Paper prepared for the first meeting of the European Association for Survey Research, Barcelona, 2005, July 18-21.
Please do not quote this version without permission of the authors

[1]A Unified Model for the Survey Response Process

Estimating the stability and crystallization of public opinion

William M. van der Veld[2] & Willem E. Saris[3]

Abstract

In the past, survey researchers assumed that the respondents held opinions which could be obtained by asking the proper questions. This idea - the "file drawing model" for survey response - was criticized strongly by Converse in 1964, but also strongly defended by Achen (1975), Erikson (1978, 1979), and Judd et al. (1981). Zaller (1992) introduced the idea that people create their response on the spot when they are asked a question on the basis of the most salient considerations. The saliency of the different considerations can vary with the context in which the question is asked and the formulation of the question itself. This approach has gained a lot of attention in the last 20 years (Tourangeau et al., 2000; Tourangeau and Rasinski, 1988; Wilson and Hodges, 1991). In this study a model has been developed that unifies these ideas into one model (van der Veld and Saris, 2004). For this purpose a theory is needed that describes how a question leads to a response. This theory should account for the fact that the response could be different from the opinion, due to the (bad) quality of the measurement procedure. Furthermore, the model should allow that an opinion is made up of pre-existing or stable considerations and temporary or unique considerations. A structural equations model that follows from the mathematical description of the survey response for an individual has been developed. With this structural equations model it is possible to decompose the variance of responses into three different components that also can be linked to commonly used concepts in survey research. The variance of the responses will be decomposed in variance due to pre-existing considerations, temporary considerations, and measurement error. These three variance components can be linked to opinion stability, opinion crystallization, and measurement quality respectively.

1 *Data analyzed in this paper comes from the Russian Socio-Economic Transition Panel (1992-1999) that was funded by the Dutch Organization for Scientific Research (NWO). We acknowledge support from the Netherlands Institute of Advanced Studies (NIAS) where we resided during the academic year of 2003-2004, and worked on this paper. In addition we acknowledge the support of the Amsterdam School of Communication Research (ASCoR).*
2 *William van der Veld is assistant professor in Methodology of the Social Sciences at department of Communication Science of the University of Amsterdam.*
3 *Willem E. Saris is Full Professor in Methodology of the Social Sciences at the University of Amsterdam and professor at the ESADE business school of Ramon Llull University in Barcelona.*

Historical Background

The quality of survey data has been criticized ever since survey research. This criticism became even more acute when Philip Converse (1964) found that most people did not provide responses to survey items from an underlying stable and consistent belief system. He found that most people seem not very confident about the responses they provide when the question concerns political issues. One is therefore inclined to question the value of the response, i.e. whether the response can be trusted as an indication of an opinion. And consequentially whether that opinion has any substantive value. This suspicion was supported by the fact that responses on the same question across time were very unstable, indicating random changes of opinions. And also by the fact that scores on logically related variables were inconsistent. This was found to be true for the larger part of the population. On the other hand, a small part of the population did have a so-called stable and consistent belief system. Because of this discrepancy Philip Converse named this model "black and white" model.

The conclusion that measured opinions are essentially worthless is an unwelcome conclusion because normative theories in the field of political science require that the public have some political knowledge, at least enough to cast a deliberate vote. Therefore not everybody (Achen, 1975; Erikson, 1978, 1979; Judd et al., 1980; Feldman, 1989) could accept the conclusion that most people have no opinion because they give very inconsistent responses. In 1975 Christopher Achen came with an alternative explanation for the response inconsistency. He claimed that survey items are vague and therefore respondents have problems selecting the proper response alternative. This leads to mistakes, and a response could thus contain measurement error. If that measurement error is separated from the response, then the opinion remains stable. In order to separate measurement error from the response he used a quasi-simplex model. Analyses of the data for the same political issues showed that opinions were very stable after correction for measurement error, even when responses were very inconsistent as Philip Converse (1964) found.

Although both approaches give different explanations for response inconsistency, they agree on the nature of the opinion. That is, in both approaches it is assumed that opinions rest in memory (or not when the opinion does not exist), waiting to be expressed by means of a survey item. However, having an opinion stored in memory leads to puzzles when the implications are traced out. The assumption of *having an opinion* implies that for all questions imaginable an opinion is stored in a person's brain. And even though our storage capacity may be adequate, it is very unlikely that this is so. Moreover, how could it be possible to have an opinion stored about something that is still to come (Boynton, 1995), or that will never happen or that does not exist? This idea undermines the bases of the old response models where it is assumed that the opinion is stored in memory. In addition, measurement error as an explanation for response inconsistency was not accepted universally. For example John Zaller and Stanley Feldman (1992) called measurement error just another term for unexplained variance. In his book *'The nature and origins of mass opinion'* John Zaller (1992) claims that:

> *…These findings* [the inconsistencies] *are more than methodological curiosities. They seriously undermine the conventional view that surveys are passive measures of "what the public really believes." More ominously, they raise the fear that a large fraction of what opinion researchers measure is either random or systematic "noise". … (Zaller 1992, p. 35)*

Having established that *having an opinion* stored in memory for all questions is unlikely and that *measurement error* is the admission of the weakness that one is unable to explain what a response consists of, other explanations are necessary. John Zaller and Stanley Feldman (1992) came up with a different model for the survey response. They claim that respondents create opinions *on the spot* each time a survey item is asked. Response inconsistency is explained by possible differences in question wording and changes in context. Thus, when a respondent is probed twice for the same opinion, differences may occur due to the fact that different considerations are triggered with which the opinion will be created. In this approach it is implicitly assumed that the opinion created *on the spot* has an exact match among the possible response alternatives, suggesting that measurement be without measurement error. In other words, there is no difference between the response and the opinion.

All three models for the survey response have their merits. Philip Converse's model stresses that responses to political survey items often lack consistency and stability. Christopher Achen's model stresses that a response could contain measurement error, and due to that the response is not necessarily the same as the opinion. And John Zaller and Stanley Feldman's model stresses that the salient considerations are mainly unique considerations that can have a major impact on the response. In order to develop a comprehensive model for the survey response these important views should be brought together.

Overview

We will first provide a new comprehensive theory for the survey response that emerges out of the existing theories of the survey response. That is a theory unifying the ideas of Converse, Achen, and Zaller and Feldman. We then formulated a structural model to test the theory, and we will show how this model unifies the ideas of Converse, Achen, and Zaller and Feldman. Then we will present the data that we use to test the model. We have compared this data with the NES panel data (1956-1960). With this data we have estimated the parameters of the structural model that govern the survey response process, these parameters are stability and the crystallization of an opinion, and the quality of the survey measurement instruments. In addition we will discuss identification and design issues in order to anticipate possible shortcomings of the structural model. Finally, we will discuss the results of the analyses, as well as the implications of the model and theory.

Theoretical framework

Our starting point is that we assume that words or combinations of words in a survey item are associated with one or more nodes in a semantic network (Quillian, 1966; Smith, 1998). When a survey item is presented some words of the survey item immediately trigger the activation of nodes in that semantic network. While reading or hearing the *question* the activation of nodes spreads in a way that is directed by the question wording and by (strong) associations with already activated nodes. Furthermore, the connection of all activated nodes represents the respondent's opinion. However, it is not clear how that works. We therefore assume that during the spreading of activation evaluations are made, either deliberate or automatic (Wegner and Bargh, 1998) that are combined into a representation of a summary judgment, which we call the opinion. We can see that this is a one-step process where it is not possible to identify different phases in the answering of a survey item such as the interpretation [or comprehension], the retrieval [or information-searching], and an evaluation [or judgment] phase. In the literature on the survey response the distinction between these different phases is often made (Schwarz et al. 1998; Sudman et al., 1996; Tourangeau et al. 2000). Although such a distinction is convenient for the sake of understanding the process, we think - as we pointed out here - that it is implausible to make these distinctions.

The next point is that we assume that a survey item is a combination of a *question* and a measurement method. A *question* is directly related to the object of study; it is the *premier* stimulus probing for an opinion. A measurement method is related to the wording of the survey item and is determined by the choices that we make about the form of the response scale, the place of the question in the questionnaire, the presence of an introduction, the choice of the data collection mode, and all other possible choices that must be made when designing a survey item (Andrews, 1984; Scherpenzeel and Saris, 1997). It is important to note that the *question* and the measurement method are not independent entities; they cannot be separated completely from each other. However, it has been found that responses differ for the same question if the formulation of the question is slightly different (Schuman and Presser, 1981; Andrews, 1984; Scherpenzeel and Saris, 1997; Holleman, 2000), suggesting that different measurement methods cause different measurement errors. But not different opinions, provided that the *questions* remains the same.

In accordance with what we mentioned earlier we assume that the opinion is a combination of evaluations. These evaluations are made concurrent to the spreading of activation, and are - for a specific *question* - always made in one dimension. This might for example be a positive-negative dimension, but it might also be in another dimension such as good-bad. The evaluation function transforms - what we shall call for the sake of simplicity - bits of information into real numbers. Because the information might be in any dimension, and the result of the transformation should be in only one dimension, there are multiple functions necessary transforming a score on different dimensions to one dimension. Below this function T is presented. The d in the subscript of the function represents the specific dimension of the information, that is, the arguments are found in that dimension. Hence the information (I) uses the same superscript. With this function it is possible to transform associations with apples and pears to a general scale for the tastiness of fruit, so to say.

$$T_d: I^d \rightarrow R$$

An object has a position on a dimension I^d which will be denoted by $I^d{}_i$. We make the following restriction that the subjective ordering of the information is the same as the ordering of the numerical representations. In other words, when $I^d{}_1$ has an equal or lower position on the dimension I^d than $I^d{}_2$, then[i]:

$$T_d(I^d{}_1) \leq T_d(I^d{}_2)$$

Above we mentioned that the opinion is the combination of evaluations, but we did not yet specify how the evaluations are combined. One could combine the evaluations in several ways, for example with mean-functions or additive functions (Von Neumann and Morgenstern, 1947; Anderson 1970, 1981, 1996; Münnich, 1998; Luce, 2000). We suggest that the evaluations should be combined by addition of the evaluations. This is not uncommon, in utility-theory the utilities are also found to be additive. We can now write for the opinion:

$$O = T_1(I^1{}_1) + \ldots + T_2(I^2{}_1) + \ldots = \sum T_n(I^n{}_i) \qquad \textit{Equation 1}$$

This theory does not imply that people are consciously adding the evaluations. It could well be that most of the summation is masked but still generates a number representing the automatic evaluations. This generated number could in turn be adjusted with evaluations that consciously come to mind after or even during the masked summation.

In the introduction we mentioned that historically there are two ways of looking at the survey response. The first is that the opinion corresponding to a response is stored in memory and the second is that the opinion is created in memory. We pursue our goal to unify these ideas and assume that the opinion is the combination of two different sets of evaluations. There is information which is evaluated over and over again and generates the same result when a *question* is asked repeatedly; we call this the stable part of the opinion. This stable part corresponds to the notion that opinions are stored in memory, we can think of attitudes, or belief systems, ideology, etc. And second, there is information that is evaluated only once, at least within a certain time frame, and generates a unique result; we call this the unique part of the opinion. This unique part corresponds to the notion that opinions are created at the time the question is asked or *on the spot*. The opinion now consists of a stable and a unique part. However the stable part could also change, i.e. people might learn new information. New information can have an enduring effect and thus we need an extra set of evaluations that is new at some point in time and becomes part of the stable part from that moment on, again within a certain time frame.

The opinion now consists of three parts that we conveniently label S, N, and U that refer respectively to evaluations that are stable, new, or unique. Hence, there is no overlap of the information in the different parts. For example when I_1 belongs to the stable set then it cannot belong to either the unique or new set. Since all evaluated information determines the opinion and the information can be split up in three distinct parts (S, N, and U). Each of these evaluations can be written as equation 1, taken into account a limited set of non-overlapping information. The opinion is determined by the sum of all information (equation 1) which can also be written as the sum of the three evaluations (S, N, and U).

$$O = S + N + U \qquad \textit{Equation 2}$$

We will present an example to clarify how we think this model works. Imagine that we ask you about your satisfaction with your social contacts. Normally, your social contacts are relatives, friends, neighbors, and colleagues. Across the years, you could make new friends, switch jobs and get to know new colleagues, get new relatives by marriage or birth, et cetera. You can also lose friends, colleagues, or relatives due to disputes, decease, moving, or for other reasons, and there are also contacts which prevail only for a very short time, for example during a holiday or during a party etc. Some of these contacts are positively (+) evaluated and others that are negatively (-) evaluated. Some of these relations continue for a long time (S), some are new but enduring (N), and some are new but only for a very short time (U) so they are unique in a sense. When the survey item is posed you will make a total evaluation that is the sum of the evaluations of all relations, hence we can also say that it is the sum of the evaluations of the contacts in the different groups specified above (S, N, and U)[ii].

Now that we have explained how the opinion can be formulated when a survey item is asked we turn to the response. We have already mentioned that we assume that a survey item is a combination of a *question* and a measurement method. This *question* is the premier stimulus probing for the opinion. The combination of the measurement method and the *question*, which we call the survey item, is used to obtain a response associated with the opinion. Note that in this formulation the response might be different from the opinion due to the measurement method used. We are convinced that the opinion (at a specific point in time) exists independently of the measurement method. Let us illustrate this point with an example (Schwarz et al. 1991). In a split-ballot experiment the *question* was asked *how successful have you been in life so far*. The only difference between the two groups was that one group received an 11-point scale ranging from 0-10 and the other group that ranged from -5 to +5. Given that the randomization procedure works, we do not expect any differences between the two groups regarding their averaged perceived success (latent) in life so that the opinion is the same for each group. The average response score (observed) on the other hand could differ between the groups due to the different measurement method, even after a linear transformation of the scores to a common scale. The group that received the measurement method with the negative numbers reported themselves more successful on average. Our explanation for this deviation is that connected with each measurement method there is a specific error. The consequence is that the same opinion could be expressed in different ways when different measurement methods are used, we refer to this as a matching process. In this matching process errors are made that are connected to the measurement method. For more details about matching processes we refer to Saris (1987a, 1987b) for continuous variables and to Anderson (1982) for categorical variables. We propose here to use a formulation (Lord and Novick, 1968) that is as simple as possible[iii].

$$R = O + e \qquad \textit{Equation 3}$$

In equation 3, O is the opinion and R is the response. The difference between the opinion and the response is the measurement error e. These measurement errors are made in the matching process connected to a specific measurement method.

This completes the theoretical framework, where we have explained (1) how the opinion is the result of the combination of evaluations (integration), and (2) how the response is the result of the matching of an opinion with a response option (matching). The whole theory is summarized in equation 4, where equation 2 is substituted in equation 3.

$$R = S + N + U + e \qquad \textit{Equation 4}$$

<u>Model specification and design</u>

The theoretical framework (equation 4) is presented as a model for a single individual. In addition to that, we would like this model to apply to the whole population, so that the contribution of the different elements (S, N, and U) to the response is made explicit.

In order to study the process over time we must introduce an index for the time. The reason is that uniqueness and stability can only be studied across time. In addition, each measurement method could have its own effect on the response therefore we introduce a measurement method index (m). We assume that for all individuals the response process is the same, which implies that equation 4 applies to all respondents. And we further assume that the opinion of all individuals is expressed on the same dimension, so that responses can be meaningfully compared across individuals. This allows us to reformulate the model with random variables (equation 5). Note that 'the same response process for all individuals' does not imply that they all use the same information. As a matter of fact the information and the evaluations connected to the information should differ from individual to individual otherwise the variables in the model would be constants.

$$R_{tm} = S_{t-1} + N_t + U_t + e_{tm} \qquad \textit{Equation 5}$$

The indices of the variable R_{tm} in equation 5 refer to the fact that the response variable R is observed at some point t in time with measurement method m. We can also see that the response variable is partly determined by the variable S from a previous point in time (t-1). We could change this formulation by substituting S_t for $S_{t-1} + N_t$. We have applied these substitutions in the formulation of the model.

$$\begin{aligned} S_t &= S_{t-1} + N_t \\ O_t &= S_t + U_t \qquad \textit{Model 1} \\ R_{tm} &= O_t + e_{tm} \end{aligned}$$

In the first equation of model 1 we see the variable S_t, this variable represents evaluations of information that is stable for each individual. In the remaining of this essay we will refer to this variable as the stable opinion. We can see that this variable is the addition of the stable opinion at a previous point in time and the variable N_t. The latter variable represents evaluations of information that for each individual is new at that point in time and that will remain at least until the next point in time. We will refer to this variable as the new stable opinion. Furthermore we have introduced the variable O_t that represents the evaluation of information that for each individual is both stable (and new) and unique. In the remainder of this essay we will refer to this variable as the opinion. The variable U_t represents the evaluation of information that for each individual is unique at that time, we will refer to this variable as the unique opinion in the remainder of this essay. And in the last equation of the model we recognize the response variable R_{tm}, which represents the opinion expressed on a specific response scale m. It can be derived that the opinion is obtained by subtracting the measurement (e_{tm}) error from the response.

In order to make the link between the model presented here and quality criteria often used in research the variables in this model are standardized. To standardize the variables they first have to be expressed as deviation scores, so that their means will be zero. We denote these variables with a 'd' in the superscript. Then we can standardize (Saris and Stronkhorst, 1984) all the variables - except for N_t, U_t, and e_{tm} - which lead to the following equations (model 2) by dividing them by their standard deviation. In doing so, the effect parameters that were equal to 1 in model 1 become equal to the ratio of the standard deviation of the causal variable divided by the standard deviation of the effect variable. This result of standardization is presented in model 2

$$\begin{aligned} S^*_t &= s_{t,t-1} * S^*_{t-1} + N'_t \\ O^*_t &= c_t * S^*_t + U'_t \qquad \textit{Model 2} \\ R^*_{tm} &= q_{tm} * O^*_t + e'_{tm} \end{aligned}$$

Where:

$S^*_t = S^d_t / \sigma_{St}$ $\quad s_{t,t-1} = \sigma_{St-1} / \sigma_{St}$ $\quad N'_t = [1 / \sigma_{St}] * N_t$

$O^*_t = O^d_t / \sigma_{Ot}$ $\quad c_t = \sigma_{St} / \sigma_{Ot}$ $\quad U'_t = [1 / \sigma_{Ot}] * U_t$

$R^*_{tm} = R^d_{tm} / \sigma_{Rt}$ $\quad q_{tm} = \sigma_{Ot} / \sigma_{Rt}$ $\quad e'_{tm} = [1 / \sigma_{Rtm}] * e_{tm}$

$S^*_{t-1} = S^d_{t-1} / \sigma_{St-1}$

The coefficients in model 2 are linked to commonly used concepts in survey research:

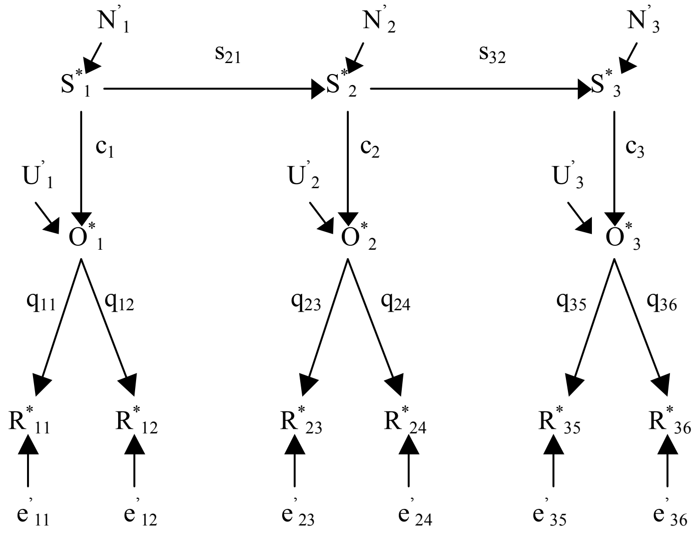
- The coefficient q_{tm} is the quality coefficient of measurement method m at time t. The square of this coefficient is the strength of the relationship between the opinion and the response, and q^2_{tm} we call the *quality of the measurement instrument*. By definition, the quality is the complement of measurement error variance. This follows from the formula: $q^2_{tm} = 1\text{-var}(e_{tm})$.
- The coefficient c_t is called the crystallization coefficient of the stable opinion at time t. The square of this coefficient is the strength of the relationship between the stable opinion and the opinion at time t, and c^2_t we call the *crystallization of the opinion*. By definition, the crystallization is the complement of the variance of the unique opinion. This follows from the formula: $c^2_t = 1\text{-var}(U'_t)$.
- The coefficient $s_{t,t-1}$ is the stability coefficient of the stable opinion. The square of this coefficient is the strength of the relationship between the stable opinion at time t-1 and t, and $s^2_{t,t-1}$ we generally call the *stability of the stable opinion*. By definition, the stability is the complement of the variance of the new stable opinion. The variance of N indicates the amount of change in the stable part. This follows from the formula: $s^2_{t,t-1} = 1\text{-var}(N'_t)$.

We can rewrite model 2 into the same form as equation 5 so that the contribution of each variable to the response is made explicit. Equation 6 shows how much the measurement error (e'_{tm}), the unique opinion (U'_t), the stable opinion at time t-1 (S^*_{t-1}), and the new stable opinion (N'_t) contribute to the response.

$$R^*_{tm} = q_{tm} * c_t * s_{t,t-1} * S^*_{t-1} + q_{tm} * c_t * N'_t + q_{tm} * U'_t + e'_{tm} \qquad \textit{Equation 6}$$

It would be attractive if the parameters of model 2 could be estimated. This requires a design and a set of assumptions. First the assumptions regarding the variance, mean and relations between the variables; it is assumed that the variables at the right hand side of the equation (equation 5) are independent[iv] of each other. The argumentation for this is given earlier in this section, and boils down to the fundamental idea that there cannot be overlap within individuals for sets of evaluations (stable, new, and unique), and also that the opinion exists independently of the measurement procedure. Second, the design to estimate the parameters. A design that allows the estimation of the parameters contains six observations using the same *question*, in a three-wave panel. In each wave this *question* should be repeated[v]. Note that this does not mean that the measurement method should be the same. Data from that design could be analyzed with the path model presented in figure 1. Feldman suggested a similar model in 1989; however that model was not identified because only 1 indicator was specified for each opinion.

Figure 1: Path model of the standardized response model.



To interpret the path model given in figure 1 we will start at the bottom. Respondents are asked to provide two responses in each wave (interview) on the same *question*, but not necessarily the same survey item. So the measurement methods may differ within and across the waves. The responses to a specific survey item at time t observed with measurement method m are denoted by R^*_{tm}. The responses contain measurement errors (e'_{tm}). If the responses are corrected for measurement errors, one obtains the opinion (O^*_t) at time t. The relationship between the opinion and the response is the quality coefficient q_{tm}.

The opinion at time t is denoted by O^*_t. The relation between the opinion and the stable opinion is the crystallization coefficient c_t. The opinion at time t is the (weighted, due to the coefficients) addition of the unique opinion (U'_t) and the stable opinion (S^*_t).

For each interview it is assumed that there is only one unique opinion (U'_t). This idea is in disagreement with what Zaller and Feldman (1992) suggested. They suggested that each time a survey question is asked the opinion will be constructed *on the spot* from unique considerations, because the context has changed. The context in which a survey question is asked is much more than the interview alone, so the change of context within an interview is rather small, compared to the change of context across the interviews. Furthermore, it should not be the intention to change the opinion within an interview. Finally, even if there is a change of opinion within the interview, i.e. different unique opinion are taken into account, then this change is just a very small fraction of the change that occurs across the interviews. Hence we might make an error by assuming that there is only one unique opinion (per respondent) within an interview, but that error will be very small.

The stable opinion at time t is denoted by S^*_t. The relation between the stable opinion at time t and time t-1 is the stability coefficient s_t. The stable opinion is the (weighted, due to the coefficients) addition of the stable opinion S^*_{t-1} at time t-1 and the new stable opinion (N'_t).

This model is in complete agreement with model 2 extended, however, to a set of 6 measures in three different waves. In the section estimation and identification we will show how this model can be identified and estimated. An empirical test of the model is presented in the section results. We shall now discuss how this model relates to the ideas suggested by Converse (1964), Achen (1975), Zaller and Feldman (1992).

Ideas of Converse (1964), Achen (1975), Zaller and Feldman (1992) unified

Model 2 unifies the arguments made by Converse (1964), Achen (1975), and Zaller and Feldman (1992) in one single model. When we apply their arguments to our model, we can generate the same results. And these are the 'problematic' finding that the correlation across time between the same variable is close to zero. Note that this result can only be found in panel surveys. Converse made the argument that the close-to-zero correlation was an indication that most people did not have an opinion. Achen claimed that low correlations were an indication that measurement instruments were of poor quality. And finally, Zaller and Feldman suggested that low correlations[vi] were an indication that people created *opinions on the spot*, whenever asked a question. We will show that all the arguments mentioned here can also follow from the model we have proposed. This will be shown using the expression (equation 7) for the across time correlation between R_{11} and R_{23} that follows from the path model shown in figure 1. Note that the response variables are measured at different points in time, with different measurement procedures.

$$\rho_{R11,R23} = q_{11} * c_1 * s_{21} * c_2 * q_{23}. \qquad \textit{Equation 7}$$

Converse (1964) attributed the small correlations across time to the fact that (most) people do not have a stable opinion. If we assume that Converse was right then the stability (s_{21}) should be 0. If we substitute this value in equation 7 then correlation between R^*_{11} and R^*_{23} becomes 0. This value is a little exaggerated compared to Converse's findings, but it shows what happens when there is (almost) no stability.

Achen (1975) attributed the small correlations across time to the fact that measurement instruments were not perfect. In addition he found that opinions were not as unstable as previously thought (Converse, 1964). If we assume that Achen was right, then quality (q_{11} and q_{23}) of the measurement instrument must be low and at the same time the stability (s_{21}) of the opinion must be high. For example when the quality is 0.5, the stability is 0.9 and we substitute these values in equation 7, then the correlation ($\rho_{R11,R23}$) across time will be 0.22 or smaller. The maximum, in this special case 0.22, is attained when the crystallization equals 1, however we would expect that the crystallization is usually lower than 1. We can see that it is possible to reproduce Achen's results and conclusions with our model.

Zaller and Feldman (1992) attributed the small correlations across time to the fact that people create *opinions on the spot*. If we assume that these authors are right, then the response is almost completely determined by the unique considerations. In other words, the crystallization is then approximately 0[vii]. When in equation 7 the value 0 is substituted for c_1 and c_2, the correlation ($\rho_{R11,R23}$) across time also becomes zero. This shows that our model also covers the suggestions of Zaller and Feldman.

The advantage of our model is that it takes all three ideas into account. It would be even nicer if we could show that the different parameters of model 2 could be estimated so that we can suggest which characteristics of a measure causes the fact that the variables are not so highly correlated through time as we might expect. Next we will show that the parameters of the model in figure 1 are identified and that the coefficients can be estimated. We will illustrate how this is done using data from a Russian panel study.

Data

The data that we use are collected in the RUSSET panel, which is the **Ru**ssian **s**ocio-**e**conomic **t**ransition panel. This study has been done by a group of Dutch researchers in collaboration with the Institute for Comparative Social Research (CESSI) in Moscow. The panel existed from 1992 until 1999 and consisted of 7 waves with an interval of approximately one year. The universe covered is the population aged 18 and above living permanently in Russian territory. The area probability sampling method for national surveys (Kish, 1965) has been used to draw a multistage sample. The method of administration of the survey was face-to-face interviewing (paper and pencil) in the respondent's home. More information about the RUSSET panel can be found on http://www.vanderveld.nl.

In the RUSSET panel variables have been measured that largely follow the design we need for our model. We picked 7 different variables that are cover a variation of issues, from satisfaction with housing to policies on minorities. All these variables were, in agreement with the design, observed six times, twice in wave 3, twice in wave 4, and twice in wave 7 of the panel. Across and within the waves we have used the same core *questions* to measure our variables. The formulations of these core *question*s are given in table 1, exact formulations can be found on http://www.vanderveld.nl.

Table 1: Formulation of the questions used in this study.

Proud Russian	As a Russian there is not much to be proud of
Minority education	National minorities should be allowed to have education in their own language
Minority medical access	Minority groups who live here illegally, should not get access to medical and educational facilities
Satisfaction housing	How satisfied are you with your housing conditions at present?
Satisfaction contacts	How satisfied are you with your social contacts at present?
Disappointing people	Most people are disappointing once you get to know them better
Influence over life	I have little influence over the things that happen to me

For all seven different variables we can estimate the model. We will refer to these sets of variables as datasets, although they are not datasets in the natural sense. So in total we have 7 datasets, each containing six variables observed with the same *question*. For each dataset we have estimated the covariance matrix with PRELIS2E (Jöreskog and Sörbom, 1996). Sample sizes and case deletion methods for each dataset can be found in appendix A. The correlations matrices and univariate statistics can be downloaded from http://www.vanderveld.nl. The covariance matrices were analyzed using the maximum likelihood estimation procedure in LISREL 8E (Jöreskog and Sörbom, 1993). This procedure was applied to the data even though some variables are not normally distributed. However, the robustness studies of Anderson and Amemiya (1988), Satorra and Bentler (1990), and Satorra (1992) have shown that this so called “quasi maximum likelihood” estimator is robust under quite general conditions.

Comparison of the RUSSET data with the NES data

The RUSSET panel is a panel study focused on socio-economic transitions. During the period just after the collapse of the Soviet Union many social, political, economical, and cultural changes have occurred that influenced all Russians' living conditions. The panel was used to capture the individual changes and living conditions during that period. In this respect the RUSSET data is different from the National Election Studies, because the NES studies are meant to track political attitudes. In table 2 we presented some figures from the NES panel ('56-'60) and the RUSSET panel. The NES data are taken from Achen's (1975) article on mass attitudes. The RUSSET data are the same data that are used in this article.

The figures in the column 'across' are the averages of the correlations between the three waves. The figures in the column 'within' are the averages of the correlations of the repeated observations within each wave. The variables from the NES panel are different from the RUSSET panel, which makes variable-wise comparison of the correlations not possible. We therefore can only look at the size of the correlations. The low correlations in the NES panel, especially on the *housing and electricity* issue, lead Converse (1964) to develop his Black and White model. He took these low correlations as indications of 'non-attitudes'. It appears that for the RUSSET data, the correlations are also in the same range as the *housing and electricity* issue in the NES panel[viii]. The simple thought would be to assume – in line with the black and white model - that (most) Russians have no attitudes on the issues. But such an assumption seems implausible given the size of the within wave correlations for most issues. If Russians do not have an opinion, and every response would be an opinion created *on the spot* à la Zaller and Feldman (1992), then the average within wave correlations would be a lot lower than is the case in table 2. The average within wave correlation is in fact rather high and does not indicate non-attitudes or opinions created *on the spot*. Some people might say that the within wave correlations are biased by so-called memory effects, which means that the response of the previous question is recalled when the second (repeated) question is asked. However, several studies have been done to detect the impact of questionnaire design on memory effects and the results were that memory effects can be minimized if:

1. The time between the repeated measures should be more than 20 minutes, to put it differently there should be more than 40 questions asked before repeating a measure (Van Meurs and Saris, 1995);
2. The questions between the repeated measures should be largely on similar topics (Van Meurs and Saris, 1995); and,
3. Different response scales (measurement methods) are used for the repeated measure (Van der Veld and Saris, 2000).

In the RUSSET panel we have taken these considerations into account when the questionnaires were designed. Therefore the presence of large memory effects in our data will be quite unlikely. So, the repeated measures are feasible when design constraints are taken into account, and they are a necessary feature of our design to be able to distinguish between temporary and stable components in the response. The NES data, do not provide this feature, therefore, we used the RUSSET panel data.

Table 2: Average Pearson correlations of NES and RUSSET panel.

NES PANEL 1956-1958-1960*		RUSSET PANEL 1994-1997-1999		
	Across**		**Across****	**Within*****
Maintain army overseas	0.31	Proud Russian	0,13	0,44
Foreign aid	0.36	Disappointing people	0,16	0,67
Gov. intervention in housing & electricity	0.37	Minority medical access	0,17	0,79
Isolationism	0.39	Minority education	0,20	0,73
Federal aid to education	0.45	Satisfaction contacts	0,22	0,74
Federal assistance to negroes	0.50	Influence over life	0,22	0,60
Party identification	0.83	Satisfaction housing	0,42	0,85

* Source Achen (1975)

** Average Pearson correlation across three waves of panel data

*** Within each wave, three in total, there are repeated observations. These three correlations, between the repeated observations, are averaged.

Identification and Estimation

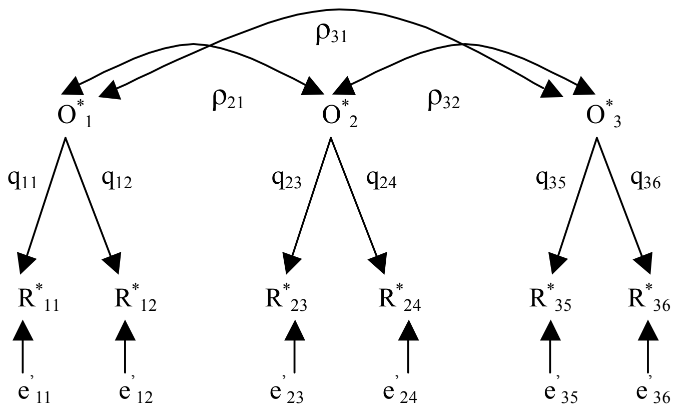
The estimation of the parameters of our model - presented in Figure 1 - will be made with LISREL (Jöreskog and Sörbom, 1993), which is a program for the analyses of covariance structures and not for the analyses of correlation (standardized covariance) structures (Cudeck, 1989; Jöreskog, 1970). Because model 2 is a standardized model, there might be cause for concern when we want to estimate the parameters. In order to avoid problems with respect to this issue we will analyze covariance matrices. Parameter estimates produced with that procedure are standardized to obtain the parameters that we are interested in.

The path model could be estimated in 1 step, but we used a two-step procedure to estimate the parameters of the model. In the first step the measurement model, shown in figure 2, will be estimated. In the second step the structural model, shown in figure 3, will be analyzed. The analysis will be on the estimated relations between the opinions in the measurement model (step 1). The reader should see that if the model in figure 3 is put on top of the model in figure 2 a model emerges that is similar to the model in figure 1, so with the two-step procedure we have strictly separated the measurement and the structural model. The advantage is that possible misspecifications in the measurement model do not propagate through the structural model. Two-step procedures have also been used and/or advocated among others by Anderson and Gerbing (1988), Ping (1998, 2001), and Jöreskog (1998). These authors also estimate the measurement model in the first step. In the second step they estimate the whole model including the measurement model, but they either fix the parameters of the measurement model or use the specifications of the measurement model suggested by step 1.

<u>Step 1: The measurement model</u>

In the measurement model, presented in figure 2, there are six observed variables observed at three different points in time (t=1, t=2, and t=3). The metric - unit of measurement - of the opinions is unspecified by definition; therefore the model is not identified. In order to specify a metric for the opinions the relation between the opinion in each wave and the first observed variable in each wave is fixed to 1. The effect of this is that the metric of each opinion is equal to the metric of the first observed variable in each wave. Now the model is identified and the parameters can be estimated. It is important that the estimates are correct (no misspecifications), because the estimated relations between the opinions are used in step two as input for the structural model. If errors are present in these relations, then the estimates of the structural model will be wrong.

Figure 2: The measurement model.



<u>Additional assumption to deal with a misspecification</u>

If the design of the study is correct, then the model will not be misspecified. In reality the available data is not in agreement with our design on the point that instead of different measurement method, we used the same measurement procedure in some instances. For the variables satisfaction contacts and satisfaction housing the measurement procedures of the first observed variable in wave 3 and 4 are the exactly the same. For the other variables the measurement procedures of the first observed variable in wave 3 and 7 are the same. We used a standard procedure to deal with this misspecification of the model (figure 2). That is when identical survey questions are used, the random errors can also be assumed identical. The argumentation behind this constraint is based upon the assumption that a specific measurement instrument (survey question) has a specific measurement error connected to it (Lord and Novick, 1968; Heise, 1969).

<u>Estimation of and results from the measurement model</u>

The results of the analyses are presented in table 3. We have analyzed the seven variables with the extra restriction (discussed above) on the variances of the random errors for specific observed variables. The random errors are connected to the quality coefficients in the following way: q^2_{tm} = 1-var(e_{tm}). In table 3 the columns of some quality coefficients are with a light grey background. For each variable there are cells with a light grey background the random errors connected to those cells are set equal (Lord and Novick, 1968; Heise, 1969).

It turns out that with the extra restrictions most datasets yield acceptable results. However, for 'Minority education' and ' Minority medical access' the error variances connected to q_{22} were negative. Fortunately the negative error variances were not significantly different from zero, therefore we fixed the negative error variances to zero and re-analyzed the datasets again. The results are now acceptable. We can see that there is a large variety in the estimates of the quality coefficients. When we consider the quality of the measurement instruments used, the worst quality found is 0.25 (which is 0.50^2) and the best quality found is 1. Furthermore, the average quality of the measurement instruments used in this study is 0.68. This means that at average the strength of the relation between the opinion and the response is 0.68, which is low compared to what Scherpenzeel (1995a) found in an analysis of the quality of satisfaction measures across Europe. This does not mean that in Russia data is of low quality on average.

It might be that the measurement procedures we have used in this study have a negative effect on the quality. In the last three columns, the standardized relations between the opinions are presented. The variances the opinions (O^*_t) are not presented here, because after standardization they are 1 by definition. We can see that the correlations are just higher than the correlations presented in table 2. Which should be the case (although the correlations in table 2 are averaged), because here we have corrected the observed correlations for measurement error.

Table 3: Completely standardized results of the measurement model.

	χ^2	df	q_{11}	q_{12}	q_{21}	q_{22}	q_{31}	q_{32}	ρ_{21}	ρ_{31}	ρ_{32}
Proud Russian	5.72	7	0.70	-0.52	0.79	-0.50	0.68	-0.84	0.36	0.16	0.19
Minority education	5.28	8	0.95	0.73	0.84	**1.00**	0.96	0.70	0.38	0.15	0.23
Minority medical access	6.55	8	0.81	0.97	0.88	**1.00**	0.77	0.91	0.22	0.21	0.26
Satisfaction housing	11.94	7	0.88	0.97	0.90	0.93	0.93	0.94	0.61	0.42	0.51
Satisfaction contacts	6.15	7	0.82	0.88	0.84	0.86	0.87	0.87	0.33	0.27	0.30
Disappointing people	5.49	7	0.83	-0.77	0.92	-0.79	0.79	0.82	0.30	0.17	0.21
Influence over life	8.83	7	0.83	-0.81	0.81	-0.81	0.78	-0.60	0.39	0.23	0.37

Step 2: The structural model

The structural model (figure 3) is known in the literature as the quasi-simplex model. Many people (Heise, 1969; Wiley & Wiley, 1970; Achen, 1975; Wheaton et al., 1977; Alwin & Krosnick, 1991; Coenders et al., 1999) have used this model to estimate the stability and reliability however we use this model to estimate the stability and the crystallization of an opinion.

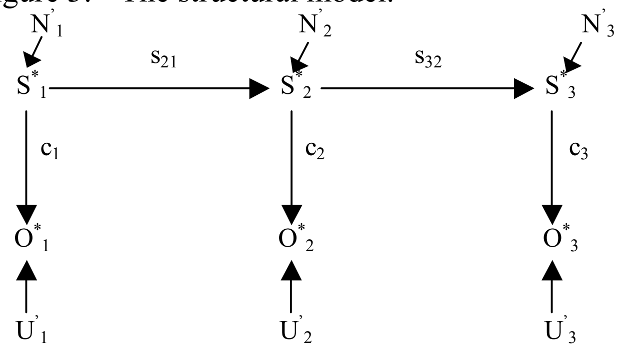
Because the covariance matrix is analyzed the metric (unit of measurement) of the stable opinions is unspecified. In order to specify a metric for the stable opinions the relation between the stable opinion and the opinion is fixed at 1 in each wave for the estimation of the unstandardized coefficients. The effect of this is that the metric of each stable opinion is equal to the metric of the opinion in each wave, which is in effect the metric of the first observed variable in each wave (see step 1). The structural model (figure 3) has 11 parameters to be estimated while there are only three variances and three covariances (from step 1); hence there are 6 elements as the empirical basis for the estimation. Because the metric for the stable opinions is specified, the number of parameters to be estimated is reduced by three (the three variances of the stable opinions). This results in 8 parameters to be estimated and only 6 known elements[ix]. Therefore the parameters of the structural model (figure 3) are not identified without further restrictions. We will make the following additional assumptions concerning the stable, new stable, and unique opinions in the population to solve the problem of under-identification.

1. For starters, we use the common procedure (Wiley & Wiley, 1970) in quasi-simplex models, which means that we assume that the variances of the unique opinions are equal to each other across the interviews. This assumption will lead to an increase of 2 degrees of freedom, because only 1 variance has to be estimated (and not 3), and therefore the model is just identified. The substantive meaning of this assumption is that across the interviews the variation of the unique opinion remains the same. This does not mean that for each individual the same unique opinion is used at each interview. On the contrary, what is unique for each individual should deviate greatly from interview to interview. If not, then the unique opinion is no longer a variable but a constant.
2. Because there is still no test of the model (it is only just identified) we will add more

restrictions to the model which are also made plausible. First of all, we assume that S_t is a fixed part of the S_{t-1}, and as a result the stability coefficients are equal across time. The substantive meaning would be that stability is a stable characteristic of the topic under investigation. This assumption results in an increase of 1 degree of freedom. Second, we assume that across the interviews the variance of the new stable opinions (N_t) is the same. The substantive meaning of this assumption is that for each individual 'new stable' information will come in at a constant pace. This assumption results in an increase of 1 degree of freedom.

If all these assumptions are made, then the structural can be estimated with 2 degrees of freedom.

Figure 3: The structural model.



<u>Limitation of the structural model</u>

In the structural model (figure 3) it is assumed that all covariance between the opinions is explained by the structural model. For example the covariance between the opinion at the first and the third interview can only be caused by the stable opinion S_1 through the stable opinion S_3. To put it differently, there is no direct relation between the first and the third moment. In real life, it is however not uncommon that we forget something that is recalled as an important feature at a later point in time. In other words, if a question is asked three times, it is possible that we forget an important aspect of the opinion that we thought of the previous time, and will recall the next time. This could mean that either the unique opinion for an individual is not as unique as we have previously assumed, or that the new stable opinion for an individual is not as new as we have previously assumed.

This misspecification[x] cannot be solved by introducing a covariance between U_1 and U'_3 ($\rho_{u1,u3}$), or between N'_1 and N'_3 ($\rho_{N1,N3}$), or by introducing a direct effect between S^*_1 and S^*_3 (s_{31}). The reason for this is that these parameters are not identified. The consequences of the absence of a direct relation between the opinion at time 1 and time 3, given that such an effect is present in reality, has been studied by Coenders et al. (1999). They have shown that only minor violations cause great instability of the estimates of the crystallization coefficients. In addition standard errors can become very large, and many analyses will yield inadmissible parameter estimates. As a consequence, if such a relationship between time 1 and 3 exists, then our structural model is the wrong model. We do not want to explore this further, but a possible solution is to assume a different structural model, for example a single factor model which could indicate a predisposition.

The structural model, whether it is misspecified or not, will also fail when the correlations between the opinions (ρ_{21}, ρ_{32}, and ρ_{31}) are close to zero. This is because it will lead to a situation were the model parameters – although mathematically defined – cannot be estimated, this situation is called empirical under-identification. In that case there are three theoretical possibilities. Either one or both of the stability coefficients are zero, or the crystallization coefficients are zero, or both are very small. It should be clear that it that case we could conclude without any estimation procedure that the opinion is created *on the spot* and/or there was a huge change of opinion.

Estimation of and results from the structural model
We have analyzed the seven variables with the extra restrictions (additional assumptions) so that we have a test for the structural model. Of course this also implies that we assume that the change process has a certain structure, which we discussed above. The completely standardized results of these analyses are presented in table 4.

In the first column the name of the variable is given, which refers to the question in table 1. In each row, the estimates for each dataset are presented. At first glance most analyses yield an acceptable chi-square fit index, meaning that we accept the specified models. However, for 'satisfaction housing' the chi-square hints at rejecting the model. It has, however, been proven that the chi-square is (besides the commonly recognized sample size problems) dependent on the size of the effect parameters in the model (Saris & Satorra, 1985; Saris, Satorra, & Sörbom, 1987c). That is, the larger the estimates of the effect parameters the more likely the model is to be rejected. It can be seen that for 'satisfaction housing' the sample size is much larger than for the most other variables (appendix A) and it can be seen that the estimates of the crystallization coefficients are the highest of all variables, which could explain why this variable has a significant chi-square value.

The results of the equality constraints (the additional assumptions to identify the model) are easily detected, apart from small deviations caused by the standardization, when the estimates in table 4 are compared. The estimates of the crystallization coefficients are (almost because of standardization) the same across the waves for each variable, and the estimates of the stability coefficient are also (almost because of standardization) the same between two successive interviews. Still, there is a large variety in the estimates of the crystallization coefficients and stability coefficients across variables. When we consider the crystallization of the opinions, the lowest crystallization found is 0.25 (which is 0.50^2) and the highest crystallization found is 0.76. Furthermore, the average crystallization of the opinions studied here is 0.49. When the stability is evaluated, we find that the lowest stability is 0.25 (for 'Minority education') and the highest stability is 0.90 (for 'Minority medical access'). The average stability is 0.51. On average both the stability and the crystallization are quite low, which means that the effect of the unique opinion and the effect of the new stable opinion is quite large.

It is difficult to go into further detail about these estimates. For starters, it difficult to say something about each variable with respect to the size of the estimates, because we are not experts on the issues that were measured. In addition, there is (yet) no comparison material. What however can be done is to look at the results as a whole; what we can learn from them and whether they make sense. From a substantive point of view some of these results make sense. For example satisfaction and policy issues are completely different topics. According to Converse (1964) most people lack an opinion on policy issues. In contrast one would expect that most people know how satisfied they are. If we compare the average crystallization for satisfaction and minority policies we find that indeed the average crystallization of satisfaction is higher (difference =0.14) than for minority policies, and the same holds for the stability (difference=0.10). For other issues, it is not so easy to make a comparison, because it is less clear to us whether it should be expected that one should be higher than the other. In addition, the estimates have standard errors, and if they are taken into account the differences may disappear. So because we don't know whether or not there should be differences, it makes no sense to discuss all estimates in detail. Still one thing that asks for some explanation is the variable 'minority medical access'. It happens that the stability is very high, higher than for all the other variables. But this is a policy issue and therefore we would not expect this to happen. However, stability as Converse (1964, 1970) saw it is a combination (in our model) of crystallization and stability. Would we combine those two parameters into a new parameter, call it *Converse-stability*, which is the square of the product of crystallization and stability, then a different picture appears in line with Converse's expectations. That is for 'minority medical access' the *Converse-stability* is very low: 0.21. So, in this case the stability might be very high, which could be explained with a predisposition that one cannot refuse medical help to anybody, but at the same time the crystallization might be very low, which could be explained by mixed feelings about minorities, and or the structure of the health-care system. However, this is educated guessing because we are not experts on the field of minorities and the Russian Federal health system. But what counts here is that a low *Converse-stability* does not imply that there is no stable predisposition behind the response.

Table 4: Completely standardized results of the structural model.

	χ^2	df	c_1	c_2	c_3	s_{21}	s_{32}
Proud Russian	4.56	2	0.69	0.67	0.66	0.63	0.60
Minority education	3.32	2	0.78	0.77	0.77	0.51	0.50
Minority medical access	0.19	2	0.50	0.51	0.52	0.91	0.91
Satisfaction housing	16.56	2	0.87	0.87	0.87	0.75	0.75
Satisfaction contacts	0.93	2	0.62	0.61	0.60	0.85	0.85
Disappointing people	1.44	2	0.64	0.62	0.62	0.66	0.65
Influence over life	0.00	2	0.78	0.78	0.78	0.62	0.62

Discussion of the implications

We have estimated the parameters for 7 different variables using data from the RUSSET panel in two steps. In the first step we have estimated the quality coefficients. There were 42 estimates of the quality coefficient in the range 0.5 to 1.0, and the average quality of the measurements instruments used in this study is 0.70. In the second step we have estimated the crystallization and the stability coefficients. There were 21 estimates of the crystallization coefficient in the range 0.5 to 0.87, and the average crystallization of the opinions was 0.49. There were 12 estimates of the stability coefficient in the range 0.50 to 0.91, and the average stability of the opinions was 0.51. We could discuss these figures based on their sizes but that would force us to use relative terms like high, mediocre, and low estimates. Because there is no other material for comparison than this study, and because terms like mediocre have no fixed meaning, we go in another direction. A more fruitful discussion follows when we use the estimates to determine the percentage of explained variance of an observed variable and the shared variance of two observed variables. For example, it is interesting to know how much the stable opinion contributes to the response R^* (the * indicated that observed response variable is standardized) and how this is determined.

The explained response variance

When we want to know how much the stable opinion (S_t^*) contributes to the response R_{tm}^* we have to square product of the path coefficients in the model (figure 1) between them. So, the contribution equals $c_t^2 * q_{tm}^2$. When this product is multiplied by 100 we get the percentage of the explained variance. This contribution can also be referred to as the strength of the relationship between the stable opinion and the response R^*. We could derive the explained variance of the response R^* due to the stable opinion at a previous point in time in a similar way. In table 5 the percentage of explained variance of the response R^* due to the stable opinion of 'proud Russian' is given. The presented figures show that 23% of the response R_{11} is explained by S^*_1. However, when we look at the two successive waves, we see that this percentage drops to 3%. This means that only 3% of the variance of the response at the third wave (R^*_{31}) is explained by S^*_1 (the stable opinion in the first wave). In summary the stable opinion of 'proud Russian' has hardly any effect on the response, especially not over time, but for other topics this may be very different. This example, however, shows that the stability, the crystallization, and the quality all have an effect on the strength of the relation between the stable opinion and the response.

Table 5: Percentage of response variance explained by the stable opinions of 'proud Russian'.

	S^*_1	S^*_2	S^*_3
R^*_{11}	23 %	-	-
R^*_{21}	11 %	28 %	-
R^*_{31}	3 %	7 %	20 %

It is also possible to compute the percentage explained variance of the response due to the opinion (O^*_t), this would be equal to the square of the quality coefficient (q_{tm}^2) multiplied by 100. In table 6 these figures are presented. It should be clear that these percentages are higher than in table 5, because we do not take the crystallization and the stability into account. Still we can observe that there is a large discrepancy between the opinion and the response at a specific point in time. For example in the first wave, only 49% of the variance in the response is explained by the opinion. The rest, that is 51% of the variance, is explained by measurement error.

The lesson here is that we should strive for measurement instruments with high quality, so that the agreement between the opinion and the response is as large as possible (close to 100%). For a scientific procedure for the development of high quality measurement instruments we refer to Scherpenzeel (1995b), Veld van der W. M. et al (2001), and Saris et al. (2002).

Table 6: Percentage of response variance explained by the opinions for 'proud Russian'.

	O^*_1	O^*_2	O^*_3
R^*_{11}	49 %	-	-
R^*_{21}	-	62 %	-
R^*_{31}	-	-	46 %

There is another lesson to be learned regarding the crystallization and stability. The crystallization and stability are characteristics of the variable under investigation. This variable is again a characteristic of the public (Allport, 1937). Thus, in order to change, or strengthen the crystallization and the stability, we have to educate the public. Brand advertisement is a good example of such a strengthening procedure. Alternatively, one could try to make the public more interested or engaged in political matters, so that the public will and can search for relevant information. For political scientists, public opinion researchers, political parties, lobby groups, etc. there are also other methods available for informing the public, such as deliberate polling or the choice questionnaire (Neijens, 1987, 2004; Bütschi, 2004). These procedures provide the public with relevant knowledge, enabling them to form a considered opinion.

The shared response variance: correlations at different levels in the model

We can also discuss, besides explained variance, the shared variance at different levels in the model (figure 1). For example the percentage of shared variance between the stable opinions of subsequent waves equals the square of the stability coefficient ($s_{j,j-1}$), and the shared variance between the responses R^*_{11} and R^*_{21} equals $q_{11}^2 * c_1^2 * s_{21}^2 * c_2^2 * q_{21}^2$. In table 7 the percentage of shared variance at different levels in the model (figure 1) is presented for 'proud Russian'. It should be clear that the lower we get in the model (figure 1), the lower the percentage of shared variance is. The shared variance is related to the correlation in the following way: the square root of the percentage of shared variance is the correlation coefficient, hence the correlation (in the population) equals: $\rho_{R11,R21} = q_{11} * c_1 * s_{21} * c_2 * q_{21}$. What we observe are the correlations between the response variables and these correlations are for many political topics rather low. And also for 'proud Russian' are these correlations low, namely √0.03 (=0.18), √0.01 (=0.10), and √0.02 (=0.14). Should we follow Converse (1964) then these low correlations would be due to the absence of a stable opinion. It follows from table 7 that the correlation between the first and second stable opinion is 0.6 (= √0.4), and that is far from zero. On the other hand, the correlation between the first and the second opinion is 0.3 (=√0.09), this is a lot closer to zero.

So, it depends on what level we are (stable opinion or opinion) whether Converse's conclusion was correct for this specific topic. Should we follow Achen (1975) then these low (0.18, 0.10, and 0.14) correlations between the observed variables would be due to measurement error, and the stability would be high. The correlation between the first and second stable opinion is 0.6 (= √0.4) which is *rather* high. So in that respect - low observed correlations and still high opinion stability - Achen was right. On the other hand, should we look at the stability of the opinion (0.3 = √0.09), then Achen's conclusion was far from correct. Should we follow Zaller (1992) then these low (0.18, 0.10, and 0.14) correlations would be due to poorly crystallized stable opinions. For 'proud Russian' the estimates of crystallization coefficients are 0.69, 0.67, and 0.66, which is far from zero (zero=not crystallized). On the other hand, even when the crystallization coefficients are far from zero, the unique opinion plays an important role and explains more than 50 % ($1-0.69^2$=52%) of the variance in the opinion (O^*_{t}) of 'proud Russian'.

Table 7: Percentage of shared variance of S, O, and R for 'proud Russian'.

	S_1	S_2	S_3
S_1	-	-	-
S_2	40 %	-	-
S_3	14 %	36 %	-

	O_1	O_2	O_3
O_1	-	-	-
O_2	9 %	-	-
O_3	3 %	7 %	-

	R_{11}	R_{21}	R_{31}
R_{11}	-	-	-
R_{21}	3 %	-	-
R_{31}	1 %	2 %	-

The lesson here is that the shared variance, or the population correlation[xi], is defined as a function of three components: opinion stability, opinion crystallization, and measurement quality. This means that the cause of response instability (low across time correlation) cannot and should not be uniquely attributed to unique considerations (Zaller and Feldman, 1992), or measurement error (Achen, 1975), or non-attitudes (Converse, 1964), but by a combination of these elements.

Opinion Strength vs. our model

Our model could also produce an indication of attitude or opinion strength if we follow the definition of Petty and Krosnick (1993) who define attitude strength by two phenomena. They assert that a strong attitude should be stable, i.e. resistant to change and persistent over time, and crystallized i.e. have a strong impact on thinking and behavior. In this formulation, one can recognize two components – stability and crystallization – of our model. It follows from this definition that in our model a strong attitude would mean that both the stability and the crystallization are high. Petty and Krosnick suggest that stability will be high if people think of the issue at stake as important and if they have considerable relevant knowledge and are certain about the attitude. The first component will produce stability by selective perception, while knowledge and certainty will produce stability because they can easily generate counter arguments. Crystallization will be higher if the accessibility of the attitude in memory is high and if the attitude components have a low degree of ambivalence. For more details on this issue we refer to Petty and Krosnick (1993), and Bizer et al. (2004). When an estimate of attitude strength is needed Bizer et al. propose that different measures should be used that predict the stability and crystallization of an attitude. A difficulty with that approach is that one needs to measure a multitude of strength related attributes to predict the strength of the attitude for each individual. In our approach the attitude strength is represented by just two parameters – the crystallization and stability parameters - of the model. A possible drawback of our approach is that we cannot obtain individual measures of the strength of the opinion, however, we could estimate the crystallization and stability for different (homogeneous) subgroups in the population.

Summary

The theory we suggest unifies the important arguments that explain response inconsistency. Our starting point is the survey item itself. The *question* triggers the activation of nodes in a semantic network (Quillian, 1966) - a person's mind. The network that is activated *somehow* represents the respondent's opinion. Where *somehow* refers to an integration process that yields information in terms of a numerical degree of positive and negative, pro or con, etc. We have distinguished 3 different types of sets of nodes in the integration process. There are the nodes that are activated each time the *question* is asked. There are nodes that will be activated at some point in time and then every time the *question* is asked. And there are nodes that are activated only once regardless of how many times the *question* is repeated. The result of the integration process of the three different sets is what we call the opinion at a specific point in time. This opinion is however not the same as the response. In order to produce a response, the respondent has to match his opinion to a response option. We assume that this matching process is stochastic in nature, and therefore errors are possible. The result of the matching process, the response, thus contains the opinion and an error due to the measurement procedure. This model has been translated to a structural equations model for the population. This structural equations model contains parameters that are linked to commonly used concepts in survey research such as the stability of the opinion, the crystallization of the opinion, and the reliability of the response. A design to estimate these parameters contains six observations with the same *question* across three different interviews in a panel. We have used 7 different variables from the RUSSET panel to estimate the parameters of the model. This did not yield serious problems, indicating that for these variables the response process is in agreement with our ideas. However, it cannot be expected that this will be the case for all variables. That is, the theory is very general and does cover the response process for all type of variables, opinions, facts, predispositions, etc. But, the model that we use to estimate the parameters has certain limitations, and that might cause problems for some variables. We have anticipated these limitations in the paragraphs *Additional assumption to deal with a misspecification* and *limitations of the structural model* Even though we have discussed these limitations they require further study so that an even more general model could evolve.

References

Achen C. H. 1975. "Mass political attitudes and the survey response." *American Political Science Review* 69:1218-31.

Allport F. H. 1937. "Toward a science of public opinion." *Public Opinion Quarterly*. 1(1):6-23.

Alwin D. F. and J. A. Krosnick. 1991. "The reliability of survey attitude measurement: the influence of question and respondent attributes." *Sociological Methods and Research*, 20(1):139-81.

Anderson J. C. and D. W. Gerbing. 1988. "Structural equation modeling in practice: A review and recommended two-step approach." *Psychological Bulletin* 103:411-23.

Anderson, J. C., and D. W. Gerbing. 1984. 'The effect of sampling error on convergence, improper solutions and goodness-of-fit indices for maximum likelihood confirmatory factor analysis.' *Psychometrika* 49: 155-173.

Anderson N. H. 1982. "Cognitive algebra and social psychophysics." In B. Wegener. *Social attitudes and psychophysical measurement*. Hillsdale, New Jersey, Lawrence Erlbaum Associates Publishers.

Anderson N. H. 1996. *A functional theory of cognition*. Lawrence Erlbaum Associates Inc.

Anderson N. H. 1981. *Foundations of information integration theory*. New York: Academic Press.

Anderson N. H. 1970. "Functional measurement and psychophysical judgment." P*sychological Review* 77:153-70.

Anderson T. W. and Y. Amemiya. 1988. "The asymptotic normal distribution of estimators in factor analysis under general conditions." *The Annals of Statistics* 16:759–771.

Andrews F.M. 1984. "Construct validity and error components of survey measures: a structural modeling approach." *Public Opinion Quarterly* 48(2):409-42.

Bizer Y., P. S. Visser, M. K. Berent, J. A. Krosnick. 2004. "Exploring the Latent Structure of Strength-Related Attitude Attributes." In: W. E. Saris and P. M. Sniderman. *Studies in Public Opinion: Gauging Attitudes, Nonattitudes, Measurement Error, and Change*. Princeton, N.J.: Princeton University Press.

Boynton G. R. 1995. "Computational modeling: A computational model of a survey respondent." In: M. Lodge and K. M. McGraw. *Political Judgment: Structure and process*. Ann Arbor: The University of Michigan Press. Pp. 229-48.

Bütschi D. 2004. "The role of information on considered opinions. The example of the "Choice questionnaire." In: W. E. Saris and P. M. Sniderman. *Studies in Public Opinion: Gauging Attitudes, Nonattitudes, Measurement Error, and Change*. Princeton, N.J.: Princeton University Press.

Coenders G., W. E. Saris, J. M. Batista-Foguet, A. Andreenkova. 1999. "Stability of three-wave simplex estimates of reliability." *Structural Equation Modeling* 6:135-157

Converse P. E. 1974. "Non-attitudes and American public opinion: Comments: The status of nonattitudes." *The American Political Science Review* 68(2):650-660.

Converse P. E. 1964. "The Nature of Belief Systems." In: D.E. Apter. *Ideology and Discontent*. New York: Free Press. Pp. 206-61.

Cudeck R. 1989. "Analyses of correlation matrices using covariance structure models."

Psychological Bulletin 105(2):317-327.
Erikson R. S. 1979. "The SRC panel data and mass political attitudes." *British Journal of Political Science* 9:89-14.
Erikson R. S. 1978. "Analyzing one variable three wave panel data: A comparison of two models." *Political Methodology* 2:151-66.
Feldman S. 1989. "Measuring issue preferences: The problem of response instability." *Political Analysis*. 1:25-60.
Heise D. R. 1969. "Separating reliability and stability in test-retest correlation." *American Sociological Review* 34:93-01.
Hochschild J. 1981. *What's fair? Americans' attitudes toward distributive justice*. Cambridge: Harvard University press.
Holleman B. 2000. *The forbid/allow asymmetry. On the cognitive mechanisms underlying wording effects in surveys*. Rodopi, Amsterdam - Atlanta.
Jöreskog K. G. 1998. "Interaction and nonlinear modeling: Issues and approaches." In: R. E. Schumacker and G. E. Marcoulides. *Interaction and nonlinear effects in structural equation modeling*. Lawrence Erlbaum associates, Inc.. Pp. 63-100.
Jöreskog K. G. and D. Sörbom. 1996. *PRELIS® 2 User's reference guide*. Chicago (Il.): Scientific Software International.
Jöreskog K. G. and D. Sörbom. 1993. *LISREL® 8 User's reference guide*. Chicago (Il.): Scientific Software International.
Jöreskog K. G. 1970. "A general method for analysis of covariance structures." *Biometrika* 57:239-251.
Judd C. M., J. A. Krosnick and M. A. Milburn. 1981. "Political involvement and attitude structure in the general public." *American Sociological Review* 46:660-69.
Judd C. M. and M. A. Milburn. 1980. "The structure of attitude systems in the general public. Comparisons of a structural equation model." *American Sociological Review* 45:627-43.
Kish L. 1965. *Survey Sampling*. New York: J. W. Wiley and Sons Inc.
Krosnick J. A. and D. F. Alwin. 1987. "An evaluation of a cognitive theory of response-order effects in survey measurement." *Public Opinion Quarterly* 51:201-19.
Lodge M. and K. M. McGraw eds. 1995. *Political Judgment: Structure and process*. Ann Arbor: The University of Michigan Press.
Lord L. M. and M. R. Novick. 1968. *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley.
Luce R.D. 2000. *Utility of Gains and Losses: Measurement Theoretical and Experimental Approaches*. Mahwah, NJ: Erlbaum
Marsh, H. W. 1989. 'Confirmatory factor analysis of multitrait multimethod data: many problems and few solutions.' *Applied Psychological Measurement* 13: 335-361.
Münnich Á. 1998. *Judgment and Choice*. Ph.D. dissertation, University of Amsterdam.
Neijens, P. 2004. "Coping with the non attitudes phenomenon: a survey research approach." In: W. E. Saris and P. M. Sniderman. *Studies in Public Opinion: Gauging Attitudes, Nonattitudes, Measurement Error, and Change*. Princeton, N.J.: Princeton University Press.
Neijens P. 1987. The Choice Questionnaire. Design and evaluation of an instrument for collecting informed opinions of a population. *Amsterdam: Free University Press.*
Petty R. E. and J. A. Krosnick. 1993. *Attitude strength: antecedents and consequences.* Hillsdale: Erlbaum.
Ping R. 2001. "Latent variable interaction and quadratic effect estimation: a two-step technique using structural equation." (http://www.wright.edu/robert.ping/) Updated from Ping R. 1996. *Psychological Bulletin* 119:166-175.
Ping R. 1998. "EQS and LISREL Examples Using Survey Data." In: R. E. Schumacker and G. E. Marcoulides. *Interaction and nonlinear effects in structural equation modeling*.

Lawrence Erlbaum associates, Inc.. Pp. 63-100.
Quillian M. R. 1966. *Semantic Memory*. Cambridge, MA: Bolt, Beranak, and Newman.
Rindskopf, D. 1984. 'Structural equation models: empirical identification, Heywood cases, and related problems.' *Sociological Methods & Research* 13: 109-119.
Saris W. E., W. M. van der Veld, and I. M. Gallhofer. 2004. "Development and Improvement of Questionnaires Using Predictions of Reliability and Validity." In: S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, and E. Singer. *Methods for Testing and Evaluating Survey Questionnaires*. Wiley and Sons, New York.
Saris W. E. and F. M. Andrews. 1991. "Evaluation of measurement instruments using a structural modeling approach." In: P. P. Biemer, R. M. Groves, L. E. Lyberg, et al. *Measurement errors in surveys*. New York: Wiley. Pp. 575-99.
Saris W. E. 1987a. *Variation in response functions: A source of measurement error*. Amsterdam: Sociometric Research Foundation (SRF).
Saris W. E. 1987b. "Individual response functions and correlations between judgments." In: W. E. Saris and I.N. Gallhofer. *Sociometric research*. Vol. 1. *Data collection and scaling*. London: MacMillan.
Saris W. E., A. Satorra, D. Sörbom. 1987c. "The detection and correction of specification errors in structural equation models." *Sociological Methodology* 17:105-129.
Saris W. E. and A. Satorra. 1985. "Power of the likelihood ratio test in covariance structure analysis." *Psychometrika* 50:83-90.
Saris W. E. and H. Stronkhorst. 1984. *Causal modeling in non-experimental research: An introduction to the LISREL approach*. Amsterdam: Sociometric Research Foundation, Amsterdam (SRF).
Satorra A.: 1992, "Asymptotic robust inferences in the analysis of mean and covariance Structures." *Sociological Methodology* 22:249–278.
Satorra A. and P. M. Bentler. 1990. "Model conditions for asymptotic robustness in the analysis of linear relations." *Computational Statistics & Data Analysis* 10:235-249.
Scherpenzeel A. C. and W. E. Saris. 1997. "The validity and reliability of survey items: a meta-analysis of MTMM Studies." *Sociological Methods and Research* 25(3):341-83.
Scherpenzeel A. C. 1995a. "Meta-analysis of a European comparative study." In: W. E. Saris and Á. Münnich. *The multitrait multimethod approach to evaluate measurement instruments*. Budapest: Eötvös University Press. Pp. 225-242.
Scherpenzeel A. C. 1995b. *A question of quality. Evaluating survey items by multitrait-multimethod studies*. Leidschendam: KPN Research. Thesis University of Amsterdam.
Schuman H. and S. Presser. 1981. *Questions and answers in attitude surveys: Experiments on question form, wording and context.* New York: Wiley.
Schwarz N., R. M. Groves, H. Schuman. 1998. "Survey Methods." In: D. Gilbert, S. Fiske, and G. Lindzey. *Handbook of Social Psychology*. 4th ed., Vol. 1, pp143-179. New York: McGraw-Hill.
Schwarz N., B. Knauper, H-J Hippler, E. Noelle-Neumann, and L. Clark. 1991. "Rating scales: numeric values may change the meaning of scale labels." *Public Opinion Quarterly*, Vol. 55 (4): 570-82.
Smith E. R. 1998. "Mental Representation and Memory." In: D. Gilbert, S. Fiske, and G. Lindzey. *Handbook of Social Psychology*. 4th ed., Vol. 1, pp391-445. New York: McGraw-Hill.
Sudman S., N. M. Bradburn, N. Schwarz. 1996. *Thinking about answers*. San Francisco: Jossey-Bass.
Tourangeau R., L. J. Rips, K. Rasinski. 2000. *The Psychology of the Survey Response*. Cambridge University Press.
Tourangeau R. and K. Rasinski. 1988. Cognitive processes underlying context effects in

attitude measurement.” *Psychological Bulletin*. 103:299-314.
Veld, van der W. M. and W. E. Saris. 2004. “Separation of reliability, validity, stability, and opinion crystallization” In: W. E. Saris and P. M. Sniderman. *Studies in Public Opinion: Gauging Attitudes, Nonattitudes, Measurement Error, and Change*. Princeton, N.J.: Princeton University Press.
Veld, van der W. M. and W. E. Saris. 2000. "Problems in Russian opinion polls: Memory effects, acquiescence and awareness." *Kwantitatieve Methoden*. 64:87-16.
Veld, van der W. M., W. E. Saris, and I. N. Gallhofer. 2001. “Prediction of quality of survey items before data are collected.” In: J. Blasius, J. Hox, et al. *Proceedings of the Fifth International Conference on Logic and Methodology*. 2nd ed. Leverkusen (Germany): Verlag Leske + Budrich.
Van Meurs, A. and W. E. Saris. 1995. “Memory effects in MTMM studies.” In W. E. Saris and Á. Münnich. *The Multitrait Multimethod approach to Evaluate Measurement Instruments*. Budapest (Hungary): Eötvös University Press. Pp. 89-102.
Von Neumann J. and O. Morgenstern. 1947. *Theory of Games and Economic Behavior.* Princeton, NJ: Princeton University Press.
Wegner D. M. and J. A. Bargh. 1998. “Control and Automaticity in Social Life.” In: D. Gilbert, S. Fiske, and G. Lindzey. *Handbook of Social Psychology*. 4th ed., Vol. 1, pp446-496. New York: McGraw-Hill.
Wheaton B., B. Muthen, D. F. Alwin, G. F. Summers. 1977. “Assessing Reliability and Stability in Panel Models.” *Sociological Methodology* 8:84-136.
Wiley D. E. and J. A. Wiley. 1970. "The estimation of measurement error in panel data." *American Sociological Review* 35:112-17.
Wilson T. D. and S. D. Hodges. 1991. “Attitudes as temporary constructions.” In: A. Tesser and L. Martin (eds). *The Construction of Social Judgment*. Hillsdale, N.J.: Erlbaum.
Zaller J. R. 1992. *The Nature and Origins of Mass Opinion.* Cambridge: Cambridge University Press.
Zaller J. R. and S. Feldman. 1992. "A simple theory of the survey response: Answering questions versus revealing preferences." *American Journal of Political Science* 36(3):579-16.

APPENDIX A

Description and explanation of the sample size of the datasets in this study

Table B1: Sample size for each *question* and topic after listwise deletion.

question	**n**
Proud Russian	264
Minority education	223
Minority medical access	217
Satisfaction housing	1119
Satisfaction contacts	1074
Disappointing people	319
Influence over life	310

The table above shows the sample sizes, after listwise deletion, for *questions* we have analyzed. The number of respondents that participated in panel wave 3, 4, and 7 from which our data originate are respectively 2273, 2073 and 1618. The interviews were administered in respectively autumn 1995, spring 1997, and autumn 1999. So, between each interview there is approximately 18 months. For the variables 'Proud Russian', 'Minority education', 'Minority medical access', 'Disappointing people', and 'Influence over life' there are relatively few cases compared to the number of respondents that participated in the three waves. This has several reasons. First of all and most important, these topics were part of methodological experiments that have been conducted within random subgroups of the respondents that completed wave 2 (N=2807). Respondents were randomly assigned to the different groups that were asked different survey item*s*. This resulted in subgroups of approximately 600 respondents before wave 3 was administered. Second, only those respondents who said they had the Russian nationality were allowed to answer the survey items. Third, also item non-response was also an important factor, due to the listwise deletion procedure, that resulted in a loss of cases. Finally, the subgroups diminished due to panel attrition.

One may be worried about the small sample sizes for the analyses. If item non-response and panel attrition are somehow connected to the topics under investigation it will bias our findings. Unfortunately there is no good way to test whether this is the case. However, the relations between the different variables under investigation do not change across time, thus within each wave they are approximately the same. This could be an indication that dropout and item non-response is not related to the variables under investigation. Note, however, that this is no real proof, even if one corrects for measurement error. This is because it might be that the 'true' relations have changed in the population. Furthermore, we have tried some imputation methods (MCAR, MAR) but that did not have any effect on our results, therefore we continued with the data at hand.

i The subscript on I denotes individual elements in that set.

ii The example presented is related to social contacts. One might also think about judging the current minister of internal affairs. Unique evaluations might be current events, enduring evaluations might be his party affiliation, and new evaluations might be some of the unique evaluations that have a large impact. One should that it is more difficult to use this example as an explanation of the model. The reason for that is that what is stable for one person might be unique for the other. This is no problem in the model, but it is a problem in an example because everybody could find reasons to disagree with it. Hence we favor use the simple example of social contacts.

iii The simple formulation is chosen but we are aware that we should have incorporated the systematic effect of the measurement method. Consequentially, a better formulation would be R = O + m + e. However, that would complicate the story to follow a lot. This however means that some of our assumptions are wrong, So, by chosing for the simple formulation, we now that some assumptions we make are wrong. On the other hand, that will only be so if m is larger than 0. Thus that is an extra assumption: m=0, that however is in contradiction with the example presented earlier in the paragraph.

iv The assumptions made are the following:

$$\mathrm{Cov}(N'_{t},N'_{t+x}) = \mathrm{Cov}(S^{*}_{t-x},N'_{t}) = \mathrm{Cov}(U'_{t},N'_{t}) = \mathrm{Cov}(U'_{t+x},N'_{t}) = 0$$
$$\mathrm{Cov}(e'_{tm},N'_{t}) = \mathrm{Cov}(e'_{t+xm},N'_{t}) = \mathrm{Cov}(U'_{t},U'_{t+x}) = \mathrm{Cov}(S^{*}_{t-x},U'_{t}) = 0$$
$$\mathrm{Cov}(S^{*}_{t},U'_{t}) = \mathrm{Cov}(O^{*}_{t+x},U'_{t}) = \mathrm{Cov}(e'_{tm},U'_{t}) = \mathrm{Cov}(e'_{t+xm},U'_{t}) = 0$$
$$\mathrm{Cov}(e'_{tm},e'_{t+xm}) = \mathrm{Cov}(S^{*}_{t},e'_{tm}) = \mathrm{Cov}(S^{*}_{t-x},e'_{tm}) = \mathrm{Cov}(O^{*}_{t},e'_{tm}) = \mathrm{Cov}(O^{*}_{t-x},e'_{tm}) = 0$$
$$\mathrm{Mean}(e'_{tm}) = 0$$

v A design with repeated observations within the wave can lead to a practical problem concerning the independence of the repeated observations. That is when memory effects play a role in the second observation, then the observations are said to be dependent. To prevent these memory effects one should make the time between the repeated observations as long as possible so that the previous response cannot be remembered. However, the time between the observations cannot be too long, because then the opinion (O^{*}_{t}) may change. Van Meurs and Saris (1995) have shown that memory effects are virtually gone after 20 minutes if similar questions have been asked in between the two observations and the respondents have no extreme opinions. In this study the first measure is always observed at the very beginning of the interview and the second observation at the end. The average length of the questionnaires that are used in this study was approximately 50 minutes. It is therefore fair to assume that problems due to memory effects will not occur.

vi Note that we do not mean low correlations in a general sense. Here we specifically refer to correlations of repeated measures of the same variable.

vii By definition, the crystallization is the complement of the variance of the outcome variable U'_{t} that represents the unique considerations in the model. This follows from the formula: $c^{2}_{t} = 1\text{-var}(U'_{t})$.

viii For comparison, we have also compared Dutch panel data with the RUSSET panel data (Van der Veld and Saris, 2000), and found that there were differences, but those differences were smaller than with the NES panel data.

ix The reader should know that the identification issues discussed here are standard issues in the identification of the quasi-simplex.

x The misspecification refers to the fact that there is no direct relation between the first stable opinion and either the third stable opinion or new opinion.

[xi] For this model the shared variance is the correlation because the coefficients are standardized.