set of rules that are typically of the type: ‘IF *Condition X* holds for the working memory, THEN the working memory is modified with *Action Y*’. Actions may add intermediate results to the working memory or remove results.

Sometimes, the short-term memory store is not limited explicitly, but results (‘chunks’ of data) are given an ‘activation value’ that increases sharply when a rule is triggered and that subsequently decays with time. In this way, activations can spread through a ‘network’ of triggering rules. A difference between rule-based spreading activation and neural processes in neural networks, is that single rules can perform more complicated operations than neurons (e.g. logical operations such as AND, OR and XOR—exclusive or—or higher-order logic). Complex cognitive architectures can be built with rules, for instance the ACT-R system by Anderson (1993; Anderson & Lebiere, 1998). ACT-R offers a rule-based framework in which specific models can be formulated. The nice aspect of this approach is that existing cognitive modules can be re-used in new models. The timing and other constants of these modules can often remain the same, limiting the number of free parameters. Learning in these systems is achieved by tuning rule parameters and through the addition of new rules by the programmer.

An example of a recent rule-based model, based on ACT-R, is a model of political cognition by Kim et al. (2004), named *John Q. Public*. The essence of their model is the working hypothesis that accessibility of an issue (e.g. some political fact) in long-term memory *at the moment of answering* will determine the respondent's internal perception about the association between a political candidate and issue position. This momentary accessibility of the issue is determined by cognitive and affective factors, including affective congruence. This concept refers how similar two concepts (or a candidate and a concept) are from an evaluative point of view. Thus, thinking about a negative concept like cancer, facilitates retrieval of other negatively valued concepts. The model was tested on real data of candidate evaluation.

In a more recent paper, Kim et al. (2007) tested their model via the 'artificial subjects' method mentioned above on data by Lau and Redlawsk (2001; Redlawsk, 2002). In this experiment, subjects processed a different set of campaign information about four fictitious but plausible candidates. As a result, each subject processed different information about each candidate. Intermittently they took a number of polls. Kim et al. fed the model with transcribed versions of the data used in the Lau–Redlawsk experiment, creating the same number of artificial subjects that took part in the original experiment. Each new piece of campaign information processed by the model affected the knowledge structure of the artificial subjects. They found, as they had expected, that the more an issue in memory got activated at the moment a survey question was asked, the stronger the reported perception about the association between the candidate and the issue. In general, the results of the model correlated well with the data.

### Mathematical models

A disadvantage of rule-based models is that they can become very complex, making it progressively harder to evaluate their behaviour. Learning neural networks suffer from the same problem with the added disadvantage that, whereas rule-based models can at least print a trace of all rules that fired during processing, a neural system can be a hard to disentangle network of weights and activations. Mathematical models try to avoid this problem by sacrificing many details of cognitive processing, characterizing major components of the cognitive system with just a handful of parameters.

For example, a neural network could be used to model forgetting by training a network with many different patterns. Forgetting might be induced by interference from new learning or we could add more and more noise (random values) to the weights of the network to simulate the progression of time. This would allow us to study many details of exactly what aspects of pattern interference play a role and how forgetting progresses over time. A mathematical model of forgetting would aim to characterize the forgetting process in such a way that its characteristics can be proven mathematically. Such proofs may be long and complicated. Sometimes they are impossible to find and the approach must be abandoned (or computational approximation may be used). Ideally, after proofs have been derived, we end up with an equation that describes, for example, the shape of forgetting. If such a result is obtained, it can offer considerable insight into fundamental properties of the processes and phenomena studied. It is also a powerful tool for practitioners, because they can use it to fit their data, predict the shape of the data, signal outliers in the data, etc. Most mathematical models are developed within a specific formalism of the larger field of mathematics, for example, linear (matrix) algebra or differential equations.

An example is the model for answering attitude questions (see Chessa and Holleman, 2007). The model aims to characterize the dynamics of answering (contrastive) attitude questions, that is, the storage, retrieval or recall of attitudinal information and the response production. The model builds on to the Memory Chain Model by Chessa and Murre (2002), a neurobiologically inspired mathematical model of memory encoding, forgetting, retrieval and recall. The assumptions of this model are based on very highly simplified neural events and structures, modelled using a specific mathematical framework (for mathematically inclined readers: non-homogeneous point processes). A point process is a probabilistic model for the occurrence of 'events' in space and/or time, which are represented as 'points'. In the neurosciences, point processes have been used to describe series of firing neurons over time (Abeles, 1991). The variable of interest in the Memory Chain Model is the strength of a memory representation of some stored item, which is quantified as the number of 'points' in memory over time. In the model for answering attitude questions, the strength of a memory representation refers to the attitude(s) the question intends to measure. The new model not only includes the strength of this representation, but also the valence of the evaluation (either positive or negative). Furthermore, the new model contains equations for modelling the response process. The behaviour described in the model was derived in a purely analytical manner (i.e. without computer simulations), parameters were estimated on part of a data set containing responses and reaction times to attitude questions that were worded positively. The result is a bunch of equations that directly specify, for example, reading time and answering time distributions to these questions. These equations were then fitted to another subset of the empirical data (equivalent questions worded negatively), which showed that the model could in fact adequately predict response times and response distributions for those questions. Mathematical models can be very powerful when there are many interacting variables (such as the strength of the attitude representation, the valence of that representation, the question length, etc.) that can be mapped onto the parameters of the model.

Though cognitive models such as mentioned above are starting to make inroads in to survey research, in our view a limitation is that there is often only a very weak relation between the theories used in different models or in different sub parts of the same model. Even if the same words are used, they often mean different things. This leads to many isolated models with little clue as to how they can be applied together to explain the multifaceted process of answering survey questions. This will be elaborated on in Section Towards an Integrated Model of Question Answering. Furthermore, for an adequate use of most modelling techniques discussed here, specific skills and experience are used. One way to acquire that while keeping in touch with the specific needs and goals of survey practice, is to work together in multidisciplinary teams (Section Discussion). But first the methodological toolkit of the cognitive researcher will be discussed in some more detail to obtain a good view on the broad range of research methods available to the CASM researcher to develop and test cognitive models.

## THE TOOLBOX OF COGNITIVE SURVEY RESEARCH

Researchers who study the cognitive processes underlying survey responses have a large toolbox of methods at their disposal. All of those tools can be used to test or further develop the models discussed in the previous section. For example, conclusions drawn from

eye-movement data can be used to describe respondents' behaviours or cognitive activities with a connectionist model, but think-aloud data can serve this purpose too. Each tool has its specific advantages and shortcomings. Therefore it is important for CASM researchers to be able to use more than just one tool to study a given phenomenon from several angles. First of all, there are the survey data themselves. These can be aggregated over persons and allow comparisons between groups. This can make available to the researcher the phenomenon that he wants to explain, such as the effect of the inclusion of a Don't Know-option: one random sub-sample answers a question with a DK-option, the other random sub-sample answers the equivalent question without such an option, and comparison of the answers between groups shows whether the use of a DK-option influences respondents' answering behaviour. Aggregated data can also be used to test certain cognitive hypotheses. For example, people's attitude strength is generally hypothesized to influence whether they are sensitive to characteristics of the question wording. This idea can be tested by using known groups of respondents with a strong and with a weak attitude towards a given issue, and comparing aggregated data (e.g. Wegener, Downing, Krosnick, & Petty, 1995).

For hypotheses concerning cognitive processes, a within-person design is usually stronger than a between-groups design, as these processes take place within one person. One can use survey questions to obtain data within persons and compare the answers. Repeated measurements can be obtained within persons to estimate which part of the answer reflects a consistent 'true score'—an attitude—and which part reflects measurement error (e.g. Andrews, 1984; Van der Veld & Saris, 2004). A disadvantage of within-persons designs is, however that repeated measurements may themselves influence the opinions reported. More complicated designs, for example, with rotating question orders, can be used to control for those effects.

Cognitive interviewing (Willis, 2005) and behaviour coding (Fowler & Cannell, 1996) can also be used to glean something of respondents' cognitive processes. In a cognitive interviewing, a questionnaire pre-testing method, concurrent or retrospective think-aloud tasks produce reports of the respondents' thoughts either while answering the survey question or directly after. In many survey organizations cognitive labs pre-test their questionnaires using cognitive interviewing. Difficulties in understanding a question's meaning will often be verbalized, but a lot of debate and research is still going on about best practice and the assessment of results (Presser et al., 2004). Under which circumstances does verbalization alter respondents' thought-processes (Willis, 2004)? How does interviewer behaviour influence the number and type of problems respondents raise (Conrad and Blair, 2004)? Also behaviour coding or interaction coding will give insight into the process of question answering. As it consists of monitoring interviews or reviewing taped interviews or transcripts of the interviewer's and respondent's verbal and non-verbal behaviour, it shows the ways in which respondent and interviewer negotiate about a question's meaning or about how to translate a given opinion or behaviour into a pre-coded answering option (e.g. Dijkstra, 1999; Holbrook, Cho, & Johnson, 2006; Van der Zouwen et al., 1991; Van der Zouwen & Smit, 2004). Not all aspects of every problem can be verbalized, however, or will become visible in the interaction.

Reaction times are yet another measure used to gain insight into cognitive processes. They are, for example, assumed to reflect the cognitive difficulty of a task (Fazio, 1990); the longer it takes to perform a task, in milliseconds or seconds, the more cognitive effort is assumed to be invested. So, if respondents take longer to answer a question about the frequency of a behaviour that took place various times in the last year, and take less time to

answer a question about a behaviour that only happened occasionally a short time ago, one could infer that the answer to the first question took longer because it took more effort or more mental steps were needed to retrieve it from Long Term Memory (e.g. Bassili, 1996; Mulligan, Grant, Mockabee, & Monson, 2003; see also Yan & Tourangeau, 2007, for a recent example). On the other hand, very short reaction times can indicate that no cognitive effort took place at all, and strong satisficing<sup>2</sup> took place (Krosnick, 1991). An advantage of both interaction coding and reaction time measurements is that the data can be obtained unobtrusively, without distorting the actual task of question answering. A disadvantage of reaction time measures is that longer reaction times do not necessarily only reflect prolonged cognitive processing (although recently research has shown that ‘Goldilocks ranges’ can be identified outside of which responses are more likely to be a less valid answer to the question (see Ehlen, Schober, & Conrad, 2007) and that it is difficult to obtain very precise reaction time measurements outside the research lab.<sup>3</sup>

Survey questions, interaction or behavioural coding and reaction times are now established measures of cognitive processing in survey research. Newer methods from cognitive psychology are less widely used, but could well be added to the survey researchers’ toolbox.

Eye-tracking is one of those newer methods. With eye-tracking, one literally follows people’s eye movements. The idea is that the points that are fixated on longer, are points that receive more attention and thus require more cognitive processing. So eye-tracking can provide insight into the words in a question that take longer to process, perhaps because they are more difficult to understand. Also, eye-tracking could be used to investigate to what extent question answering is indeed the linear process Tourangeau et al.’s model seems to suggest. Do respondents attend to the answering options only after they have formed a judgment? Or do the response options play a role earlier in the process of question answering as well, for example, during the comprehension stage to disambiguate the meanings of the question text? A disadvantage of eye-tracking is that the data analysis is rather elaborate. A positive development is that the devices needed to measure eye-movements are getting increasingly less obtrusive. For recent examples of eye-tracking applied to survey research, see Graesser, Cai, Louwerse, and Daniel (2006) and Galesic et al. (2006). Graesser et al. (2006), for example, used eye-tracking to obtain on-line insight into the process of question comprehension. They found that difficult words (unfamiliar technical terms) had longer first fixation durations and more fixations than content words that were defined to be non-technical terms in their research. They found it difficult, however, to detect vague or imprecise terms based on their eye-tracking data.

Physiological measures too can give important insight into the processes involved in question answering, or the nature of the underlying attitudes. If an issue causes arousal, this is detectable by fluctuations in skin temperature and subtle signs of sweatiness. Galvanic skin responses are one way of measuring this. Heart rate measurements, pupil dilations, or cortisol levels also provide indications of stress or arousal. Physiological measures could

<sup>2</sup>With satisficing the respondent offers responses that will seem reasonable to the interviewer without any memory search or information integration taking place.

<sup>3</sup>Reaction times can be gathered through a variety of methods: in telephone interviews interviewers can be asked to manually time the pause between question off-set and answer on-set, or it can be done through voice-recognition software in computer-assisted telephone interviews (CATI). In web surveys, reaction times are measured routinely nowadays, but it is much easier to record server-side times than it is to measure the actual time respondents spend on the task; whereas in self-administered surveys on a computer or laptop (CAPI) software (such as E-prime) can be used to measure reading times and answering times rather precisely.

help to distinguish sensitive questions and, through that, predict socially desirable answering behaviour. Of course one would need to have well thought-out experimental designs to be able to interpret these measures. If a question on abortion is posed and respondents show strong arousal reactions, it would not be certain at all whether that would be a reflection of the strong feelings towards abortion, or of strong feelings about the interviewer or about something else in the research context.

One could also turn to brain imaging tools such as EEG and functional Magnetic Resonance Imaging (fMRI). EEG measures the electrical activity on the surface of the cerebral cortex. The time resolution is very good: it is possible to pinpoint the onset of different processing stages as reflected in characteristic wave patterns with millisecond precision. It is much harder to localize the source of the recorded waves, typically leaving an uncertainty in location of several centimetres. fMRI has the opposite characteristics. Where EEG is particularly suitable to track cognitive processes with millisecond precision, fMRI is suitable to chart the structure of the neural groups and systems activated in the brain. fMRI measures the abundance of oxygenated blood in the brain and because of the rather slow blood response, only processes that leave a neural trace at a scale of several seconds can, therefore, be measured. A great advantage of this technique, however, is that it is possible to localize neural responses with millimetre precision.

With imaging tools, it is possible to measure in the brain where and when neural processing takes place. Our knowledge of the relationship between neural and cognitive processing has increased dramatically in the past 10 years. For some types of activities, it is now known where in the brain they are located, so if precise hypotheses are formulated about why a question on 'abortion' should cause a strong emotion or other cognitive response, this could be checked by using these methods. Examples of successful applications of this approach can be found in the budding field of 'social cognitive neuroscience' (e.g. Lieberman, Schreiber, & Ochsner, 2003; Ochsner & Lieberman, 2001).

As mentioned earlier, EEG measurements allow observation of cognitive processes as they unfold over time. It is known that certain cognitive processes correspond to characteristic EEG shapes. For example, when a semantically unexpected word is encountered in a sentence, there is an increased negativity in the EEG signal. This negative wave form is observed about 400 milliseconds after the onset of the unexpected word and is therefore called the N400. In the sentence 'When I drive my car, I like to listen to ice cream on the radio', 'ice cream' will cause an N400 response (Kutas & Hillyard, 1980; see also Van Berkum, 2004). In a similar vein, Cacioppo, Crites, Berntson, and Coles (1993) have shown that if subjects encounter an attitude object that is unexpected, that is, different in evaluation compared to other elements in a list, an EEG shape called P300 will occur. So, if a list of people is offered, such as 'Osama, Hitler, Gandhi, Mao', subjects will show a P300 on 'Gandhi' as soon as they encounter that word.

Such research opens paths for disentangling the many processes involved in answering survey questions. For example, one wonders whether at a neural level a Democrat integrates words differently from a Republican in politically loaded sentences such as 'In general, congress should increase spending'. Does the Democrat exhibit an N400 response here? (Morris, Squires, Taber, & Lodge, 2003). And if so, can we still make a case for a separate third stage, 'Deriving a Judgment', in the model by Tourangeau et al. (2000)? Recently, research suggests that respondents who hold favourable attitudes towards an issue, show an EEG response if the evaluative term in the question is negative about the issue ('I think abortion is bad'), whereas a control group of respondents who hold negative attitudes towards the issue, show an EEG response with the opposite question ('I think

abortion is good') (Van Berkum et al., 2006; Van Berkum, Holleman, Murre, Nieuwland, & Otter, 2007).

Another line of research is to use fMRI to gain insight into the neural representations of political concepts or other topics relevant to survey research. Though this may seem far-fetched, work by Haxby et al. (2001) shows that high-level visual categories, like cats, scissors, chairs, houses, etc. can be distinguished reliably on the basis of fMRI-recorded brain activations. This research suggests that these categories are represented in a distributed pattern throughout brain areas involved in semantic processing (mainly in the temporal cortex). Analyses by Hanson, Matsuka, James, and Haxby (2004) show that a category hierarchy can be reproduced on the basis of these fMRI representations. It would be interesting to measure, for example, the neural political landscape in the same manner, comparing subjects with opposing views. Can we predict from the fMRI pattern what political category is being presented (e.g. as a sentence or short film fragment)? Gun control, abortion, poverty? Can we construct a neural political landscape, like the study by Hanson et al. did for a selection of visual categories? Do such landscapes differ in systematic ways between Democrats and Republicans?

Many more possibilities for research can be adduced. For example, one could try to gain more insight into the nature of deliberative processing and satisficing, where questions are answered with the least possible mental effort (Krosnick, 1991). In fMRI research, active processing activates certain characteristic areas of the brain, notably in the frontal cortex. Can we characterize satisficing and deliberative processing in neural terms using fMRI-based imaging?

In principle, every method from the toolbox discussed here can be used in combination with each of the cognitive models discussed in the previous section. But every method from the toolbox discussed here has its own strong and weak points. Headway can only be made if one and the same phenomenon is studied and explained with various tools. For example, difficulties in the response mapping stage (the stage at which a judgment is translated into an answer) may be detectable in a respondent's verbal protocols, or in interactions between the respondent and the interviewer. But this is not necessarily so. Hence, it is interesting to also use experimental designs in which mapping difficulty is manipulated, to see whether reaction time data show consistent patterns as to which question formats are more difficult to answer than others. By using various tools to study one and the same phenomenon, one is also forced to define central concepts in a way that is less specific for the method used. For example, processing speed in survey reaction time studies may be defined as the time passed between the end of the question and the offset of an answer, whereas in eye-tracking studies it may be used to denote the first fixation durations on a word. And both may be used as an indicator of processing difficulty.

This being the case, it is important for survey researchers looking for cognitive explanations of their response effects or instability in their data, and researchers who want to describe processes underlying survey questions in cognitive terms, to be familiar with a variety of cognitive tools and to be able to use them in order to triangulate findings on the phenomena they want to study.

## TOWARDS AN INTEGRATED MODEL OF QUESTION ANSWERING

How do respondents' survey answers reflect the attitude a researcher aimed to measure? What interactional, emotional and even neurobiological processes underlie such an

answer? In order to build a coherent theory of question answering that reflects all aspects of such a process and that models the relations between different (sub)systems, an approach of 'hierarchical modelling' is needed.

Ideally, models are developed at different levels of description, for example: (1) an aggregate, behavioural level, (2) a cognitive level and (3) a neural systems. Level 1 describes surveys in a high-level, abstract fashion, as for example in structural equation models. Level 2 aims to target, for example, how processes of survey answering unfold in subjects, which can be described with box-and-arrow models, or rule-based models. Level 3 also addresses the neurobiology, but in a high-level manner. Such models may have a module that represents some of the functionality of the frontal cortex, hippocampus (memory centre), or temporal cortex (where meaning is stored). This could be done in a connectionist model. A fourth level could be added to target neurophysiological data, which might have a role for certain neurotransmitters and neuromodulators. Other (intermediate) levels may also be distinguished, such as the social level (Lieberman et al., 2003) or a genetic level (in a study of hereditary patterns).

The different levels are typically studied using different methods, have different criteria for validity, and must be validated with different types of data. They also tend to use different names for the same concept. The 'hierarchical modelling' approach proposes to reformulate models such that it becomes transparent what a given concept means at different levels. Thus, the concept 'processing speed' may appear in level 1 survey models as a parameter that somehow distinguishes groups of young and old subjects (old subjects having slower processing). At a cognitive level, it may be decomposed into various sub processes with different processing speeds for specific stages and processes. At a neural systems levels, different processing speeds may emerge as a result of more or less noise in the neural processing and because of variations in the quality of the connections. In models of older subjects, for example, there may be more noise and fewer and weaker connections. Such models would allow detailed predictions about the relationship between certain task conditions and the state of the neural architectures. Thus, we see that the blanket term 'processing speed' gains more and more detail as we descend the levels of description. 'Zooming out', however, we are left with a single parameter that roughly characterizes the total effect of low-level interactions.

Several problems can be identified that hamper this process of searching for connecting theories. First, one major problem in connecting surveys to cognitive models is not only that we typically have to use different formalisms to express modelling concepts, but also that the different levels refer to different types of data. We are thus not only faced with several levels of theorizing, but also with divergent data sources. For example, survey research tends to operate at a high level of abstraction, whereas cognitive theories refer to the details of cognitive processes as they unfold over time. Thus, a survey theory might incorporate a 'strength' parameter for attitude in certain groups, whereas a cognitive theory would focus on how attitude strength increases and declines over time as a function of certain stimulus and task conditions. At a neural systems level, we might even try to model the involvement of emotion areas, such as the amygdala, or areas involved in motivation. A second problem is that in order to give general advice on aspects of questionnaire design, one needs general insights into the mechanisms of question answering and the way in which specific aspects of the communication situation interact to influence an answer. Experiments on the process of question answering, however, tend to focus on the effect of one specific factor in the communicative situation and not on the complete picture of the answering process. A third issue to be resolved is: to what extent can concepts that are

meaningful to survey researchers, such as ‘validity’, ‘reliability’ or ‘error’, be related to cognitive concepts or explanations in a way that is useful for an integrative approach?

Even the seemingly simplest phenomenon requires a lot of work if you really want to understand what is going on during question answering *and* if you want to be able to come to some sort of general advice. For example, in survey research a well-known problem is that seemingly equivalent questions that are worded slightly differently, often lead to strongly divergent answers. If you ask ‘Do you think the government should forbid abortion?’ fewer respondents will be likely to endorse this question than the number of respondents who will answer negatively when you ask ‘Do you think the government should allow abortion?’ Logically, if you say ‘yes’ to forbid, you should say ‘no’ to allow—but in a survey context many more respondents say ‘no, not forbid’ than ‘yes allow’ (e.g. Hippler & Schwarz, 1986; Holleman, 2000; Reuband, 2000; Rugg, 1941; Schuman & Presser, 1981; Waterplas, Billiet, & Loosveldt, 1988).

To decide which of both questions, or rather answers to these questions, is a valid reflection of respondents’ true attitudes towards abortion, one would like to know the cognitive processes that *cause* the difference in the answers. The predominant explanation for the asymmetry offered in the literature, was that although the verbs are opposites, the connotations of the verbs are rather extreme: to forbid seems too harsh, whereas to allow may sound as if one wants to encourage a deviant behaviour (Schuman & Presser, 1981)—which would cause respondents to rather say ‘no’ to both verbs.

A first problem, of course, is that the forbid/allow asymmetry is observed on an aggregated, behavioural level, whereas to explain it, one wants to focus on the individual answering behaviour of respondents and the cognitive principles explaining that behaviour. The ‘connotations explanation’ leaves implicit whether it is meant to describe the cognitive mechanisms behind each individual’s behaviour, or whether it intends to give an aggregated explanation after the fact. It is suggested here that the connotations of words are context independent: they should be there all the time, for all language users. Consequently, is every individual assumed to avoid extremeness? In that case, the forbid/allow asymmetry should entail that *nobody* answers affirmatively to forbid questions or allow questions. If only *some* respondents are assumed to be likely to avoid answering extremely, the subgroups of respondents this holds for should be specified. Next, the hypothesis should be tested on an individual level, for example, by within-respondents measures (Holleman, 2006).

What is more, the semantic connotations explanation should be related to more general models of question answering, such as Tourangeau et al.’s box-and-arrow model. This can be done, for example, by using a simplified model in which two stages are distinguished: one cognitive stage of question processing (including activation, formation or retrieval of a judgment), and one more communicative stage of translating that judgment into a response. The difference in forbid/allow answers can reflect that different interpretations of the question text were construed due to other connotations being activated, or can reflect that different attitudes are retrieved or formed (towards the attitude object as well as to ‘forbidding things’ in one question version, and towards the attitude object as well as to ‘allowing things’ in the other). In that case, forbid/allow questions do not measure the same attitudes: it was wrong to assume that the percentage of ‘no’-answers to the one question should be similar to the percentage of ‘yes’-answers to the other). It could also be the case, however that forbid/allow questions do measure similar attitudes, but that those attitudes are expressed differently on the answering scale due to the use of ‘forbid’ or ‘allow’ in the question text. In that case, a given evaluation of the attitude object can be reflected by a ‘no,

not forbid' as well as by a 'yes, allow'. Analyses through a structural modelling approach<sup>4</sup> show that the first possibility can be rejected: differences in forbid/allow answers arise when similar attitudes are translated to the answering scale, due to the use of forbid/allow (Holleman, 1999). Once that was established, further experiments could be carried out, in order to find out how the meanings of answering options shift due to the use of forbid/allow, within-respondents, and how the meanings of yes and no to forbid/allow questions are dependent not only on forbid/allow, but also on the issue at hand and of respondent characteristics (Holleman, 2006).

What we mean to show with this example, is that first the phenomenon to be explained has to be modelled cognitively and has to be mapped onto the general process of question answering. Only then does it become possible to formulate precise hypotheses about the effects of a certain survey, interviewer or respondent characteristic and to test these on the proper level of aggregation. Based on these results, it is possible to start thinking about generalizations of these findings to other phenomena and about the practical implications of these findings.

So one has to zoom in on a survey phenomenon, such as the use and the effects of forbid vs. allow, and on the cognitive mechanisms explaining those in some detail before any results are obtained that can be related to other survey phenomena, or to survey practice, which is an aggregate level. The forbid/allow research sketched above was extended by measuring reaction times to track processes of question answering in order to triangulate the results obtained through structural modelling. A mathematical model describing the cognitive processes at hand was developed to test the results (Chessa & Holleman, 2007). Currently, research is being conducted in which the answers to contrastive questions are related to processes at the neural level: do evaluative contrastive terms that are often used in surveys (such as good/bad, forbid/allow, positive/negative) instantly activate a measurable brain response during interpretation of the question, as shown by an EEG (N400 or P300) response? To what extent can the types of stages in the process of question answering as distinguished by Tourangeau et al. be discriminated at a neural level? At the moment these latter issues are too fundamental to be of any use for survey practice right away, but in the longer run the neural level will help us to understand what attitudes and opinions are.

Connecting levels in a vertical manner is crucial for the development of the integrated viewpoint proposed here. Such connections can be set-up through multi-level theories centred around specific concepts. Examples abound in the natural sciences. The concept of 'temperature', for example, used to be a high-level measure of heat and cold, but gained new meaning at the lower, molecular level with the development of statistical mechanics. There are now so many connections across and between levels in physics that many theorists aim to develop so called Grand Unified Theories. Both cognitive science and survey research are far removed from such lofty goals and it is unclear whether such an endeavour is even feasible for cognitive science. Nonetheless, much is to be gained even by trying, and the more we achieve similar connections between levels, the more we will be able to use data and insights from one level to help guide and constrain theory and practice on another.

<sup>4</sup>A structural modeling approach here refers to a confirmatory factor analysis which distinguishes between true scores, reflecting attitudes and error scores, reflecting systematic variance due to the manipulation of the question wording, as well as variance due to specific question content (see Andrews, 1984 for an example).

## DISCUSSION

In this paper, we have sketched some directions for future research in cognitive survey research that we feel will help to gain a more integrated view of the cognitive processes involved in answering survey questions, and the relation between respondents' answers and their 'true' attitudes or behaviour. The essence of an integrated model is that the central concepts in such a model obtain meaning in more than just one research context. That is, the meaning of 'attitude strength' in a social theory would refer to differences between groups, and is different from the concept of attitude strength in a neurobiological model which might be implemented in terms of certain brain areas involved in motivation or emotion (e.g. the amygdala). In an integrated model those two concepts would be related to each other, so that differences in activation of specific simulated brain regions (groups of nerve cells) can be related to characteristic differences between specific social groups. Instead of aiming for explanations of rather 'local' survey problems, research should aim to build related models that link the processes at a social level in a specific context to the processes at the cognitive level and, if possible, the processes at the level of the brain.

In integrated theories, models at each level pose constraints on the other levels, top-down as well as bottom-up (Lieberman et al., 2003). The more detailed and formal the integrated model is, the clearer the constraints on the other levels will be: a box-and-arrow model typically poses fewer and perhaps less clear-cut constraints on the other levels than a connections model, a rule-based model, or a mathematical model.

We have not yet reached the stage of integrated theories, as in the natural sciences. Social and cognitive sciences are of course much younger research fields. They also differ from the natural sciences in many ways, with social and behavioural phenomena being less well-defined and more difficult to isolate. Nonetheless, *striving* for integrated theories will help to improve the advancement of cognitive theory as well as survey practice.

The process of developing more integrated theories in the CASM field, the study of CASM, will help the cognitive psychologists involved to test their theories and develop them further in cognitive and communicative tasks that take them out of lab. Stepping out of the lab and practicing science in the real-world is a scary prospect for many cognitive psychologists. Survey research is a nice compromise between the highly controlled nature of laboratory experiments and the slings and arrows of the real-world. Survey research is a controlled, but natural language use situation. Cognitive aspects of survey research, thus, provide cognitive researchers with excellent opportunities to obtain insight into cognitive processing or cognitive characteristics by manipulating input variation (i.e. question wording, question order, relations between text and visual context) and measuring output characteristics (such as answers to questions, reaction times, verbal interactions, etc.).

A quest for integrated theories in the CASM field will help CASM researchers to come to a better and more complete understanding of the cognitive mechanisms at hand during question answering and of the interplay between these mechanisms. It will help to formulate better definitions of central concepts and processes, to achieve a better understanding of the relations between phenomena to be explained, and hence it will ultimately lead to better advice for survey practice.

Different types of models typically require different types of data to develop and test them. This is not a problem because the fullest understanding of a phenomenon or concept is obtained best by approaching it from different angles and by measuring it with different methods, as each method reveals specific aspects of the phenomenon of interest. Furthermore, every measurement tool has its flaws and constraints. Therefore in this paper,

we suggest using multiple methods to triangulate in order to remedy some of the limitations of each separate method.

We sketched a broad range of cognitive measures that can be used to study a large variety of specific survey phenomena; classic methods or measures, such as reaction times, and more recent neurocognitive methods and tools. For example, a survey issue such as strong satisficing can be examined with reaction times (very short RT's indicating the respondent just looked for a quick plausible answer, but not for the best answer), but also with neurocognitive methods such as fMRI (which would show that the frontal cortex is not active while satisficing). To examine which part of a certain question is processed most thoroughly, eye-tracking can be used, as a longer gaze is associated with deeper processing in this method. Galesic et al. (2006) show, for example, that contrast effects between question content and pictures arise when respondents actually processed the pictures, whereas assimilation effects occurred if they did not. Eye-tracking could also investigate whether concepts that are congruent with respondents' attitudes are processed more thoroughly than concepts that are incongruent. This would show that the stage of question interpretation and attitude retrieval are more intertwined than the model of Tourangeau et al. suggests. EEG techniques can also be used to explore this idea. To determine which questions are about sensitive or emotional issues for certain respondents, psychophysical measures such as galvanic skin responses can be of use, whereas interaction coding will provide insight into the verbal cues (intonation patterns) and non-verbal cues (such as facial expressions) indicating that respondents find an issue very sensitive.

An overview of cognitive models and methods and their applications in survey research cannot be complete, and we did not aim to be. We do hope to have inspired CASM researchers and cognitive psychologists to use some of the recent methods of cognitive science, to triangulate findings, to work towards integrative models and to continue studying phenomena that are relevant in survey research.

A prerequisite for the types of research sketched here is interdisciplinary cooperation. Teamwork is needed to formulate new research questions that are relevant to all fields included and to get hold of the theories that are both relevant to that problem as well as up to date. Teamwork is called for to formulate hypotheses at different theoretic levels, to link the various models and to design the proper ways to test these models and hypotheses with a range of tools. As members of an interdisciplinary research team, speaking from experience, we can say that quite some time is needed to get acquainted with each other's fields: psychologists will have to learn about the practical problems that plague survey research, cognitive psychologists will have to learn about neuroscience methods and its accompanying neuroanatomical vocabulary, neuroscientists will have to learn about attitudes, survey measurement and so on. But it is well worth the effort. We hope that in the coming years many other cognitive psychologists, neuroscientists, social psychologists, linguists, political scientists and survey methodologists will attempt to collaborate and will help to complete the bridge.

## ACKNOWLEDGEMENTS

We started this work on the relation between neurobiological processes and survey research during our stay in 2003–2004 at the Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS) in Wassenaar, The Netherlands. We thank Willem Saris for bringing together the interdisciplinary group 'A new view on survey research' at

NIAS, and we thank NIAS for the opportunity to cooperate on this research. Last but not least, we thank Bob Belli and Fred Conrad, editors for this paper, and two anonymous reviewers for their elaborate and constructive comments on an earlier version of this manuscript; and Milton Lodge and Daphne van Weijen for their useful feedback. Of course all remaining errors are our own.

## REFERENCES

- Abeles, M. (1991). *Corticonics*. Cambridge: Cambridge University Press.
- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R., & Lebiere, C. (1998). *Atomic components of thought*. Hillsdale, NJ: Erlbaum.
- Andrews, F. M. (1984). Construct validity and error components of survey measures: A structural modeling approach. *Public Opinion Quarterly*, 48, 409–442.
- Ayidiya, S. A., & McClendon, M. J. (1990). Response effects in mail surveys. *Public Opinion Quarterly*, 54, 229–247.
- Bassili, J. N. (1996). The how and why of response latency measurement in telephone surveys. In N. Schwarz, & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 319–346). San Francisco: Jossey-Bass.
- Belli, R. F. (2005). Editorial. Announcing a special issue on the cognitive aspects of survey methodology. *Applied Cognitive Psychology*, 19, 245–247.
- Cacioppo, J. T., Crites, S. L., Berntson, G. G., & Coles, M. G. H. (1993). If attitudes affect how stimuli are processed, should they not affect the event-related brain potential? *Psychological Science*, 4, 108–112.
- Callegaro, M. (2005). Origins and development of the cognitive models of answering questions in survey research. *First Annual Meeting of the European Association for Survey Research (EASR)*, Barcelona July 18–22.
- Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on interviewing techniques. *Sociological Methodology*, 12, 389–437.
- Chessa, A. G., & Holleman, B. C. (2007). Answering attitudinal questions: Modeling the response process underlying contrastive questions. *Applied Cognitive Psychology*, 21, 203–225.
- Chessa, A. G., & Murre, J. M. J. (2002). A model of learning and forgetting I: The forgetting curve. Amsterdam: *NeuroMod Technical Report 02-01*.
- Conrad, F., & Blair, J. (2004). Data quality in cognitive interviews. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, et al. (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 67–89). New York: Wiley Publishers.
- Converse, P. E. (1964). The nature of belief systems. In D. E. Apter (Ed.), *Ideology and discontent* (pp. 206–261) New York: Free Press.
- Dijkstra, W. (1999). A new method for studying verbal interactions in survey interviews. *Journal of Official Statistics*, 15, 67–85.
- Draisma, A. R., & Dijkstra, W. (2004). Response latency and (para) linguistic expressions as indicators of response error. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, et al. (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 131–149). New York: Wiley Publishers.
- Ehlen, P., Schober, M. F., & Conrad, F. G. (2007). Modeling speech disfluency to predict conceptual misalignment in speech survey interfaces. *Discourse Processes*.
- Fazio, R. H. (1990). A practical guide to the use of response latency in social psychological research. In C. Hendrik, & M. S. Clark (Eds.), *Research methods in personality and social psychology* (pp. 74–97). Newbury Park, CA: Sage.
- Fowler, F. J., & Cannell, C. F. (1996). Using behavioral coding to identify cognitive problems with survey questions. In S. Schwarz, & S. Schuman (Eds.), *Answering questions* (pp. 15–36). San Francisco: Jossey-Bass.
- Galesic, M., Tourangeau, R., Yan, T., & Ye, C. (2006). Use of eye-tracking for studying survey response processes. *61st AAPOR Conference*, Montreal, May 18–21.

- Graesser, A. C., Cai, Z., Louwerse, M. M., & Daniel, F. (2006). Question Understanding Aid (QUAID). A web facility that tests question comprehensibility. *Public Opinion Quarterly*, 70, 3–22.
- Hanson, S. J., Matsuka, T., James, V., & Haxby, J. V. (2004). Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: Is there a “face” area? *NeuroImage*, 23, 156–166.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293, 2425–2430.
- Hippler, H.-J., & Schwarz, N. (1986). Not forbidding isn't allowing: The cognitive basis of the forbid/allow asymmetry. *Public Opinion Quarterly*, 50, 87–96.
- Hippler, H.-J., & Schwarz, N. (1987). Response effects in surveys. In H.-J. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 102–122). New York: Springer Verlag.
- Hippler, H. J., Schwarz, N., & Sudman, S. (1987). *Social information processing and survey methodology*. New York: Springer Verlag.
- Holbrook, A. L., Cho, Y. I., & Johnson, T. (2006). The impact of question and respondent characteristics on comprehension and mapping difficulties. *Public Opinion Quarterly*, 70, 565–595.
- Holleman, B. C. (1999). The Nature of the forbid/allow asymmetry. Two correlational studies. *Sociological Methods & Research*, 28, 209–244.
- Holleman, B. C. (2000). *The forbid/allow asymmetry. On the cognitive mechanisms underlying wording effects in surveys*. Amsterdam/Atlanta: Rodopi Publishers.
- Holleman, B. C. (2006). The meanings of ‘Yes’ and ‘No’. An explanation for the Forbid/Allow asymmetry. *Quality & Quantity*, 40, 1–38.
- Jabine, T. B., Straf, M. L., Tanur, J. M., & Tourangeau, R. (Eds.). (1984). *Cognitive aspects of survey methodology: Building a bridge between disciplines*. Washington DC: National Academy Press.
- Jobe, J. B., & Loftus, E. F. (Eds.). (1991). Cognition and survey measurement: History and overview [Special Issue]. *Applied Cognitive Psychology*, 5, 175–192.
- Kim, S.-Y., Lodge, M., & Taber, C. (2004). A computational model of political preferences. The dynamics of political candidate evaluation. In *Proceedings of the sixth international conference on cognitive modelling* (pp. 362–363). Pittsburgh, PA: Carnegie Mellon University/University of Pittsburgh.
- Kim, S.-Y., Lodge, M., & Taber, C. (2007). A model of survey response: Perceptions about candidate issue positions. Paper presented at the Annual meeting of the American Political Science Association 2007, Chicago, Aug 29–Sep.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213–236.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203–205.
- Lau, R. R., & Redlawsk, D. P. (2001). Advantages and disadvantages of cognitive heuristics in political decision making. *American Journal of Political Science*, 45, 951–971.
- Lieberman, M. D., Schreiber, D., & Ochsner, K. N. (2003). Is political cognition like riding a bicycle? How cognitive neuroscience can inform research on political thinking. *Political Psychology*, 24, 681–704.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception. Part 1: An account of basic findings. *Psychological Review*, 88, 375–405.
- Molenaar, N. J. (1982). Response-effects of “formal” characteristics of questions. In W. Dijkstra, & J. van der Zouwen (Eds.), *Response behaviour in the survey-interview* (pp. 49–89). London: Academic Press.
- Molenaar, N. J. (1991). Recent methodological studies on survey questioning. *Quality and Quantity*, 25, 167–187.
- Morris, J. P., Squires, N. K., Taber, C. S., & Lodge, M. (2003). Activation of political attitudes: A psychophysiological examination of the hot cognition hypothesis. *Political Psychology*, 24, 727–745.
- Mulligan, K., Grant, J. T., Mockabee, S. T., & Monson, J. Q. (2003). Response latency methodology for survey research: Measurement and modeling strategies. *Political Analysis*, 11, 289–301.

- Ochsner, K. N., & Lieberman, M. D. (2001). The emergence of social cognitive neuroscience. *American Psychologist*, *56*, 717–734.
- Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., et al. (2004). Methods for testing and evaluating survey questionnaires. *Public Opinion Quarterly*, *68*, 109–130.
- Redlawsk, D. P. (2002). Hot cognition or cool consideration? Testing the effects of motivated reasoning on political decision making. *The Journal of Politics*, *64*, 1021–1044.
- Reuband, K.-H. (2000). “Erlauben” vs. “Nicht erlauben” oder “verbieten”? Wie sich unterschiedliche Frage-Alternativen auf das Antwortverhalten auswirken. *ZA-Information*, *48*, ZUMA Mannheim.
- Rugg, D. (1941). Experiments in wording questions II. *Public Opinion Quarterly*, *5*, 91–92.
- Schuman, H., & Presser, S. (1981/1996). *Questions and answers in attitude surveys. Experiments on question form, wording, and context*. Thousand Oaks, CA: Sage Publications (Originally published in New York: Academic Press).
- Schwarz, N. & Sudman, S. (Eds.) (1995). *Answering questions*. San Francisco: Jossey-Bass.
- Schwarz, N., Knauper, B., Hippler, H.-J., Noelle-Neumann, E., & Clark, L. (1991). Rating scales: Numeric values may change the meaning of scale labels. *Public Opinion Quarterly*, *55*, 570–582.
- Schwarz, N., & Bohner, G. (2001). The construction of attitudes. In A. Tesser, & N. Schwarz (Eds.), *Blackwell handbook of social psychology: Intraindividual processes* (Vol. 1, pp. 436–457) Oxford: Blackwell.
- Smith, E. R. (1999). New connectionist models of mental representation: Implications for survey research. In M. G. Sirken, D. J. Herrmann, S. Schechter, N. Schwarz, J. M. Tanur, & R. Tourangeau (Eds.), *Cognition and survey research* (pp. 251–264). New York: Wiley & Sons.
- Sniderman, P. M., & Theriault, S. M. (2004). The structure of political argument and the logic of issue framing. In W. E. Saris, & P. M. Sniderman (Eds.), *Studies in public opinion, attitudes, nonattitudes, measurement error, and change* (pp. 133–165). New Jersey: Princeton University Press.
- Strack, F., & Martin, L. L. (1987). Thinking, judging and communicating. A process account of context effects in attitude surveys. In H. J. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 123–148). New York: Springer Verlag.
- Sudman, S., & Bradburn, N. M. (1974). *Response effects in surveys: A review and synthesis*. Chicago: Aldine.
- Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers. The application of cognitive processes to survey methodology*. San Francisco: Jossey Bass Publishers.
- Tanur, J. M. (1992). *Questions about questions: Inquiries into the cognitive basis of surveys*. New York: Russell Sage.
- Tourangeau, R., & Rasinski, K. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, *103*, 299–314.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.
- Tourangeau, R., Couper, M. P., & Conrad, F. (2004). Spacing, position, and order. Interpretative heuristics for visual features of survey questions. *Public Opinion Quarterly*, *68*, 368–393.
- Van Berkum, J. (2004). Sentence comprehension in a wider discourse: Can we use ERPs to keep track of things? In M. Carreiras, & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eye tracking, ERP and beyond* (pp. 229–270). London/NY: Psychology Press.
- Van Berkum, J., Holleman, B. C., Murre, J., & Nieuwland, M. (2006). Studying question answering with brain potentials. *Paper presented at the Discourse, Communication & Cognition Conference (DCC)*, Utrecht University, Nov 9 2006.
- Van Berkum, J., Holleman, B. C., Murre, J., Nieuwland, M., & Otter, M. (2007). When do people disagree with an opinion statement? An ERP-study on opinion poll comprehension. *Poster presentation at the Annual Meeting of Cognitive Neuroscience (CNS)*, May 4–8 2007, New York.
- Van Overwalle, F., & Siebler, F. (2005). A connectionist model of attitude formation and change. *Personality and Social Psychology Review*, *8*, 28–61.
- Van der Veld, W., & Saris, W. E. (2004). Separation of error, method effects, instability and attitude strength. In W. E. Saris, & P. M. Sniderman (Eds.), *Studies in public opinion. attitudes, nonattitudes, measurement error, and change* (pp. 37–59). New Jersey: Princeton University Press.
- Van der Zouwen, J., Dijkstra, W., & Smit, J. H. (1991). Studying respondent-interviewer interaction: The relationship between interviewing style, interviewer behavior, and response behavior. In P. P.

- Biemer, R. M. Groves, L. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 419–438). New York: John Wiley & Sons.
- Van der Zouwen, J., & Smit, J. H. (2004). Evaluating survey questions by analyzing patterns of behavior codes and question-answer sequences. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, et al. (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 109–131). New York: Wiley Publishers.
- Waterplas, L., Billiet, J., & Loosveldt, G. (1988). De verbieden versus niet toelaten asymmetrie. Een stabiel formuleringseffect in survey-onderzoek? [The forbid/allow asymmetry. A stable wording effect in survey research?] In: *Mens en Maatschappij*, 63, 399–415.
- Wegener, D. T., Downing, J., Krosnick, J. A., & Petty, R. E. (1995). Measures and manipulations of strength-related properties of attitudes: Current practice and future directions. In R. E. Petty, & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 455–489). Hillsdale, NJ: Erlbaum.
- Willis, G. B. (2004). Cognitive interviewing revisited: A useful technique, in theory? In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, et al. (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 23–43). New York: Wiley Publishers.
- Willis, G. B. (2005). *Cognitive interviewing. A tool for improving questionnaire design*. Thousand Oaks: Sage Publications.
- Yan, T., & Tourangeau, R. (2007). Fast times and easy questions: the effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology*, (in press). DOI: 10.1002/acp1331