Recall Accuracy and its Determinants

Eijk, C. van der; Niemöller, B.

Acta Politica
tijdschrift voor politicologie

Kernredactie
drs. J. Th. J. van den Berg, drs. H. W. Blom, dr. B. J. S. Hoetjes, drs. K. Koch (secretaris), drs. O. Schmidt, prof. dr. J. K. De Vree

Redactieraad
Jhr. drs. G. van Benthem van de Bergh, prof. dr. H. R. van Gunsteren, drs. R. P. van den Helm, prof. dr. A. Hoogerwerf, drs. J. A. M. Klaver, dr. A. E. van Niekerk, dr. J. van Putten, dr. U. Rosenthal, mr. dr. I. Th. M. Snellen, dr. J. A. A. Thomassen

Bijdragen
en andere mededelingen voor de redactie zende men aan de redactie-secretaris, drs. K. Koch, p.a. Volkenrechtelijk Instituut van de Rijksuniversiteit Utrecht, Janskerkhof 16, 3512 BM Utrecht (030 - 31 87 49); de kopij dient persklaar te worden ingezonden, in zesvoud en in machineschrift

Boeken ter recensie
en boekbesprekingen zende men aan de boekenredacteur, drs. H. W. Blom, Erasmus Universiteit H 2-16, Burgemeester Oudlaan 50, 3062 PA Rotterdam; terugzending van ongevraagde recensie-exemplaren kan niet plaatsvinden

Administratie
Voor abonnementen en advertenties richte men zich tot de uitgever, Boompers boeken- en tijdschriftenuitgeverij bv, Postbus 58, 7940 AB Meppel (05220/ 54306); postrekening 2756509

Acta Politica verschijnt driemaandelijks; de abonnementsprijs bedraagt *f* 62,—; voor instellingen *f* 86,—; voor studenten *f* 47,— (maximaal gedurende vijf achtereenvolgende jaren).
Voor het buitenland gelden aangepaste tarieven die op aanvraag te verkrijgen zijn.
Losse nummers *f* 19,50
Advertentietarieven: 1/1 pagina *f* 300,—; 1/2 pagina *f* 170.—.

Indien men het abonnementsgeld gireert voor een ander gelieve men op het girostrookje duidelijk naam en adres van de abonnee te vermelden, ter vermijding van dubbele incasso

Recall Accuracy and its Determinants *

C. van der Eijk and B. Niemöller

1 Introduction

Analyzing individual voting behavior has been for many years one of the more popular concerns of empirical political scientists. Since elections are a regularly recurring phenomenon, *changes* in voting behavior are also an important topic for study. Numerous publications have appeared on stable, changing and floating voters.[1] The influence of the findings of this kind of research on theories about all kinds of political phenomena can hardly be overstated. Conclusions about the dynamics of individual voting behavior have their ramifications in theories of system stability and performance, on our perceptions and interpretations of changing political needs, demands and issues, on our understanding of the determinants of a single voting decision by a voter, and on our perception of the potential for change in party support and resulting viable coalitions.

How are changes in individual voting behavior generally observed? Since the advent of modern, large scale survey sampling there is no need anymore to make inferences on the basis of aggregate election results[2], for we can tackle the problem directly for each individual in a sample. Changes in individual voting behavior can be measured by keeping track of respondents over the course of several elections, and asking them each time how they voted (panel study). Notwithstanding the straightforwardness of this approach, it has several drawbacks which explain its relatively infrequent use: prohibitive costs, the high loss of respondents (panel mortality) caused by the long period of time between elections, and the lack of generalizability to the electorate-at-large as the panel grows older (new cohorts entering the electorate are not included in an existing panel). Instead of using panels, researchers have resorted to measuring changes in voting behavior by asking respondents not only to report on their most recent choice, but also to recall what they did at one or more previous elections. Combining this information yields an easy

* The data analyzed in this report have been collected with financial aid of ZWO the Netherlands Organization for the Advancement of Pure Research), grants nrs. 43-11, 43-30 and 43-73.

classification in stable and changing voters. Most of what passes as 'common knowledge' about voting behavior and its stability has been based on data of this second kind.[3]

When we compare the panel- and recall method of assessing changes in voting choice, it is evident that a necessary condition for yielding equivalent results is an unbiased recall of past behavior. As far as we know, this assumption has rarely been seriously questioned in the framework of voting research[4], even though research on other phenomena suggests that recall data cannot readily be accepted as accurate.[5]

It is just this question – the accuracy of recall data on voting – which we challenge in this paper. For this purpose we are using data obtained from a panel of 509 respondents. Each person in this sample has been interviewed 5 times, covering 3 different elections for the second chamber of parliament in the Netherlands, i.e. the elections of 1971, 1972 and 1977.[6] After each of these elections our respondents have been asked whether or not they voted, and if so, for which party. At the same time they were asked whether, and if so in which way, they voted at previous elections. This dataset thus contains data on individual voting change of the panel type as well as of the recall type. This enables us to compare the results of these two approaches.[7] Figure 1 illustrates which data on voting behavior in the three elections are available for each of our respondents.

Figure 1: Information on voting behavior at different elections, available in our data[8]

		data collected in the year		
		1971	1972	1977
Information	1967	recall 71-67	—	—
concerns	1971	vote 71	recall 72-71	recall 77-71
election	1972	—	vote 72	recall 77-72
in the year	1977	—	—	vote 77

In section 2 of this paper we will show that the assumption of unbiased recall doesn't stand up to an empirical test. We will give an appraisal of the proportion of unjustified classifications of stability of voting behavior which would result if we were to rely solely on recall data. In section 3 we will investigate which factors are correlated with a faulty recall, in section 4 we will select the more important of these factors, and in section 5 we will combine these into a log-linear model, explaining the incidence of faulty recall.

2 Faulty Recall-delineating the Nature and Scope of the Problem

2.1 Aggregate comparison of different reports on voting in 1971

As illustrated in figure 1, our respondents have been asked to report their 1971 voting behavior at three different moments: in 1971, right after the elections, in 1972 and in 1977. Obviously, if recall data are accurate, the distribution of answers should be identical, or only slightly different on these three occasions. This doesn't turn out to be the case as can be seen in table 1.[9]

Table 1: Marginal distribution of voting behavior, as reported in 1971 (vote 71), 1972 (recall 72-71), and 1977 (recall 77-71)

	vote 71	recall 72-71	recall 77-71	number of people giving the same answer all 3 times
didn't vote	43	36	25	6
not yet entitled to vote	—	1	—	—
CPN	10	13	7	2
PSP	3	3	4	1
PvdA	111	116	138	78
PPR	7	4	9	1
D'66	37	44	18	14
DS'70	26	17	8	4
KVP	118	117	118	84
ARP	31	34	35	23
CHU	31	22	25	17
VVD	40	57	57	31
BP	3	8	4	—
GPV	7	5	6	4
SGP	7	6	5	4
RKPN	1	—	—	—
other parties	9	4	1	—
no answer/don't know	25	22	49	—
total	509	509	509	269 (53% of 509)

Given the differences between the columns of table 1 it seems justified to question the reliability of recall data. Apart from this general conclusion some specific remarks can be made:

* the number of people reporting not to have voted in the 1971 election drops sharply between 1971 and 1977;

* especially in 1977 the number of people unable to give an answer to the question as to what they did in 1971 is relatively large (almost 10%);

* with the exception of the KVP, the number of people reporting that they

voted for a certain party in 1971 changes considerably over time. This phenomenon can be observed for all kinds of parties: left wing, center, right wing, confessional and secular;[10]

* the last column of table 1 indicates the number of people which gave the same answer all 3 times. Only 53% of our sample is completely consistent in its recall. Evidently, the marginal distributions suggest greater stability than exists at the individual level.[11]

2.2 Individual comparison of different reports on voting in 1971

An adequate comparison between original response and recall necessarily involves comparisons for each respondent separately. Such comparisons can be made by crosstabulating vote 71 with recall 72-71 and recall 77-71 respectively. The results of such analyses are reported in tables 2 and 3. Consistent responses appear in the main diagonal of these tables, all other cells signify inconsistencies; we have to be aware of the fact that consistency is a necessary, but not sufficient condition for accurate recall.[12]

In table 2 we find 151 respondents (i.e. 30%) outside the diagonal; in table 3 we find 200 (39%) inconsistent responses. Checking the plausibility of these figures by considering the 1972 election, and comparing vote 72 with recall 77-72, leads to 164 cases which are inconsistent in their response (32%). The number of inconsistencies seems to increase monotonely with increasing time span here as well. For the moment we will not indulge in any analysis of the structure of the inconsistencies which appear in tables 2 and 3. Our first aim is to explain the incidence of recall inconsistencies and consequently all inconsistencies are equivalent to us. The only thing worth mentioning is that the stability in the number of KVP responses (see table 1) seems to be a mere fluke. Not a single party seems to be excepted from the phenomenon that many of its voters do not remember having voted for it.

The data so far lead to the conclusion that in 30 to 40% of the cases a recall question yields different answers from an equivalent question asked shortly after the relevant election. This proportion is the more striking when we bear in mind that our respondents have been interviewed five times on voting behavior and other political topics. If being interviewed repeatedly has any effect on response, it definitely means that the chance on inconsistencies will not be greater than for 'normal' respondents in election surveys who are interviewed only once.

It might be argued that our account above is a bit on the pessimistic side, as it includes in the inconsistents those voters who are unable to recall what party they voted for, and honestly report that they don't know. Taking this into account we created four new variables, indicating the consistency of

Table 2: Voting behavior in 1971, as reported in 1971 (vote 71) and in 1972 (recall 72-71)

vote 1971	didn't vote	too young	CPN	PSP	PvdA	PPR	D66	DS70	KVP	ARP	CHU	VVD	BP	GPV	SGP	RKPN	other	NA/DK	total
didn't vote	17	—	1	11	—	3	3	1	2	—	1	—	1	—	1	—	—	2	36
too young	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
CPN	1	—	7	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	13
PSP	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
PvdA	11	—	1	88	—	2	2	3	2	2	2	1	—	—	—	—	—	4	116
PPR	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	1	—	4
D66	3	—	—	—	—	2	28	—	2	2	1	—	—	—	—	—	—	1	44
DS70	—	—	—	—	—	—	—	13	—	—	—	—	—	—	—	—	—	—	17
KVP	1	—	—	—	—	—	2	5	99	2	—	—	—	—	—	—	—	—	117
ARP	—	—	—	—	—	—	—	—	2	26	—	—	—	—	—	—	—	—	34
CHU	—	—	—	—	—	—	—	—	—	—	21	—	—	—	—	—	—	—	22
VVD	—	—	—	—	—	—	1	—	—	—	2	36	—	—	—	—	—	—	57
BP	1	—	—	—	—	—	—	—	—	—	—	—	2	1	—	—	—	—	8
GPV	—	—	—	—	—	—	—	—	—	—	—	—	—	5	—	—	—	—	5
SGP	1	—	—	—	—	—	—	—	—	—	—	—	—	—	6	—	—	—	6
RKPN	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
other	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3	1	4
NA/DK	1	—	—	1	—	2	2	1	4	1	1	2	—	1	—	—	1	4	22
total	43	0	10	3	111	7	37	26	118	31	31	40	3	7	7	1	9	25	509

Table 3: Voting behavior in 1971, as reported in 1971 (vote 71), and in 1977 (recall 77-71)

vote 1971		didn't vote	CPN	PSP	PvdA	PPR	D66	DS70	KVP	ARP	CHU	VVD	BP	GPV	SGP	RKPN	other	NA/DK	total
recall 77-71	didn't vote	11	1	—	3	—	3	4	1	—	—	—	—	—	—	—	—	—	25
	CPN	2	2	—	1	2	—	—	—	—	—	—	—	—	—	—	—	—	7
	PSP	—	1	1	1	—	1	—	—	—	—	—	—	—	—	—	—	—	4
	PvdA	9	5	1	88	—	11	6	4	1	2	—	—	—	—	—	—	8	138
	PPR	1	—	1	2	3	—	—	—	1	1	—	—	—	—	—	—	—	9
	D66	—	—	—	—	—	14	1	1	1	—	—	1	—	—	—	—	—	18
	DS70	1	—	—	—	—	—	4	—	—	—	2	—	—	—	—	—	1	8
	KVP	6	—	—	4	1	2	4	94	—	—	2	—	—	—	—	1	4	118
	ARP	1	—	—	1	1	—	—	1	25	3	1	—	—	—	—	—	1	35
	CHU	—	—	—	1	—	—	1	1	—	20	—	1	—	—	—	—	—	25
	VVD	1	—	—	2	—	2	5	5	1	2	32	1	—	5	—	—	4	57
	BP	—	—	—	1	—	1	—	—	—	—	1	—	1	—	—	—	—	4
	GPV	1	—	—	—	—	—	—	—	—	—	—	—	4	—	—	—	—	6
	SGP	—	—	—	—	—	—	—	—	—	—	—	—	—	5	—	—	—	5
	RKPN	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	other	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	1
	NA/DK	10	1	—	6	—	3	1	11	2	3	1	—	1	1	—	3	6	49
	total	43	10	3	111	7	37	26	118	31	31	40	3	7	7	1	9	25	509

recall with the original response, and setting apart those cases as missing who one or more times responded with 'don't know', 'no answer', or 'don't remember'. These variables will be referred to as *Recall 2-1, Recall 3-1, RECALL, and Recall 3-2*. The first two refer to voting behavior in 1971, recalled in 1972 and 1977 respectively (1=1971; 2=1972; 3=1977). RECALL combines these two variables: only those respondents are classified as consistent who in 1972 *and* in 1977 report the same behavior as they did in 1971. The variable Recall 3-2 refers to voting behavior in 1972, recalled in 1977.[13]

Table 4 gives the distribution of our cases over these new variables. It is clearly visible that the number of inconsistent responses increases as the time span to be bridged by the recall increases. It also appears that giving a consistent recall at a certain time doesn't insure a consistent recall at another time: there are only 53% of the cases who on al 3 occasions report the same behavior at the 1971 election.

Table 4: Consistency of recall with originally reported behavior

	election recalled took place in 1971 Recall 2-1	1971 Recall 3-1	1971 RECALL	1972 Recall 3-2
consistent	354 (70%)	303 (60%)	269 (53%)	343 (67%)
inconsistent	112 (22%)	138 (27%)	157 (31%)*	135 (27%)
missing data	43 (8%)	68 (13%)	83 (16%)	31 (6%)
total	509 (100%)	509 (100%)	509 (100%)	509 (100%)
time between behavior and recall (approx.)	18 months	70 months	inap.	52 months

* In the construction of RECALL 3 different variables have been used: vote 71, recall 72-71 and recall 77-71. Of the 157 inconsistent cases 135 gave twice (out of 3 times) the same answer; the remaining 22 cases gave a different answer at every time.

We conclude that the *use of recall data is at best hazardous. Of the cases considered here we find at best 70%, and at worst 53% in agreement with their original report of voting behavior.* Only a small proportion of the remaining cases consist of 'don't knows' etc., most of them are per se faulty.

2.3 Stable and changing voters: comparing the panel- and recall approaches

In section 1 we mentioned that the stability of individual voting behavior can be observed in two ways either: by making use of repeated interviews with a certain group of respondents, or by comparing the current vote with recall of

voting behavior in the past. In sections 2.1 and 2.2 we observed that recall data are often inconsistent with the original report, right after the elections. Do these inconsistencies effect an analysis of stable and changing voters? The answer to this question is not obvious, as it is very well possible that many of the individual recall inconsistencies cancel each other out so that the aggregate picture of voting stability and change could be similar using either panel or recall data.

Before we can proceed we have to devise a means of classifying respondents as stable or changing voters, and we need such a classification for both panel data and recall data. There are 3 elections which are of interest to us, i.e. those of 1971, 1972 and 1977. Voting stability or change recorded by panel data is indicated by variables named *vote 2-1, vote 3-1* and *vote 3-2,* the numbers indicating the elections being compared (1=1971, 2=1972 and 3=1977). Corresponding variables based on recall data are termed *float 2-1, float 3-1* and *float 3-2.*[14] Figures 2 a, b and c illustrate the relationship between these new variables and the ones mentioned in figure 1.

A first comparison between the results obtained from panel and recall data can be made on the distribution of stable and changing voters. Table 5 displays the relevant information.

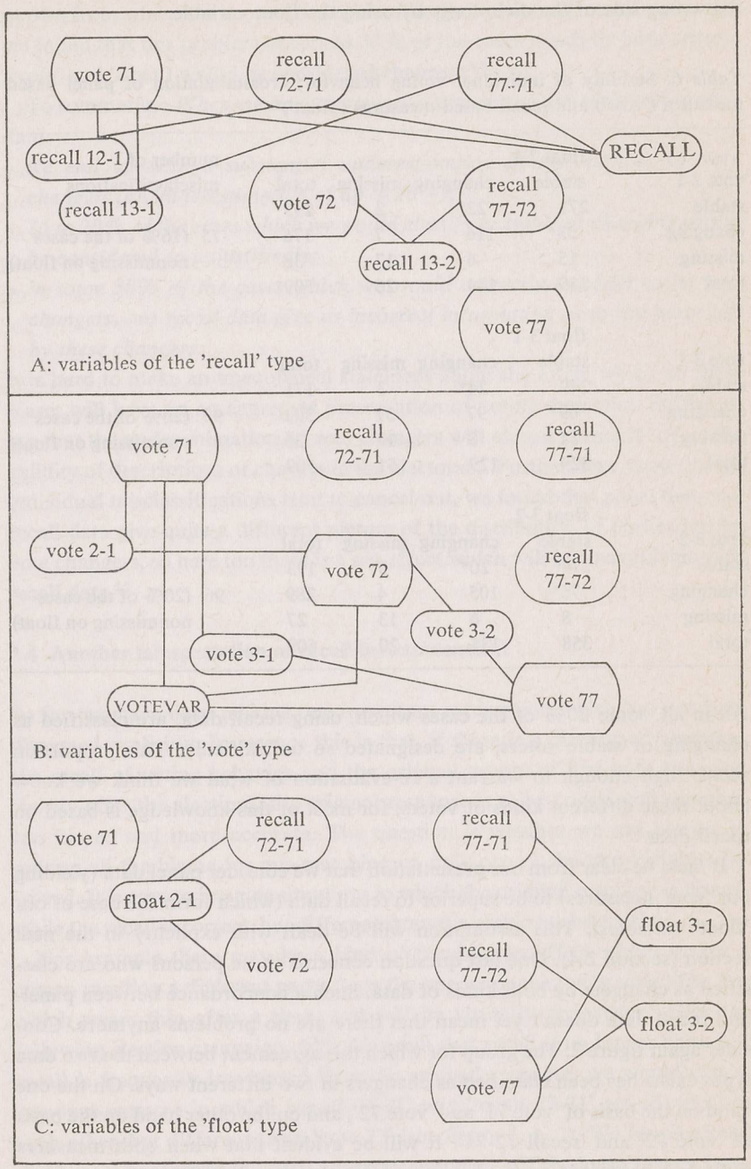
Table 5: Stability of individual voting behavior, measured with panel data (vote 2-1, vote 3-1, and vote 3-2), and with recall data (float 2-1, float 3-1 and float 3-2)

	vote 2-1	float 2-1	vote 3-1	float 3-1	vote 3-2	float 3-2
stable*	295(58%)	339(67%)	193(38%)	332(65%)	297(58%)	361(71%)
changing	176(35%)	144(28%)	280(55%)	126(25%)	185(36%)	128(25%)
misisng data**	38(7%)	26(5%)	36(7%)	51(10%)	27(5%)	20(4%)
total	509	509	509	509	509	509

* respondents who didn't vote at either election are considered stable
** one or more response 'don't know', ,no answer', or 'don't remember'

The two ways of measuring vote stability lead to clearly different results. The use of recall data (predominant in election studies) leads to a serious underestimation of the number of changers. Usually we are not only interested in the question of how many vote changers there are, but also how they may be characterized, and which parties they leave or join. The use of recall data for such an analysis entails additional hazards quite apart from the underestimation of the number of changers. The best that we might expect is that those respondents which are designated as changers using the recall method, are designated correctly.

Figure 2: The construction of the 'recall', 'vote' and 'float' variables



This however, is not the case. In table 6 we have crosstabulated the 'vote' and 'float' variables; the cells marked with an asterisk contain the cases put at the wrong side of the dichotomy by using the float variable.

Table 6: Stability of individual voting behavior; crosstabulation of panel based measures ('vote') and recall based measures ('float')

	float 2-1				number of
vote 2-1	stable	changing	missing	total	misclassifications
stable	271	22*	2	295	
changing	53*	116	7	176	75 (16% of the cases
missing	15	6	17	38	nonmissing on float)
total	339	144	26	509	
	float 3-1				
vote 3-1	stable	changing	missing	total	
stable	249	24*	3	171	
changing	69*	97	31	302	93 (20% of the cases
missing	11	8	17	36	nonmissing on float)
total	329	129	51	509	
	float 3-2				
vote 3-2	stable	changing	missing	total	
stable	274	20*	3	193	
changing	76*	105	4	289	96 (20% of the cases
missing	8	6	13	27	nonmissing on float)
total	358	131	20	509	

All in all, some 20% of the cases which, using recall data, are classified as changing or stable voters, are designated so unjustifiedly. This proportion seems high enough to warrant a re-evaluation of what we think we know about these different kinds of voters, for most of this knowledge is based on recall data.

It must be clear from our presentation that we consider panel data (yielding our 'vote' measures) to be superior to recall data (which form the base of our 'float' variables). This assumption will be dealt with explicitly in the next section (section 2.4). One last question concerns those persons who are classified as changers by both kinds of data. Such a concordance between panel- and recall data doesn't yet mean that there are no problems anymore. Consider again figure 2. The group for which this agreement between the two data types exists has been classified as changers in two different ways. On the one hand on the basis of 'vote 71' and 'vote 72', and on the other hand on the basis of 'vote 72' and 'recall 72-71'. It will be evident that when both measures indicate a change in voting behavior from the one election to the next, it is

still quite possible that 'vote 71' and 'recall 72-71' are not equivalent.[15] If that were the case the two measures would lead to a different description as to the parties from where the support for some other party comes from. Empirically we found that this problem occurs in 30% of the cases which by both criteria ('vote' and 'float') would be considered changers.[16]

To summarize: *When we investigate individual voting dynamics with recall data:*

- *we end up with a substantial underestimation of the number of vote changers (an underestimation of up to 20%);*
- *16 to 20% of the cases which we would classify as stable or changing would be considered so unjustifiedly;*
- *in some 30% of the cases which we would correctly consider to be vote changers, our recall data give us incorrect information as to the party left by these changers.*

It is hard to make an unequivocal statement about the effect which all these biases will have on an aggregate presentation of voting dynamics. Suffice to say that the underestimation of vote changers will almost certainly affect the validity of descriptions of charges in the electorate. Furthermore, even though individual misclassifications tend to cancel out, we found that panel data and recall data give quite a different picture of the distribution of parties left by vote changers, so here too there is a net effect which will be unavoidable with recall data.[17]

2.4 Another interpretation of recall inconsistencies?

So far, we have been making one important assumption which has not been discussed explicitly. In essence this is that, if there is a discrepancy between the recall of voting behavior and the original report of that vote (assessed shortly after the election), the original response is to be considered superior, less biased and more accurate. The question is whether we are correct in putting all the blame for mis-matching on only one of the two variables involved. It is easy to imagine situations in which the original response is biased, while the recall is correct, but different from the previous observation.

For example there may be a kind of bandwagon effect which makes a person mention a different party from the one he or she really voted for, but which wears thin after a time, and will be virtually nonexistent after the following election campaign, with the result that recall is no longer biased. If recall is frequently less biased than the original response, we would expect response-patterns in which 'recall 72-71' and 'recall 77-71' are identical to each other, but different from 'vote 71' (see figure 1, p. 2). We found a total of 43 cases conforming to such a response sequence. This doesn't mean how-

ever that all these cases necessarily indicate a recall which is more reliable than 'vote 71'. There are other ways in which such a pattern can originate. It could happen by chance, or for example by a recall biased towards a different party from the one the person really voted for, with bias persisting through to the time of the second recall. Nevertheless, we are convinced that there will be cases in which the recall is the better of the two responses, but we do not have any means of identifying these cases, or of estimating their number.

However, the possibility that in some cases recall yields better information about a person's behavior doesn't invalidate, or even handicap our research. We started by doubting the validity of the empirical basis of most research on the dynamics of individual voting behavior, i.e. those analyses which make use of recall data. In such analyses changers are indicated by differences between their current vote and recall of past vote. In cases where the report of current vote is not correct, these analyses are troubled by it as much as we are. All analyses of voting dynamics (panel analyses as well as recall analyses) must cope with the fact that a certain number of responses will not be 'true'. As panel analyses and recall analyses both use data which are collected right after the election, both are subject to inacurracies because of this, but we do not know of any way estimating the size of these inaccuracies with the data we have available. Some analyses of voting dynamics have to cope with the additional problem that they use recall data as well as reports on current preference; in a number of cases the recall will be faulty. Since this problem is in addition to the one just mentioned, there is no reason at all to assume that in general it alleviates the first problem, except in those cases where 'side information'[19] suggests that biasing factors distorting the original response (bandwagon effects etc.) have operated on a much larger scale in one election than in another. There is no indication that this is the case with the elections of 1971 and 1972.

Therefore, without implying that an immediate report yields flawless data, we definitely prefer such items over recall items. Furthermore, most of the factors which would lead to a faulty immediate report are general in character, and affect all kinds of data (including recall); whereas most of the factors biasing recall are specific, and do not affect other data (like the immediate report). Such factors, specific to recall, are numerous; a few obvious examples being:

1 gradual erosion of memory with the passage of time.
2 confusion caused by other elections which have taken place in the interval between the election being recalled and the time of recall. In our own case these may be provincial, regional, or municipal elections. Usually the national parties, as well as typical local groups and parties compete in these elections.

3 confusion caused by the disappearance of parties which were once voted for years ago. We think that in 1977 it is easier for a respondent to recall that he or she voted for the PvdA in 1971, than that one voted for 'Binding Rechts', a group nobody heard of anymore since their unsuccessful participation in the 1971 election.

Factors like these grow more important as time goes by, and are specific factors biasing recall items, over and above the general factors which cause deviations of measurements from a 'true score'. One final point which should be emphasized is that in the remainder of our analysis it does not really matter whether the original report or the recall is to be blamed for inconsistencies between them. In the next sections we will try to explain inconsistencies and certain patterns of inconsistent response. At least for the statistical part of these analyses it is not relevant which of the 3 responses on the 1971 vote 'causes' the inconsistency: vote 71, recall 72-71 or recall 77-71. One thing we know for sure: the nature of an election ensures that the notion of a 'true score' for the respondent's behavior in the election is a viable one. Three identical responses on the three relevant questions are a necessary but not a sufficient condition for agreement between the observed score and this true score, at the same time any deviation of this pattern is a sufficient condition for unreability in the response. It is this asymmetry between necessary and sufficient conditions, together with the other arguments mentioned in sections 2.1 through 2.4 which make us believe that we have not painted an overpessimistic picture of the problems entailed in the use of recall data.

3 Correlates of Recall Inconsistencies

3.1 Introduction

Our aims in this paper are to assess the reliability of recall data, and to gain some insights in the determinants of recall inconsistencies. In this section we will begin by pursuing the second question, investigate the relation between inconsistent recall and a number of variables which can be conceived as potential 'causes'. The central variables in this section have already been introduced in section 2.2, and are *recal 2-1* – (in)consistency between vote 71 and recall 72-71 –; *recall 3-1* – (in)consistency between vote 71 and recall 77-71 –; and *RECALL*, in which the (in)consistency of vote 71, recall 72-71 and recall 77-71 is measured. With the exception of missing values recall 2-1 and recall 3-1 divide respondents in 'consistent' and 'inconsistent'. The variable RECALL has three values: consistent; inconsistents who twice out of three times gave the same answer (abbreviated to '2 answers'); and inconsistents who gave a different answer on each occasion (abbreviated to '3 answers').[20]

The number of variables which can be used to distinguish consistents and

inconsistents using these 3 measures is only limited by one's creativity. Without pretending to have taken into account all possible factors, we've tried to cover the most plausible ones. Some of these are relatively broad and can be divided in more specific measures, for others the data allow for only one or at most very few operationalisations. We will start with explaining the general factors. More detail and the results of our analyses with those factors will follow in the subsections 3.2 through 3.7. Theoretically plausible causes for inconsistencies are:

(a) *technical problems in data preparation.* In any collection of data certain pieces of information are in a technical sense wrong. Tired interviewers sometimes mismark answers, such mistakes can also happen during the tedious and uninspiring tasks of coding and keypunching. These mistakes originate from ennui, and we can expect that the data of certain cases will become garbled, and contain, on the average, more wrong codes than those which have been interviewed, coded and keypunched by motivated and fit personnel. To the extent that our data contain such 'garbled' cases, we can expect them also to contain a disproportionate share of recall inconsistencies. This possibility will be investigated further in section 3.2.

(b) *characteristics of the interview situation.* The respondents have been interviewed 5 times in total. Probably this experience hasn't been equally pleasant for all, nor will it have been identical every time. In cases where the respondent hasn't liked the situation, and/or where the interaction between interviewer and interviewee didn't contain rapport, we feel that the response will be subject to a higher chance of being incorrect than in other cases, and thus give rise to inconsistencies.

(c) *social characteristics of the respondent.* This heading covers a rather heterogeneous group of variables, which are sometimes also referred to as background variables. None of these are expected to cause inconsistencies by themselves, but through their relationship with variables in (d) and (e) below, or because some of them (most notably age and education) are known to be related to cognitive capabilities and memory skills, both necessary in recalling past behavior.

(d) *indicators of political involvement.* Operating under the idea that one's memory works best in those areas which are felt to be important, we will look at a number of variables which we assume to be direct or indirect measures of the significance of the political realm to an individual.

(e) *participation in the electoral process.* This factor may very well belong to political involvement in general, but we've chosen to put it under its own heading because it is much more directly related to recall of *voting* behavior than general political involvement.

(f) *changes in voting behavior over consecutive elections.* Our data cover

three different, but comparable elections: 1971, 1972 and 1977. With the variables 'vote 71', 'vote 72' and 'vote 77', we have information on the respondent's voting behavior, as reported shortly after each of the relevant elections. On the basis of these (panel)data we can divide our respondents into stable and changing voters. If voting behavior changes through time, the cognitive tasks involved in recalling one's past behavior get more complex, thus increasing the chance on inconsistencies. Besides this, to acknowledge having changed allegiance may create cognitive dissonance for some voters, which can be avoided by distorted recall. Given these two reasons we will investigate the relation between recall inconsistencies and changes in voting behavior.

3.2 Potential causes: technical problems in data preparation

In section 3.1 we already explained that recall inconsistencies as we observe them in our data, do not necessarily originate in the respondent. Interviewers, coders and keypunch operators do make errors which can result in inconsistencies. This gives rise to two questions: *First* we want an estimate of the proportion of this kind of error. Should it be as prevalent as the number of recall inconsistencies observed, one could hypothesize that all these are caused by 'technical' problems. We doubt this to be the case, but we are still interested in the extent of this error, this could serve as a criterion for deciding when to stop interpreting inconsistencies in other ways.

A *second* question is whether or not cases which show 'technical' error, are also characterized by more recall inconsistencies than one could expect on the basis of chance. Such a co-occurance could be interpreted as the result of interviewers, coders or keypunch operators who were tired and bored at the time of processing these particular cases. We could then conclude that at least part of the inconsistencies can be explained by errors in datapreparation. To construct a measure for the technical unreliability we described above, one must have at one's disposal repeated measurements of a stable characteristic, the measurement of which is not vulnerable to distortions by changes in knowledge, perceptions, or moods of the respondent. We believe to have found such a characteristic with the year of birth of our respondents. At 5 different times they have been asked in which year they were born.

The answers have been coded without collapsing into age categories. On the basis of 4 of these answers we identified those cases in which the information didn't match.[21] It will be evident that we operate here under the assumption that all respondents accurately know their year of birth, and that they've felt no need to cheat in reporting it. If this assumption is justified inconsistencies between the 4 measurements of year of birth must be due to other causes than the respondent. In 28 cases (5.5% of our sample) we found differences in the

repeated measurement of year of birth. We consider this 5.5% as the highest level of accuracy to be obtained with our data.[22]

This finding immediate leads to one conclusion: the amount of recall inconsistencies (22 to 31%, see table 4) cannot plausibly be explained by 'technical' unreliability alone. Our second question concerns the relation between inconsistencies in reported year of bith and recall inconsistencies. It turns out that there is no association at all between year of birth inconsistencies and any of our recall variables (recall 2-1, recall 3-1 and RECALL).[23] This means that recall inconsistencies cannot be explained by mistakes and errors in data-preparation. Also, the interpretation of technical unreliability as a random phenomenon gains credibility.

3.3 Potential causes: the interview situation

Survey data are obtained through the interaction between an interviewer and an interviewee. To some (unknown) extent the quality of the data is dependent upon the quality of this interaction, and especially upon that aspect of it which encourages the respondent to answer the questions truthfully. One could suppose that ,if 'rapport' is missing, both partners in the interaction would be aware of it. This is relevant in so far that we can make use of two interviewer-ratings which were completed after the interview was finished. In one, the interviewer is asked to rate the cooperation of the respondent, in the other his or her estimate of the reliability (truthfulness) of the answers given.

We don't know exactly what an interviewer has in mind when completing these ratings, we only know that both ratings are very heavily skewed to the positive side. This leads us to suspect that in those few cases where a negative rating has been given, the rapport in the interaction must have been definitely lacking. It might very well be that just these situations give rise to recall inconsistencies, first because the motivation of the respondent to give correct answers will be probably low, and secondly because the interviewer may make more recording mistakes in such situations.

However plausible this argument may sound, it does not stand up empirically. Recall inconsistencies and interviewer ratings of reliability and cooperation are not significantly interrelated.[24] Of course this is no conclusive evidence refutung the general hypothesis that interaction during the interview can give rise to bad data, or, in our case, to recall inconsistencies. However, in so far as we have information on this aspect, it does not appear to be of any great importance.[25]

3.4 Potential causes: social characteristics of the respondent

3.4.1 Age. The ability to memorize is known to decline with increasing age[26], and one might expect older people to be preponderant amongst the inconsistents. Other factors however could be expected to give rise to the opposite pattern, for example importance of politics to a person seems to increase with age.[27] Furthermore, stable partisan attachments have been shown to be positively related to age, this could help to structure the cognitive elements in the recall task, thereby increasing consistency.

As will be clear from the data presented in table 7 there appears to be a strong monotone relation between consistency and age; this table displays the relation with the overall RECALL variable, the variables recall 2-1 and recall 3-1 yielded similar results.[28]

Table 7: RECALL by age. Cell entries are frequencies; between brackets % over the nonmissing cases (columnwise)

RECALL							
Age in 71	17-25	26-35	36-45	46-55	56-65	65+	total
consistent	20 (40)	58 (53)	50 (67)	62 (67)	42 (81)	37 (80)	269
'2 answers'	24 (48)	43 (39)	20 (26)	29 (31)	10 (19)	9 (20)	135
'3 answers'	6 (12)	9 (8)	5 (7)	2 (2)	— (—)	— (—)	22
missing data	15	19	17	14	8	10	83
total	65	129	92	107	60	56	509

$tau_b = -.24$ (only nonmissing cases included)

At this stage we will not attempt to interpret this association between age and recall consistency. We suspect that there is no direct causal link between the two variables but that there may be intervening variables which explain the relation.

3.4.2 Sex. A priori there is no reason to expect any difference in recall performance between men and women. There may however be indirect relationships linking sex and recall. Women on the average are less interested in political affairs at the level of party-politics[29] and they are still (marginally) less educated.

Whatever relations sex may have with other variables however, none of the recall variables (RECALL, recall 2-1 or recall 3-1) relates to it in a statistically significant way.[30]

3.4.3 Education. There are several reasons why we expect the level of educa-

tion to be related to recall consistency. First of all, education 'per se' increases intellectual and cognitive skills, thereby enhancing the ability to understand the questions correctly and the ability to perform the recall task. Furthermore, there is usually a strong relationship between the level of education and the importance attached to politics.

Surprisingly however, there is no statistically significant association between the two variables. The only outcome of any interest is that the percentage of missing values decreases slightly with increasing education, which is most clearly in relation to the variable RECALL, and slightly less pronounced in recall 2-1 and recall 3-1. We can conclude that education has a slight effect on the percentage of missing data, but for those cases which contain complete information, no significant relation with recall performance emerges.[31]

3.4.4 Class identification. Selft-rated social class appears traditionally as one of the most potent correlates of political and social attitudes and behavior in the Netherlands. This in itself is more than sufficient reason for us to investigate the relationship between recall and social class. The categorization on the class variable is based on the following question: 'It's not unusual to divide society into various social classes. Do you think you belong to a particular social class?', the respondents were presented with 5 precoded answers.[32] Little is known about the detailed mechanisms which determine the response to this question, we only know that in general the upper classes are more involved in party politics. We observed a slight tendency in the expected direction – i.e. that the upper classes are more consistent in their recall – but the relationships do not approach statistical significance. Furthermore there is a slight tendency for the data from the lower classes to contain more missing values.[33]

3.4.5 Income and urbanization. Empirically neither income nor urbanization are related to recall behavior in any systematic or statistically significant way.[34] Nevertheless, the possibility that these variables could be related with determinants of recall made it worthwile investigating.

3.5 Indicators of political involvement

The variables which we will investigate in this section are all assumed to reflect in some way feelings on the importance of politics.[35] Our central hypothesis is that persons to whom politics are important will be more consistent in their recall than those who are not interested in the subject.

3.5.1 Political interest. The variable under consideration here is the score on a 3-item scale. The items included in the scale refer to reading about political news, and participation in conversations on 'political matters'.[36] As the items are formulated rather generally the respondent has some individual latitude in intepreting what he or she considers to be 'political'. Because of this, the interest score will probably be more encompassing in its referents than the other indicators of political involvement (3.5.2 through 3.5.4).

The predicted relationship could be observed empirically, although it is rather weak for recall 2-1, a bit stronger for recall 3-1, and most outspoken for the overall variable RECALL.[37] This last relationship is displayed in table 8.

Table 8: RECALL by political interest score. Cell entries are frequencies; between brackets % over the nonmissing cases (columnwise)

	Political interest score				
RECALL	0 (low)	1	2	3 (high)	total
consistent	93 (57)	73 (64)	54 (64)	49 (74)	269
'2 answers'	56 (35)	35 (31)	28 (33)	16 (24)	135
'3 answers'	13 (8)	6 (5)	2 (2)	1 (2)	22
missing data	46	18	8	11	83
total	208	132	92	77	509

$tau_b = -.11$ (nonmissing values only)

3.5.2 Strength of partisanship. Party identification is a concept which has been successfully employed in explaining political behavior in the United States. In other countries however its use is more problematic, as it turns out to be temporally unstable and empirically barely distinguishable from vote preference.[38] The *strength* component of party identification by itself however turns out to be closely related to different forms of political behavior. Thomassen (1975) suggests that 'the intensity of party indentification really refers to the motivational strength of the vote preference at a particular moment'. We accept this hypothesis as a plausible one, but we also assume that strength of party identification contains a component of political involvement as well. In either interpretation however, strong partisan attachments would lead to a higher proportion of consistent recall.

The variable we constructed contains the following categories: member of a party[39]/strong adherent/weak adherent/leaner/no party attachment. Strictly speaking, the expectation that members of political parties are highly involved in politics is more plausible than the one that persons without party attachment are not politically involved. Because of this the relationship between in-

volvement and recall may be attenuated when using this measure for involvement. Our data show a very strong, monotone increase in consistency as the strength of partisanship increases. As we found in the case of political intrest, the relation is weakest for recall 2-1 ($tau_b = -.25$), and strongest for RECALL ($tau_b = -.33$), with recall 3-1 in between ($tau_b = .-30$). Table 9 illustrates the association with the overall RECALL variable.

Table 9: RECALL by strength of partisanship. Cell entries are frequencies; between brackets % over nonmissing values (columnwise)

	Strength of partisanship					
RECALL	no party attachment	leaner	weak adherent	strong adherent	member of a party	total
consistent	31 (42)	68 (49)	60 (71)	47 (83)	63 (89)	269
'2 answers'	37 (50)	58 (41)	22 (27)	10 (17)	8 (11)	135
'3 answers'	6 (8)	14 (10)	2 (2)	— (—)	— (—)	22
missing data	34	25	15	5	4	83
total	108	165	99	62	75	509

$tau_b = -.33$ (only nonmissing cases)

3.5.3 Sense of political efficacy. There are several reasons why we expect political efficacy to be a component of what we've termed political involvement, and thus to be related with recall behavior. First of all, political efficacy is one of the more potent factors in explaining political activity.[40] Secondly, the wording of the items[41] implies that the respondent who gives the 'positive' response affirms the importance of the individual for the electoral process and vice versa. Such an agreement with the ideology of western parliamentary arrangements is likely to be accompanied by the feeling that politics truly matters, and may be one's own role in it as well.

Given the background of findings concerning political efficacy our analyses yield rather surprising results. For all 3 variables, recall 2-1, recall 3-1, and RECALL, we found that for the efficacy scores 0 through 3 the proportion of consistent responses increases monotonely (as expected), but in all 3 instances we also found that for the highest score – 4 – this proportion drops considerably. Also, in all instances the relationship is statistically not significant ($\alpha = .05$) using a χ^2 criterion, and hovers around the critical value using tau_b. To illustrate this, table 10 shows the relationship between the efficacy score and RECALL. At the moment it is not clear to us why those respondents scoring highest on the efficacy scale are among the least consistent ones in their recall.

3.5.4 Knowledge. It might be argued that a correct recall not only requires a

Table 10: RECALL by score on 'sense of political efficacy'. Cell entries are frequencies; between brackets % over nonmissing cases (columnwise)

	'sence of political efficacy'-score					
RECALL	0 (low)	1	2	3	4 (high)	total
consistent	25 (49)	84 (62)	66 (65)	59 (76)	35 (59)	269
'2 answers'	22 (42)	42 (31)	30 (29)	19 (24)	22 (38)	135
'3 answers'	4 (8)	10 (7)	6 (6)	— (—)	2 (3)	22
missing data	8	40	13	12	10	83
total	59	176	115	90	69	509

$tau_b = -.09$ $\chi^2 = 14.43$ (nonmissing values only) $df = 8$

certain involvement in politics but also knowledge concerning the political situation in which the election to be recalled took place. The more knowledge one has, the more cues one has available for reconstructing one's past behavior. Besides this, political knowledge, especially the more complex aspects of it[42] can be regarded as an element of involvement as well. All in all we thought it useful to inquire whether there is any relationship between political knowledge and recall. Out of the data gathered in 1971 we constructed 4 measures of knowledge, all relating to domestic politics in the Netherlands in 1971.[43] These are:

a) knowledge concerning cabinet ministers: nr. of correct names mentioned

b) knowledge concerning members of parliament: nr. of correct names mentioned

c) knowledge concerning political parties: nr. of correct answers on the question which parties are represented in parliament

d) knowledge concerning the composition of the cabinet: nr. of correct coalition parties menitoned, all responses containing wrong answers in the '0' category.

None of the first 3 measures yields a significant relation with any of the recall variables. Only the last measure is significantly – though weakly – associated with them: with recall 2-1 $tau_b = .11$, with recall 3-1 $tau_b = .12$, and with RECALL $tau_b = .11$.

The difference in results between the four indicators for political knowledge are not so surprising when one bears in mind that incorrect answers do not necessarily lead to a low score on the first 3 indicators, as contrasted with the fourth. We may conclude that those parts of knowledge about politics for which combinatorial insights are necessary, do correlate significantly, though weakly, with recall performance.

3.6 Participation in the electoral process

3.6.1 Frequency of voting. Voting in itself can be regarded as an expression of political involvement. The reason why we haven't treated this factor in section 3.5 is mainly that it relates much more directly to recall of *voting* behavior than the variables reported there. Why would we expect frequency of voting to correlate with recall consistency? First of all: irregular voters have an additional task to perform, namely to recall whether or not they voted at all in the election referred to, i.e. an extra chance to make mistakes. Secondly, we expect irregular voters to be politically less involved, and as we saw before involvement leads to more consistent recall. Thirdly, as long as the majority of the electorate do vote, and voting is considered to be socially superior to nonvoting, social desirability will distort the response in a number of cases. Such a response effect leads only to consistent recall if it yields the *same* answer in all 3 instances when the respondents were interviewed on their 1971 behavior; the chance that this happens must be considered remote. An opposite effect might be brought about by respondents who, out of principle refuse to participate in elections, and who, because of this principled behavior (i.e. consciously made choices) give the same answer every time they've been asked to report their 1971 voting behavior. In our estimate this group is so small in the Dutch electorate and so prone to refuse to be interviewed at all (let alone several times in this panel) that this potential countervailing effect will be virtually nil.

Table 11 presents the data; one must bear in mind that the information concerning the participation in the 3 elections has been obtained just after each of these elections and thus involves no recall effects over any appreciable period of time.

Table 11: RECALL by vote participation. Cell entries are frequencies; between brackets % over nonmissing cases (columnwise)

nr of times voted in the elections of '71, '72, '77						
RECALL	0	1	2	3	missing data	total
consistent	—	3 (18)	10 (28)	254 (69)	2	269
'2 answers'	—	12 (70)	21 (58)	99 (27)	3	135
'3 answers'	—	2 (12)	5 (14)	15 (4)	—	22
missing data	1	3	9	28	42	83
total	1	20	45	396	47	509

$tau_b = -.30$ (only nonmissing cases included)

Table 11 displays the relationship with RECALL; recall 2-1 and recall 3-1 showed analogous relations with frequency of voting.[44] It is clear that frequent voters (the large majority in the Netherlands) do not necessarily have a correct recall of their past voting behavior. Those few respondents however who report not to have participated in one or more elections have quite a low chance to give a consistent recall.

3.6.2 Certainty of choice. In 3.6.1 we saw that participation in elections is strongly related to recall, but the explanatory power of this finding is limited by the fact that 70% of the relevant cases report having voted in all 3 elections. Within this group we can still make distinctions however based on the certainty of the choice made. In this way we can also indicate the 'motivational strength of the vote preference' which Thomassen thought was indicated by strength of partisanship (see section 3.5.2). For us the relevant election is the one of 1971. We have 3 pieces of information which relate to the certainty of the choice made in 1971:

1. before the election of April 1971 the respondents were asked whether or not they intended to vote, and if so, for which party. Those who didn't know were then asked whether they hesitated between parties, or whether they had no idea at all what they were going to do.
2. after the election, respondents were asked how long before the election they had made their ultimate choice of party.
3. after the elections respondents who had voted were asked whether they would keep voting for the same party in the next couple of years.[45]

For each of these questions, the answers can be interpreted as a measure of the certainty of the eventual choice; the second question is at face value the most direct measurement, as it pertains to the real behavior and not to intended behavior as the first one does. The third question, suggesting fickle voting in the case when one would respond not to stay with the same party, almost certainly forces some voters into an 'it depends' response which would then be interpreted in some cases incorrectly as hesitation about one's own choice. Because of these reasons we expect the second measure in particular to be associated with recall behavior, for the other two we expect that the 'doubters' will be no more consistent in their recall then those who were very certain. Our data show that all 3 variables are significantly associated with recall performance, in the direction expected. The strength of these associations with each of the recall variables is as follows (expressed in tau_b):

	recall 2-1	recall 3-1	RECALL
hesitates before election	.15	.21	.17
when decided	.33	.33	.34
will vote same in future	.20	.27	.26

The relationships are robust over the 3 measures of recall consistency, and conform to our expectation that the 'when decided' item most unequivocally reflects motivational strength of voting choice.

3.7 Changes in voting behavior

Frequent changes in voting behavior will make it more difficult to recall the choice made at a particular election. We base this hypothesis mainly upon the increasing cognitive complexity of the recall task as voting behavior gets more diverse. As is the case with our previous hypotheses, it carries with it a ceteris paribus clause. A carefully considered choice will be remembered well even if the person in question changes his preference, while a stable, but unmotivated voter will be prone to make some mistakes in recall.

Changes in voting behavior will be indicated here with panel data, i.e. the variables 'vote 71', 'vote 72', and 'vote 77' (see figure 1). Out of these data we've constructed a new variable, which we refer to as *VOTEVAR*, which is simply the number of different parties one has voted for in the 3 elections (see figure 2B).[46]

In table 12 we display the relationship between VOTEVAR and RECALL. The result is a very high tau_b of .64. Recall 2-1 and recall 3-1 also yield very strong relations.[47] For all 3 recall variables we noticed that the proportion of inconsistencies in the category of stable voters is well within the range we hypothesized in section 3.2 as being caused by unavoidable mistakes and errors in datagathering and -preparation.

Table 12: RECALL by changes in voting behavior (VOTEVAR). Cell entries are frequencies; between brackets % over nonmissing cases (columnwise)

	Votevar				
RECALL	1 party (stable)	2 parties	3 parties (different at each election)	missing data	total
consistent	205 (93)	56 (38)	6 (12)	2	269
'2 answers'	15 (7)	85 (57)	32 (62)	3	135
'3 answers'	—	8 (5)	14 (27)	—	22
missing data	5	20	16	42	83
total	225	149	52	47	509

$tau_b = .64$ (nonmissing cases only)

The extremely high association – for survey data – between recall and vote stability suggests a simple process causing recall inconsistencies: it could very well be that for some people who change in their real voting behavior the answer to the question on *past* voting is answered as if it referred to *current*

voting. This could be interpreted as either a psychological solution to lack of memory (projecting the present back into the past), or as a dissonance reduction mechanism for those who would experience cognitive dissonance at reporting changed behavior.[48] This simple and rather elegant explanation does not stand up empirically. Of all inconsistent responses, 157 in total, only 36 conform to a pattern compatible with the process described above.[49] This is not meant as a denial of the operation of such a process, but only that it doesn't operate for all persons and/or permanently (i.e. at each of the recall moments).

3.8 Summary and concluding remarks

In this section we have reviewed a large number of possible and/or probable causes of recall inconsistencies. The following conclusions can be drawn:

1. In all cases the association between a variable and each of the recall variables (recall 2-1, recall 3-1 and RECALL) are in the same direction. In almost all cases the association is strongest with RECALL and because of this we will concentrate on this variable in the following sections.

2. The incidence of recall inconsistencies cannot be explained by 'technical errors', as these could account for at most 5% of the cases. Recall inconsistencies also do not correlate with inconsistencies in the repeated measurement of year of birth (used as an indicator for 'technical errors').

3. With our data we cannot detect any noticeable effect of interview interaction on recall performance.

4. Age is strongly related to recall: older respondents recall more consistently their 1971 vote than younger ones.

5. Other social ('background') characteristics do not relate in a significant way to recall. This is the more surprising as the education-social class-income syndrome is traditionally one of the most powerful correlates of political attitudes and behavior.

6. Political involvement leads to relatively consistent recall. This is especially true with the measures for political interest and strength of partisanship. Most indicators of political knowledge are not related to recall, this is at least partly due to the poor quality of these measures. Only indicators which require some insight into 'what goes with what' are significantly (but weakly) related to recall.[50] Political efficacy turns out not to be significantly related to recall.

7. Recall consistency is positively related to frequency of vote participation, and to the certainty of the eventual vote choice of the respondent.

8. Recall inconsistencies are strongly related to changes in voting behavior. The proportion of inconsistencies who remain loyal to the same party over 3

elections is so small that it could be accounted for by 'technically' caused errors.

The lack of association between social characteristics and recall came as a surprise to us. To check whether a significant association was not being suppressed by another variable we also computed the partial associations between recall and sex, income, education or social class, controlling for 'when decided' (se esection 3.6.2), strength of partisanship, frequency of vote participation, and changes in voting behavior. These partials do not lead to a different conclusion than the one stated above, in other words, the lack of association between recall and social characteristics doesn't constitute a case of a 'spurious-zero-correlation'. The same approach has been chosen in respect to political efficacy, which was the other case of a variable which – against our hypothesis – didn't correlate significantly with recall. Here too these analyses confirm our earlier conclusions.

The analyses we've performed in this section represent a first step in an effort to explain the incidence of recall inconsistencies. The next step, in section 4, will be to investigate how much variance in RECALL can be explained by the variables which were strongly related to RECALL, and to establish to what degree these explanatory variables are overlapping. This hopefully leads to a further understanding of factors which can be considered as 'causes' of inconsistency.

4 Regression Analysis

In the previous chapter we used tabular analysis to investigate the relationship between consistency of recall behavior and a great number of other variables. We found that the variable RECALL is associated with a number of other variables. However, this method only yields information about seperate pairs of variables, where RECALL is always one of the two. In other words we do not have information on the *interrelations between the variables* which are associated with RECALL. Furthermore we cannot yet say *to what degree the variation* in recall behavior *can be explained* by these other variables. There are several methods which could provide information on these two questions. Regression analysis, analysis of variance and loglinear analysis can be thought of. The data we use however, do not conform to the assumptions underlying regression analysis and analysis of variance. Nevertheless we have chosen to perform a multiple regression analysis on RECALL for the following reasons:

(a) the analyses of chapter 3 yielded more potential predictors of RECALL than can be handled by algorithms for loglinear analysis.

(b) as we do not know what interrelations between potential predictors exist, we cannot suffice with selecting those variables from the ones we investigated in chapter 3 which are most strongly associated with RECALL (using the magnitude of tau as a selection criterion for instance). The possibility that these variables are nested precludes a selection on such a basis.

The regression analysis which will be reported in this chapter is intended to lead to a simple selection of explanatory variables, while taking their interrelations into account. On the basis of the analyses in chapter 3 we selected the following variables to be included in the regression analysis:

INTEREST	score on political interest scale, measured in 1971 (section 3.5.1)
EFFICACY	score on political efficacy scale, measured in 1971 (sec. 3.5.3)
PARTISAN	strength of partisanship, measured in 1971 (sec. 3.5.2)
AGE	(section 3.4.1)
VOTFREQ	frequency of voting in the elections of '71, '72 and '77 (sec. 3.6.1)
KNOW	knowledge on composition of government coalition (sec. 3.5.4)
WHENDEC	moment at which respondent decided upon voting behavior in 1971 election (sec. 3.6.2)
VOTFUT	expectation in '71 that one will vote for the same party in future elections (sec. 3.6.2)
VOTEVAR	stability of voting behavior in the elections of '71, '72 and '77 (sec. 3.7)

We performed two separate regression analyses to which we applied pairwise deletion of missing data.[50a] The two runs differed in the treatment of the independent variables in relation to each other:

1. all independent variables were assigned the same inclusion level, i.e. we did not postulate a certain order according to which the variables should be included in the regression equation.

2. we assigned different inclusion levels in such a way that VOTEVAR had to be included in the regression equation after all the other variables were included, all other variables were assigned the same inclusion level.

The reasoning behind performing these two different analyses is as follows. We know from chapter 3 that the association between VOTEVAR and RECALL is stronger than of any other variable with RECALL. This implies that assigning all variables the same inclusion level will put VOTEVAR as the first explanatory variable in the equation. This could obscure the explanatory power of those independent variables which are multicollinear with VOTEVAR. Such a result would not bother us if VOTEVAR could be assumed to be causally prior to the other independent variables; however, we do not think

that such an assumption is warrented or tenable.
We stated earlier that we will use the regression analysis only to assess the set of explanatory variables. Because of these limited aims we will not report the regression coefficients but only the correlations and additions to the variance explained in RECALL and to arrive at a meaningful selection of a are summarized in table 13. On the basis of the data reported in table 13 we can draw the following conclusions:

- variables contributing significantly to an explanation of RECALL: VOTEVAR, WHENDEC, VOTFREQ, EFFICACY.
- variables which do not contribute significantly: VOTFUT, INTEREST, KNOW, PARTISAN, AGE.
- the contribution of WHENDEC, EFFICACY and VOTFREQ to R^2 is very small.

Table 13: Summary table of stepwise regression analysis; dependent variable RECALL, all independent variables were assigned the same inclusion level

step	variables entered	F to enter	significance	R^2 change	simple R
1	VOTEVAR	276.17	0	.44	.67
2	WHENDEC	7.21	.008	.01	.36
3	EFFICACY	7.52	.006	.01	.11
4	VOTFREQ	5.81	.016	.01	.29
5	VOTFUT	1.94	.165	.00	.25
6	KNOW	1.84	.176	.00	.14
7	PARTISAN	.72	.398	.00	.37
8	INTEREST	.37	.546	.00	.13
9	AGE	.13	.714	.00	.28
				$R^2=.48$	

Table 14 containes the results of the second regression analysis in which the variable VOTEVAR was forced to be included in the equation last of all.

The second regression analysis yields different results. Entering VOTEVAR as the last variable in the regression equation results in

- variables contributing significantly to the explanation of RECALL: PARTISAN, WHENDEC, VOTFREQ, AGE and VOTEVAR
- variables of which the contribution is ambiguous: EFFICACY
- variables not contributing significantly: VOTFUT, KNOW, INTEREST.

variance explained (with their significance). The results of our first analysis To select out of the 9 independent variables above some to be used in sub-

Table 14: Summary table of stepwise regression analysis; dependent variable RECALL, VOTEVAR was assigned the highest inclusion level

step	variables entered	F to enter	significance	R^2 change	simple R
1	PARTISAN	55.63	.000	.14	.37
2	WHENDEC	26.97	0	.06	.36
3	VOTFREQ	23.77	.000	.05	.29
4	AGE	4.36	.037	.01	.28
5	EFFICACY	3.18	.076	.01	.11
6	VOTFUT	2.12	.146	.00	.25
7	KNOW	2.37	.124	.01	.14
8	INTREST	1.11	.292	.00	.13
9	VOTEVAR	130.86	0	.20	.67
				$R^2=.48$	

sequent analyses, we applied the following criteria:

a) in at least one of the regression analyses there must be a significant contribution to the explanation of RECALL ($\alpha=.05$)
b) in at least one of the regression analyses the contribution to R^2 must be greater than 1%.

These criteria leave us with VOTEVAR, WHENDEC, PARTISAN and VOTFREQ, besides the dependent variable, RECALL. The structure of interrelations between these 5 variables will be investigated in the following chapter.

5 Loglinear Analysis

5.1 Introduction

Chapter 3 reveiled a series of variables which are significantly associated with RECALL; in the previous chapter we selected 4 variables out of this series which turned out to be the most promising set to explain variance in recall consistency. These 4, together with RECALL, constitute the set of variables to be analysed in depth in this chapter. We will refer to them by using characters, according to the correspondences indicated in table 15.
In this chapter we will apply loglinear analysis to arrive at a better understanding of the interrelations between these 5 variables. As this kind of analysis is not as well-known as, for instance, regression analysis, we will start with a short description of what loglinear modelling is about. This is not meant to be exhaustive, or even to serve as an introduction. For that, we refer the readers to the literature on this method.[51]

Table 15: Variables to be used in loglinear analyses

symbol	variable	description	reference
P	Partisan	strength of partisanship	section 3.5.2
F	Votfreq	frequency of voting in 3 elections	section 3.6.1
V	Votevar	stability of voting choice in 3 elections	section 3.7
W	Whendec	moment at which respondent decided upon '71 voting behavior	section 3.6.2
R	RECALL	consistency of recall	section 2.2

5.2 The loglinear model

The loglinear analyses does not deal with individual scores, as in analysis of variance or regression analysis, but with grouped scores. More specifically, the unit of analysis is the frequency (f) in a cell of a crosstabulation.[52] In a case with three variables, X_i, Y_j and Z_k, with X trichotomous and Y and Z dichotomous, we have the following 12 frequencies:

		Z			
		1		2	
		Y			
		1	2	1	2
X	1	f_{111}	f_{121}	f_{112}	f_{122}
	2	f_{211}	f_{221}	f_{212}	f_{222}
	3	f_{311}	f_{321}	f_{312}	f_{322}

For a general description of the steps involved in a loglinear analysis, we turn to Reynolds' list of five consecutive operations[53]:

(1) Propose a model that might account for the observed data. A model is really an hypothesis about distributions and interrelationships among the variables in a cross-classification. (In fact, 'model' and 'hypothesis' are used synonomously). Suppose, for example, one believed that a group of variables were mutually independent. Then the cell probabilities should reflect this independence.

(2) Derive a set of expectations under the assumption that the model is true. One next asks, 'What would a set of data look like if the model were true?' Since only a sample is available, these expectations have to be estimated from the observed data. Suppose, as in the previous example, the model asserted mutual independence. One could then estimate what a sample cross-classification of a given size would look like if it were drawn from a population in which the variables were really mutually independent. Since the estimation

process can be complicated, it is not described here, but fortunately many computer programs calculate the estimates automatically.

(3) After comparing the expected observations with the observed data, decide whether or not the model is acceptable. If the observed cross-classification is indeed drawn from a population having the given model, then any discrepancies between the expected and observed data should be due to sampling error. An appropriate statistic for testing this hypothesis is the familiar goodness-of-fit chi square which compares expected and observed frequencies. In measuring the discrepancies, the basic question is, 'Can the departures between observed and expected values reasonably be attributed to chance or are they so large that the model itself seems wrong?'

(4) If the discrepancies are small enough, retain the model and go to the next step. Otherwise, return to Step 1 and propose another model. The new model should, of course, be a refinement of the previous one for the analysis to be carried out as efficiently as possible.

(5) Having accepted a model, estimate its parameters. As in any model-building enterprise, one tries to make good estimates of the parameters. The parameters in turn are translated into substantive statements and used to make predictions.

In any application the cell-frequencies which are *expected* under a specified model are of crucial importance; they will be denoted here by F_{ijk}. These frequencies can be expressed as the result of the following multiplicative model

$$F_{ijk} = \tau . \tau^{x}_{i} . \tau^{y}_{j} . \tau^{z}_{k} . \tau^{xy}_{ij} . \tau^{yz}_{jk} . \tau^{xz}_{ik} . \tau^{xyz}_{ijk} \qquad (1)$$

In this way F, comparable to a dependent variable, is expressed in terms of a set of parameters τ. This set contains the grand-mean τ, the main-effects $\tau^{x}_{i}, \tau^{y}_{j}, \tau^{z}_{k}$ and the interaction-effects $\tau^{xy}_{ij}, \tau^{yz}_{jk}, \tau^{xz}_{ik}$ and τ^{xyz}_{ijk}. The model described in expression (1) constitutes a complete description of the observed data through effect parameters estimated with the same data.[54] The description is complete, or saturated, as all possible effect-parameters are specified in the model. As such, the model is a trivial one for explanatory purposes: all the information available has to be used in the estimation of the full set of parameters. If we specify models in which some parameters of the full set are left out, the F_{ijk}'s which are predicted through the parameters specified will deviate from the observed frequencies f_{ijk}. The deviations will be larger as the associations represented by the parameters left out are stronger. In this way we can classify all kinds of models and investigate how much the predicted frequencies differ from the empirically observed data. Large deviations signal a bad fit, i.e. a model that is unsatisfactory from a descriptive point of view, and which has to be rejected. For a number of reasons, which we will not dwell upon here, loglinear analysis uses the additive model which is ob-

tained by taking the logarithm of expression (1):

$$\iota nF_{ijk} = \lambda+\lambda^{x}_{i}+\lambda^{y}_{j}+\lambda^{z}_{k}+\lambda^{xy}_{ij}+\lambda^{xz}_{ik}+\lambda^{yz}_{jk}+\lambda^{xyz}_{ijk} \quad (2)$$

where: $\lambda = \iota n\tau$ etc.

To interpret the results of loglinear analysis one doesn't have to know how the lambda-parameters are estimated. For further explanation of this aspect, interested readers are referred to the literature mentioned above. A certain number of aspects pertaining to the application of loglinear models must be mentioned however:

1. Estimates of the parameters of the saturated model are calculated from the logarithms of the observed cell-frequencies. As $\iota n0$ is not defined, empty cells (i.e. cell frequencies are zero) constitute a special problem. To solve this problem one can add a constant to all observed frequencies. The literature suggests a value of 0.5 and we will use this figure in our subsequent analyses.

2. Like the τ's, the lambda's are referred to as *effects*. As a matter of notation, superscripts are used to denote the variable(s) the effect refers to, i.e. λ^{x} is the effect which corresponds to the variable X. Usually the subscripts i, j, k, . . . are omitted. The *order* of an effect is the number of variables listed in the superscript, i.e. λ^{xyz} is a third-order effect. Finally, for reasons of convenience XYZ is usually used instead of λ^{xyz}.

3. A model is called hierarchical if the specification of an effect implies that all effects of lower order are also included, which on turn refer to subsets of the variables in the superscript of the effect we started with. In other words, if we specify in a model the third-order effect XYZ, the model will necessarily contain XY, XZ, YZ, X, Y, Z and the grand-mean. In our analysis we have used only hierarchical versions of loglinear models.

4. We already indicated that a model has to be rejected if the differences between the predicted cell-frequencies (F_{ijk}) and the observed ones (f_{ijk}) are too large. These differences can be tested for significance with the Pearson's chi-square formula, which, when applied to our example of three variables gives:

$$\chi^2 = \frac{\sum_{i.\,j.\,k}(f_{ijk} - F_{ijk})^2}{F_{ijk}}$$

An alternative statistic which is often used in loglinear analysis is the likelihood-ratio-chi-square:

$$G^2 = 2_{i.}\Sigma_{j.k}f_{ijk}\iota n\frac{f_{ijk}}{F_{ijk}}$$

Both statistics are asymptotically chi-square distributed. The G^2 statistic has the advantage that it is additive when nested models are partitioned.

Two models, P and Q are *nested* if all effects of one model, P, constitute

a subset of the effects specified in the other model, Q. The difference in G^2 between the two models can be used to test whether or not the additional effects specified in Q contribute to a significantly better prediction of frequencies. This difference is asymptotically chi-square distributed; the difference between the degrees of freedom of P and Q yields the degrees of freedom to be used in this test. As the Pearson chi-square statistic is not additive for partitioned nested models, we will use G^2 in testing the goodness of fit of our models. The sample must be 'sufficiently large' to assume that G^2 is chi-square distributed. How 'large' this is, is disputed in the literature; according to some the sample has to be at least 10 times as large as the number of cells in the cross-tabulation[55], whilst others are more lenient, and tolerate a ratio of 5 times the number of cells.[56]

5. Computer programs for loglinear analysis[57] give the user the option of testing for partial and marginal association. We have used these tests, therefore we quote from a manual for further explanation of these tests:

'The test of any effect in a log-linear model depends on which other effects are included in the model. Therefore no single test determines the relative importance of an effect. Brown (1976) suggested the use of two tests – marginal and partial association – to screen effects.

The hypothesis that the *partial association* of k factors is zero is a test of whether a significant difference exists between the fit of two hierarchical models – one is the full model of order k, and the other the model that differs from it in that the specified k-factor interaction is excluded. For example, to test the partial association of A and B (i.e., two factors), the full second order model is fitted and then the same model with λ^{AB} set to zero. The difference in the tests-of-fit is a test of partial association.

The hypothesis that the *marginal association* of k factors is zero is a test that the k factor interaction is zero in the marginal subtable formed by the k factors (i.e., summed over all other factors). For example, to test the marginal association of A and B (i.e., two factors), the two-way table indexed by A and B is formed and the two-factor interaction is tested.

The tests of marginal and partial association can be simultaneously used to screen the various interactions to determine whether they are necessary in the model for the data being used, whether they are not necessary, or whether they are questionable. In a second pass the models that contain all the necessary terms and relevant combinations of the questionable terms can be defined, and an appropriate model (or models) for the data can be rapidly chosen. This is further explained in Brown (1976).[58]'

Table 16: Table with frequencies used in loglinear analyses with 5 variables

P: (Partisan)	1: no party attachment, or leaner
	2: weak adherent, strong adherent or member
W: (Whendec)	1: decided upon vote a few months before election, or later
	2: decided earlier upon vote in '71
V: (Votevar)	1: voted always the same
	2: voted twice the same, once different
	3: voted different at each of the 3 elections
F: (Votfreq)	1: voted in 2 or less elections
	2: voted in all 3 elections
R: (RECALL)	1: consistent recall of '71 voting behavior
	2: inconsistent recall

				R	
W	V	F	P	1	2
1	1	1	1	0	0
			2	0	0
		2	1	10	3
	2	1	2	17	1
			1	1	1
			2	0	0
		2	1	15	30
			2	6	6
	3	1	1	0	6
			2	0	0
		2	1	0	16
			2	0	2
2	1	1	1	0	0
			2	0	0
		2	1	42	6
			2	132	5
	2	1	1	2	3
			2	3	1
		2	1	16	16
			2	9	15
	3	1	1	1	3
			2	0	0
		2	1	1	7
			2	2	5

for analysis 0.5 is added to each cell above. Total frequency is 383.

5.3 The P-W-V-F-R model

A primary understanding of the importance of the different effects (whether or not they are indispensible), can be gained by investigating the saturated model, i.e. the model with al possible effects included. The basis for all our subsequent analyses is the crosstabulation of the 5 variables involved. The relatively small number of respondents forced us to collapse some categories of the variables involved in order to avoid too many cell-frequencies to be estimated. Table 16 gives the observed crosstabulation and the contents of the categories in each of the variables after recoding.[59]

Because of listwise deletion of missing cases, all respondents who had a missing data score on one or more of the 5 variables involved are eliminated. This leaves us with 383 respondents for our analyses. Examining table 16 we see that the distribution of the cases over the cells of the table is rather uneven. No less than 17 cells (out of 48 in total) contain no cases at all. The ratio between the number of respondents and the number of cell frequencies to be estimated is 383/48 = 7.98, which is a ratio considered by some authors to be too low for loglinear analyses, by others as sufficient. Needless to say, the conditions for loglinear modelling are far from ideal, even after the recodings we performed to get fewer cell frequencies and fewer empty cells. Nevertheless, we think that the problems are not intolerable, and that further analyses are still worthwhile.

First of al we inspect the test of marginal and partial association of the different effects, as discussed earlier in section 5.2 (point 5). The results of these tests have been summarized in table 17.

Analysing the results reported in table 17 leads to the following observations:

– *The main-effects (1st. order effects)* with the exception of P (=Partisan) all reach significance. This result is rather trivial in our case and does not allow us to draw any substantial conclusions, as it only reflects the marginal distribution of the 5 variables involved. As we know from table 16 all variables except P are distributed very unevenly over their categories, which means that besides the grand-mean, knowledge of these marginal distributions (i.e. the 1st. order effects) contributes significantly to any prediction of the cell-frequencies. The variable P however is relatively evenly distributed; its grand-mean is 383/2=191.5, the size of the 2 categories being 183 and 200 respectively. The difference between grand-mean and these frequencies is not significant.

– *2nd order effects* PR, VR, PW, WV, FV and PV are significant effects; FW is not significant. FR, PF and WR are dubious (these effects are significant according to only one of the tests).

Table 17: Tests of marginal and partial association of the factors in a log linear model

Factor	DF	test of marginal association chi-square	probability	test of partial association chi-square	probability
P	1	1.5	.22	1.5	.22
R	1	42.9	.00	42.9	.00
F	1	335.2	.00	335.2	.00
V	2	117.6	.00	117.6	.00
W	1	60.5	.00	60.5	.00
PR	1	46.6	.00	4.4	.04
PF	1	7.6	.01	.6	.45
PV	2	66.9	.00	15.8	.00
PW	1	39.1	.00	12.2	.00
RF	1	10.7	.00	.1	.72
RV	2	166.9	.00	103.1	.00
RW	1	40.0	.00	3.5	0.06
FV	2	35.5	.00	22.7	.00
FW	1	2.1	.15	.5	.50
VW	2	56.6	.00	14.1	.00
PRF	1	.1	.79	.2	.64
PRV	2	4.3	.12	4.3	.12
PRW	1	.6	.45	1.1	.29
PFV	2	2.2	.34	1.3	.51
PFW	1	.6	.45	.4	.55
PVW	2	.7	.71	.0	.98
RFV	2	6.0	.05	5.0	.08
RFW	1	.7	.42	.1	.72
RVW	2	2.3	.31	1.6	.45
FVW	2	5.5	.07	3.9	.15
PRFV	2	.9	.64	.8	.68
PRFW	1	.0	.84	.3	.60
PRVW	2	.7	.71	.6	.75
PFVW	2	.3	.85	.2	.91
RFVW	2	.1	.94	.1	.97
PRFVW	2	.4	.80	.4	.80

– *3rd, 4th, and 5th order effects.* None of these effects reaches significance.

The next task is to delete effects from the full model and arrive at a more parsimoneous model which still fits the data. It is virtually impossible to test all possible models, therefor we use theoretical considerations and the information summarized above to guide us. In the evaluation of the models we

will not only use their G^2 (1r chisq), but also their effect parameters (i.e. the loglinear labda parameters) to find out which of the effects specified in a model possibly can be deleted. Because the results we obtained from the tests for marginal and partial association are only valid for the full model, and not for simpler ones, we think it best to simplify the model through a series of small, successive steps, taking into account at each step the lambda parameters of the effects included in the model, and the G^2 (goodness of fit). We will not report all these details, and only report the successive models tested and their fit.

5.4 To a parsimoneous model: deleting effects from the full model

In this section we wil discuss how we arrived at a parsimoneous model representing the data reported in the table of frequencies (table 16), i.e. using all 5 variables P, W, V, F, and R. As the whole analysis consists of a number of successive steps (each step being the test of a different model), we will not include all the details of ech of the models tested. In table 18 we have summarized the results of these steps.

Of the models listed in table 18 model 1 is the full model discussed in section 5.3. Model 2 (PWVR, VFR) has an almost perfect fit (note that the significances are used just the other way around as 'normal': low G^2 values mean high probabilities, i.e. a good fit).

Note in table 18 that for some models the exact probability is not reported. This is caused by the computer program we used (ECTA) which only prints the exact probability when it is under .50. This feature is based on the assumption that higher values imply redundancy in the model. With the help of chi-square tables in normal textbooks we still can get an impression of the fit of these models: the probability of model 2 is .95, the probability of model 9 is $>.75$.

On the basis of the results reported in table 18 we can conclude that model 2 has an almost perfect fit. Evidently the relations of F with P and W can be deleted, as well as the 5th order effect and all 4th order effects except PWVR. The next step is to replace the PWVR effect with all 3rd order effects included in it (model 3). The fit remains very good. In models 4 to 8 we investigate whether or not any of the 3rd order effects is necessary for a good fit. This is not the case. Even though this doesn't prove that all 3rd order effects in combination can be deleted, we didn't check all possible deletions of combinations of 2 and 3 of the 3rd order effects, and move directly to model 9. In model 9 we included only those 2nd order effects which are significant in tests on marginal and partial association for the full model (see table 17). Model 9 fits very well, even to such a degree that it seems useful to investigate

Table 18: Goodness of fit of the models tested (5 variable models)

number	model*	df	G^2	probability
1	PWVFR	0	0	1.00
2	PWVR, VFR	18	9.43	>.50
3	WVR, PWR, PWV, PVR, FVR	20	10.11	>.50
4	WVR, PWR, PWV, PVR, FV, VR, FR	23	16.10	>.50
5	WVR, PWR, PWV, —, FV, VR, FR	24	20.16	>.50
6	WVR, PWR, —, PVR, FV, VR, FR	24	16.12	>.50
7	WVR, —, PWV, PVR, FV, VR, FR	23	17.21	>.50
8	—, PWR, PWV, PVR, FV, VR, FR	24	17.62	>.50
9	PW, PR, WV, VR, FV, WR, PV	30	23.90	>.50 (≈.82)
10	PW, PR, WV, VR, FV, WR, —	32	42.01	.11
11	PW, PR, WV, VR, FV, —, PV	31	27.43	>.50 (≈.65)
12	PW, PR, WV, —, FV, WR, PV	32	132.69	.00
13	PW, PR, —, VR, FV, WR, PV	32	37.55	.23
14	PW, —, WV, VR, FV, WR, PV	31	28.22	>.50 (≈.60)
15	—, PR, WV, VR, FV, WR, PV	31	35.93	.25
16	PW, —, WV, VR, FV, —, PV	32	33.64	.39
17	PW, —, WV, VR, FV, WR, —	33	69.93	.00
18	PW, PR, —, VR, FV, WR, —	34	64.04	.00
19	—, PR, WV, VR, FV, WR, —	33	62.43	.00
20	PW, PR, WV, VR, FV, —, —	33	43.44	.11
21	P, V, W, F, R	41	373.52	.00

* As we are working with hierarchical models, all those effects of lower order are included in the model, which refer to subsets of the variables specified in the label of the model (see also section 5.2. point 3).

which effects (if any) can be deleted from it. This investigation is undertaken in models 10 through 20. Inspecting these models and their fit makes clear that not all the effects included in model 9 are of equal importance. On the basis of their fit only models 11, 13, 14, 15, and 16 qualify as acceptable. The other models fit poor, indicating that PV and/or VR cannot be deleted. No models were tested with FV deleted, as this is the only effect linking F into the structure of variables.

How can we make a choice out of the reasonably fitting models 11, 13, 14, 15, and 16? A first consideration has to do with parsimony and fit. Models 11, 14, and 15 are equally simple (expressed in df). Of these three models 11 has the best fit and is therefore preferable above the other two. Models 13 and 16 are also equal in their degrees of freedom; of these two model 16 has clearly a better fit, and is thus to be preferred. This leaves us with two models, 11 and 16, which differ in fit (G^2) as well as in simplicity (df). A simple criterion is to compute the difference in G^2 and in df, and testing this difference for signifi-

cance. As we see in table 19 the difference in fit between models 16 and 11 is $G^2=6.21$, df=1, this difference being significant. In other words: *even though model 16 fits the data acceptably, its advantage in simplicity (measured in df) over model 11 is outweighted by its significantly worse fit to the data, so model 11 is to be preferred.* In table 19 we have listed the differences in fit and df between some of the models under consideration. From this table we learn that only the WR effect and none of the others can be deleted from model 9 without significantly impairing the fit. We also see (last row) that deleting PR from model 11 significantly reduces the fit.

Table 19: comparisons between some of the models from table 18

models compared	effect deleted	df	G^2
9-11	WR	1	3.53
9-13	WV	2	13.65*
9-14	PR	1	4.32*
9-15	PW	1	12.03*
9-16	PR, WR	2	9.74*
11-16	PR	1	6.21*

* significant at p = .05

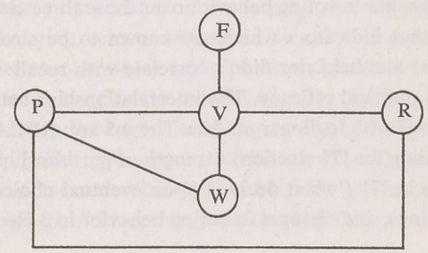
Summarizing, the reason for choosing model 11 as the best representation of the data are:

a) model 11 is superior in goodness of fit compared with all other models of equal complexity (measured in df)

b) no effects can be deleted from model 11 without a loss of fit significant in relation to the gain in simplicity (measured in df).

As model 11 contains only main-effects and second order effects we are now also able to represent it in a more visual way, as illustrated in figure 3.

Figure 3: A visual representation of model 11 (see table 18), specified by PW, PR, WV, VR, FV, PV



When interpreting a loglinear model like model 11 we have to keep in mind that it refers to the structure of a complete set of variables; as such it differs from regression analysis and causal analysis as there is not a dependent variable. We should use the results of this analysis as additional to those of the tabular and regression analysis reproted earlier. What insights were gained then by using loglinear analysis? The following conclusions are new compared with our earlier analyses:

- the effect of none of the variables is fully contained within that of any of the others, i.e. anyone of the variables has a 'unique' contribution in predicting the structure of interrelations of table 16 (the data).
- for a description of the data we can suffice with second-order effects.
- because we have only second-order effects in the model, we can use it as a starting point for causal analysis and present it as a conventional causal model with arrows between the variables.

6 Conclusions and Further Research

The results of the analyses we've presented in this paper can be readily summarized: panel data and recall data frequently yield different results in analyses of individual voting stability. This is mainly caused by the great number of recall inconsistencies we observed (at each of the recall moments only 60 to 70% of the answers were in concordance with a report given immediately following the elections). Compared with panel data recall leads to a serious underestimation of the number of vote changers; soms 16-20% of the cases which are classified as stable or changing voters on the basis of recall data are incorrectly designated so; some 30% of the cases which are correctly classified as changers are, with recall data, incorrectly placed as to the party left by them.

We examined a great number of variables to see whether they correlated with recall consistency. We were surprised to find that only a relatively small number of factors appeared to be associated at all with recall behavior. These are: age, political interest, strength of partisanship, political knowledge, certainty of voting choice in 1971, frequency of voting in the '71, '72 and '77 elections, and changes in voting behavior over those three elections. The most surprising was that indicators which are known to be strong predictors of political attitudes and behavior didn't correlate with recall: education, class, income, sex and political efficacy. The interrelationships between 5 variables have been analysed with loglinear models. These 5 are: RECALL (consistency of recall concerning the '71 election), strength of partisanship in '71, certainty of voting choice in '71 ('when decided upon eventual choice'), frequency of voting in 3 elections, and changes in voting behavior in 3 elections. The result

of this analysis was a simple model, containing all 5 variables, and specifying only some of the second order effects between the variables, and none of the higher order effects. The model fitted very well with the data.

Given the analyses reported above some additions are evidently warranted, and will be pursued by us. First of all, our loglinear models are no causal models, but general models of interrelations between variables. A conversion into a causal model explaining RECALL is necessary if we want to arrive at an explanation of inconsistent recall. Having done this, it then becomes imperative to consider the role of age in the context of such a model. We hypothesized that age by itself would not influence recall behavior, but only through a number of intervening variables. Whether the variables to be specified in a causal model perform this function adequately, or whether still other intervening variables have to be introduced will be investigated. A totally different approach to studying the reliability of recall is using latent class models in panel analysis (Wiggings, 1955). Thomassen applied a simple model out of the class of models specified by Wiggins on a 2 wave panel, after dichotomising the variables indicating voting behavior into 'voted' and 'abstained' (Thomassen, 1974). We will consider these models as a way of analysing the recall data; some preliminary analyses suggested however that the potential of these models is limited in explaining recall, as many of the assumptions on which the application of these models rests, seem to be unrealistic for the kind of data we are talking about.

We intend to widen our research beyond explanations of the incidence of recall inconsistencies which was our main concern in this report. We will pursue the following topics:

- until now we've treated all recall inconsistencies alike. However, the question can be asked whether or not each one of the possible answers that would constitute recall inconsistencies, is equally probable. Preliminary analyses (see Van der Eijk and Niemöller, 1979) indicate that this is not the case. Pursuing this question we will focus mainly on inconsistent recallers and investigate whether there is any structure to be discerned in their inconsistencies, and if so, what that structure looks like.
- our conclusion in this paper that recall data are very frequently incorrect, poses immediately the problem of reassessing what we think to know about stable and changing voters, and of individual voting dynamics, because most of the literature on these topics has been based on recall data.
- after gaining more insight into changes of party choice we will focus again on the relationship between vote changing and recall inconsistencies (see also section 3.7). The main question which we will address then is how these two variables affect each other dynamically, i.e. whether changes in voting behavior 'cause' recall inconsistencies, or whether in some cases the causal

relation is different.

- finally we will have to concern ourselves with the practical question whether or not the recall approach in measuring individual vote changes can be salvaged. If this is possible it would be well worthwhile as it is an easy way to gather data, compared with panel studies which although superior as far as quality of the data is concerned, will remain relatively rare because of the problems entailed with them. The question to be solved is how to correct for the biases contained in recall data.

Elsewhere we hope to report analyses on these topics for further research.

Notes

1 This is not the place for listing an exhaustive bibliography of literature on individual voting dynamics. Going through the contents of major professional journals such as the APSR, POQ, Am. J. Pol. Sc., Sociale Wetenschappen and Acta Politica yields scores of articles in the last decade alone which fall in this category. This just includes articles which can be classified as relevant based on their title alone, many other articles on this subject are not so easily recognized. Furthermore there is an increasing number of election studies in which vote changing is investigated.

2 See, for instance Tingsten (1937).

3 To give an impression of how rare panel data are used to describe voting dynamics we went through all 13 volumes of Acta Politica and found only 1 article in which panel data are used. We have the impression that this is not different from other relevant journals.
Another indication of the dominance of recall data can be found in a classic review of studies on floating voters (Daudt, 1961); virtually all studies reviewed use recall data, and the author himself only cursorily hints at the possibility that such data might be biased, without however taking this as a serious problem (see p. 72).

4 Some studies in which the author(s) have seriously addressed the problem of the quality of recall data are Clausen (1968), Riedel and Dunne (1969), Adamany and Dubois (1975), Thomassen (1974), Kelley and Mirer (1974), Weir (1975), Himmelweit c.s. (1978), Benewick c.s. (1969), Sudman and Bradburn (1974, p. 56).

5 See for instance Glenn (1969), Booth (1970), Schreiber (1975), Sudman and Bradburn (1974), Weiss (1968), Cahalan (1968), Tittle and Hill (1967), Luria (1973), Hilgard c.s. (1971), Anastasi (1958).

6 Documentation on the election studies and the panel study can be obtained from the codebooks, available at the Steinmetz Archives, Kleine Gartmanplantsoen 10, 1017 RR Amsterdam. These studies are (or will be) stored at the ICPSR also.

7 The dataset which we analyse in this paper consists of a panel over three consecutive national election studies in the Netherlands.
During the first election study, in 1971, a sample of the Dutch electorate was interviewed twice. The first of these waves yielded 2495 successful interviews.

At the second wave 1981 interviews were completed. The election study of 1972-73 consisted of a fresh sample. Added to this were all those respondents of the 1981 (who had already been interviewed twice) who could be located and who were willing to be interviewed again. This study also consisted of two waves. During the first of these 1325 respondents of the original 1971 sample cooperated again, after the second wave there were still 888 left, who, by then, had been interviewed four times.
During the 1977 election study a special questionnaire was designed for this panel. 509 interviews (out of the 888 left after the '73 interview) were completed. All in all, only 509 of the original 2495 respondents stayed with us during this 6 year time period, i.e. 20.4%. One might assume that in a lot of aspects these 509 would constitute a very select group due to nonrandom (biased) panel mortality. Preliminary analyses show that, surprisingly enough, in most relevant characteristics the respondents which were still with us in 1977 do not differ significantly from those who dropped out. (These analyses will be reported seperately in the near future).
In our opinion it is not too farfetched to make careful generalizations from the panel-group to the original sample (and thus, to the 1971 electorate).

8 It should be stressed that our measures of actual voting behavior (i.e. vote 71, vote 72 and vote 77) are recalled as well as the other measures, beit that the period between the event itself and the recall of it is extremely short. Given the nature of elections is western-style democracies (secret ballot), a *direct* observation of the voting act itself is impossible.

9 Readers not familiar with the Dutch partysystem and the different political parties are referred to Daalder (1966), Lijphart (1974) and Lijphart (1975).

10 This is the case as well with recall of the 1972 election. The marginal distributions of vote 72 and recall 77-72 are as follows:

11 It is easy to imagine that identical marginal distributions can be obtained be-

	vote 72	recall 77-72
didn't vote	35	21
CPN	12	7
PSP	5	4
PvdA	124	158
PPR	23	17
D'66	24	15
DS'70	21	7
KVP	103	121
ARP	30	33
CHU	21	23
VVD	64	71
BP	7	6
GPV	7	5
SGP	7	5
RKPN	4	—
other parties	5	—
no answer/don't know	17	16
total	509	509

cause all individual difference 'cancel out'. This means that marginal stability is not sufficient for concluding individual stability. On the other hand, if every individual were to give identical answers to the original question and to the recall question, then the marginal distributions will per definition be identical.

12 In the previous note we stated that marginal stability is a necessary but not sufficient condition for individual stability. Our remark on these conditions thus referred to different levels of analysis: the whole sample versus the individual respondent. When we say here that individual consistency is also a necessary but not sufficient condition for accurate recall we refer to the possibility of a discrepancy between reported behavior and actual (but unobserved) behavior. If this discrepancy is stable over the period of time covering our interviews, a respondent may give consistent answers.

13 The construction of the variables Recall 2-1, Recall 3-1 and Recall 3-2 conforms to the following expressions (illustrated for Recall 2-1 only):
if ((vote 71=recall 72-71) and (vote 71≠missing)) recall 2-1=consistent
if ((vote 71=missing) or (recall 72-71=missing)) recall 2-1=missing
if ((not (recall 2-1=consistent)) and (not (recall 2-1=missing))) recall 2-1=inconsistent
The construction of the overall variable RECALL conforms to:
if ((recall 2-1=consistent) and (recall 3-1=consistent)) RECALL=consistent
if ((recall 2-1=missing) or (recall 3-1=missing)) RECALL=missing
if ((not (RECALL=consistent)) and (not (RECALL=missing))) RECALL=inconsistent

14 Figure 2 A, B, and C show the variables out of which the created 'vote' and 'float' variables have been formed. This has been done analogous to the expres sions in the previous note. One extra caveat had to be dealt with: in 1977 the KVP, ARP and CHU did not appear on the ballot anymore, they had been combined in the CDA. Voters for the CDA in '77 who had voted for (panel data) or recalled to have voted for KVP, or ARP or CHU in previous elections were classified as stable in the 'vote' and 'float' measures.

15 Of course, this problem does not exist for the stable-stable cell of table 6. From figure 2 we know that for these respondents vote 71=vote 72=recall 72-71. For the respondents in the changing-changing cell of table 6 we know that vote 71≠ vote 72 and that vote 72≠recall 72-71, but this does not imply that vote 71= recall 72-71.

16 This information is obtained by selecting those respondents who are classified as changers by both measures and to crosstabulate for this category the original question and the answer to the recall-question. This yields: 116 respondents who are changers according to vote 2-1 and float 2-1; of this group vote 71 is different from recall 72-71 in 37 cases (32%).
For a comparison of t_1 and t_3, and of t_2 and t_3 we applied an analogous procedure, leading to similar results.

17 Individual inconsistencies might cancel out against each other, so that the *distribution* of vote changers as to the party which they left would look alike for panel and for recall data. That this is not the case can be seen from the following table which shows considerable differences (only reported for t_1-t_2).
Party left by vote changers; changers indicated by 'vote 2-1' (panel data) and by

18 Non existing – mistake in print.

'float 2-1' (recall data)

	Measure indicating voting change: vote 2-1	float 2-1 party left indicated by:
Party left	vote 71	recall 72-71
CPN	2.8%	4.2%
PSP	1.1	1.4
PvdA	17.0	16.7
PPR	2.3	.7
D'66	12.5	17.4
DS'70	9.1	4.9
KVP	15.9	18.8
ARP	4.0	5.6
CHU	7.4	2.8
VVD	4.5	6.9
BP	1.1	2.1
GPV	1.1	.7
SGP	—	—
RKPN	.6	—
else	4.0	2.8
didn't vote	16.5	15.3
total	100%	100%
	(176)	(144)

19 The term 'side information' comes from Philip E. Converse (1976), p. 16-26. The term indicates Bayesian 'a priori grounds of plausibility'.

20 We will treat RECALL as an ordinal variable.

21 The respondents have been asked 5 times to report their year of birth. In the dataset available to us 4 of these questions were coded as year of birth and one as age. We only used the year-of-birth-coded variables. If these variables weren't exactly the same, respondents were coded as inconsistent. Obviously, the more often a question on a stable characteristic (like year of birth) is repeated, the greater the number of inconsistent responses will be.

22 As far as we know there is not an agreed-upon way of estimating the proportion of punching and coding errors in general.

23 Crosstabulating RECALL and inconsistencies in reported year of birth yields:

		year of birth reports inconsistent	consistent	total
RECALL	consistent	14	255	269
	'2 answers'	10	125	135
	'3 answers'	1	21	22
	missing	3	80	83
	total	28	481	509

$\chi^2 = .86$, df = 2, prob. = .65
tau-b = −.03, prob. = .25
} only nonmissing values included

For the other recall variables we find:

	Recall 2-1	Recall 3-1	Recall 3-2
χ^2	1.95	.02	2.42
df	1	1	1
prob.	.16	.89	.12
tau-b	.08	−.02	.08
prob.	.05	.36	.04

24 Terming the ratings 'cooperation' and 'truthfullness' we observed the following associations (tau-b):

	RECALL	Recall 2-1	Recall 3-1	Recall 3-2
'cooperation'	.05	.07	.02	.07
'truthfullness'	.08″	.07	.09″	.10″

″: significant at $\alpha = .05$; in all these cases χ^2 doesn't reach significance.

25 Due to lack of insight in the meaning of these ratings we doubt their uselfulness. Furthermore, our knowledge of the interview situation is very poor as we do not have ratings for other important aspects thereof (see for instance Sudman and Bradburn (1974), p. 29 and following).

26 See Anastasi (1958) p. 249: 'with regard to memory age decrements have generally been found', and Sudman and Bradburn (1974) p. 85: 'of all the respondent variables, . . ., only age is related to memory'.
In advanced studies on the relationship between memory and age several memory factors are distinguished. One or more of these factors are used in recall tasks; the nature of the stimulus determines which combination of them is invoked. Furthermore respondents can differ from each other in this combination even if they are asked to recall the same manifest stimulus. Of course this causes that there is little agreement on the strength of the relationship between age and memory factors.

27 See for instance Verba and Nie (1972) ch. 9, and Converse (1976).

28 tau-b (recall 2-1 by age) = −.16
tau-b (recall 3-1 by age) = −.21
tau-b (recall 3-2 by age) = −.12

29 There is some discussion on the correctness of the 'common wisdom' that women are politically less involved than men. Whether or not one agrees with such a statement is partly dependent on the measures of political involvement employed. The more a definition of 'politics' is restricted to parliamentary politics and the more involvement is indicated by participation, the more the statement seems to be confirmed. Wider definitions of politics and measures of involvement which include attitudinal components usually show much less differences in political involvement between men and women.

30 χ^2 (RECALL by sex) = .35 df = 2 prob. = .84
χ^2 (Recall 2-1 by sex) = .07 df = 1 prob. = .79
χ^2 (Recall 3-1 by sex) = .15 df = 1 prob. = .70
χ^2 (Recall 3-2 by sex) = .31 df = 1 prob. = .58

31 tau-b (RECALL by education) = .05
tau-b (Recall 2-1 by education) = .03
tau-b (Recall 3-1 by education) = .06
tau-b (Recall 3-2 by education) = .01

32 The precoded answers to the question on subjective class are: upper class, upper middle class, middle class, working class, lower working class.

33 tau-b (RECALL by subjective class) = .06 (n.s.)
percentage missing cases on RECALL for upper + upper middle class 16%
middle class 14%
working class 17%
lower working class 26%

34 Urbanization and income are both associated with RECALL with a tau-b of .03, being not significant.

35 All indicators in section 3.5 only tap political involvement concerning phenomena falling under the scope of the colloquial meaning of politics, i.e. party-, parliamentary-, and government affairs. Involvement in other spheres which might be conceived of as political, but which are not commonly referred to under that heading, is not included in these indicators.

36 The three items used are:
a Can you indicate in this card how often you read the Dutch news in your paper, e.g. about wage and price policy or about government problems in The Hague?
b And again, can you indicate on this card how often you read the international news in your paper, e.g. about tensions and or negatiations between countries?
c If problems are being discussed like the wage and price policy or government problems in The Hague, do you usually join the conversation, do you listen with interest, don't you listen or aren't you interested?
A simple summation score has been used taking as 'positive' answers (nearly) always for items a and b and joins for item c. From scale analysis (method Mokken (1971)) we know that these items are unidimensional.

37 tau-b (recall 2-1 by interest) = −.09
tau-b (recall 3-1 by interest) = −.11
tau-b (recall 3-2 by interest) = −.09

38 The status of the concept of party identification in European contexts has recently been under serious consideration (see Budge c.s. (1976)). Thomassen (1976) claims that it is not a useful concept in the Dutch situation. Others have expressed doubts about the applicability of the concept of party identification in the GFR. Falter (1977) is more optimistic however about the theoretical relevance and applicability of the concept even though he too recognizes that operationalizations remain problematic. We share his opinion and think that it is applicable for the Netherlands as well. We will elaborate further upon this point elsewhere.

39 Different from the usual measures of partisanship which range from strong adherent to no attachments we also included membership of a party in our measure. In this respect our measurements differ from the one used by Thomassen (1976).

40 See for instance Mokken (1971) and Pateman (1970).

41 The 4 items combined in the efficacy score are:
a members of parliament are not really concerned about the opinions of people like me
b the political parties are only interested in my vote and not in my opinion
c people like me have no influence ate all on government policy
d so many people vote at elections that my vote doesn't count

These items have been shown repeatedly to form a unidimensional scale in the Netherlands, see for instance Daudt c.s. (1968) and Mokken (1971).

42 We think a (gradual) distinction can be made between knowledge about single stimuli, and knowledge about stimuli and their context together. The latter is obviously the more complex type. We think that operationalization d) below (see page 309 of this paper) taps the more complex aspects of political knowledge better than a), b) or c) because wrong answers are scored as 'little knowledge'. Doing this with the other question poses special problems (see note 43).

43 Because of frequent changes in cabinet ministers and especially in MP's we think that wrong answers should not be intepreted as lack of knowledge, as in most cases the answers given referred to persons who only recently left the position mentioned. As far as the political parties are concerned it is virtually impossible to give a wrong answer to this question as at the time all existing parties were represented in parliament. Only with the question on the composition of the government it is reasonable to interpret wrong answers as lack of knowledge, because at the time that coalition had been in government for 4 years without change.

44 tau-b (recall 2-1 by votfreq) = −.29
tau-b (recall 3-1 by votfreq) = −.25
tau-b (recall 3-2 by votfreq) = −.22

45 The relevant questions are:
a 1 As you know, parliamentary elections will be held in April 1971. Do you intend to vote in any case, are you certainly not going to vote or don't you know yet? (hesitate is 'don't know')
2 If intending to vote: For what party do you intend to vote? (hesitate is 'don't know')
b When did you decide to vote for this party, a few days before the elections, a few weeks before the elections, a few months before or had you made up your mind earlier about how to vote?
c Do you think you will continue to vote for this party in the future?

46 VOTEVAR has been constructed according to the following expressions (in this order) (see also figures 1 and on pages 290 and 297 of this paper):
if ((vote 71 = vote 72 = vote 77) and (vote 71 ≠ missing)) VOTEVAR=stable
if ((vote 71 = vote 72) and (vote 71 = KVP or ARP or CHU) and (vote 77 = CDA)) VOTEVAR=stable
if ((vote 71 = missing) or (vote 72 = missing) or (vote 77 = missing)) VOTEVAR=missing
if ((vote 71 ≠ vote 72) and (vote 71 ≠ vote 77) and (vote 72 ≠ vote 77)) VOTEVAR='3 parties'
if ((VOTEVAR≠stable) and (VOTEVAR≠missing) and (VOTEVAR≠'3 parties')) VOTEVAR='2 parties'

47 tau-b (recall 2-1 by VOTEVAR) = .46
tau-b (recall 3-1 by VOTEVAR) = .62
tau-b (recall 3-2 by VOTEVAR) = .49

48 This hypothesis implies that while Recall 2-1 is inconsistent, Float 2-1 indicates stable voting. The responses on 1971 voting (vote 71, recall 72-71 and recall 77-71) can be thought of as a response pattern consisting of 3 items. If we use A to indicate the first of these three, and all recalls are consistent, then we have a

response pattern AAA. If the answers at all 3 times are different from each other (i.e. RECALL='3 answers'), then the pattern is ABC. In between these patterns lie those which belong to the RECALL category '2 answers'. These patterns are AAB, ABA and ABB. Looking at the inconsistencies this way we created a new variable REC123 indicating these response patterns. We can perform an analogous operation on the string of three answers indicating voting behavior at the three different elections (vote 71, vote 72 and vote 77). We termed this variable VOTE123, and here too response patterns are AAA, AAB, ABA, ABB and ABC.

As the two strings of three items have the first item in common (i.e. vote 71 which is used in the definition of REC123 as well as in the definition of VOTE 123) the party indicated by A in both strings is the same one. This is not necessarily the case with the B or C. For instance: if REC123=ABA and VOTE123=ABA, the answers to the three items comprising REC123 can be VVD-KVP-VVD, while the answers to the thre eitems forming VOTE123 can be VVD-PvdA-VVD.

The hypothesis mentioned in section 3.7 of this paper would imply that the response patterns of REC123 and VOTE 123 are identical, and furthermore that in those cases the meaning of the B and C symbols is the same too.

49 The complete crosstabulation between REC123 and VOTE 123 is as follows:

	VOTE123						
REC123	AAA	AAB	ABA	ABB	ABC	missing	total
AAA	205	32	17	7	6	2	269
AAB	7	*14*	9	14	18	3	65
ABA	8	1	*9*	5	4	—	27
ABB	—	7	5	*21*	10	—	43
ABC	—	—	4	4	*14*	—	22
missing	5	10	3	7	16	42	83
total	225	64	47	58	68	47	509

The frequencies underlined are potentially supportive of the hypothesis on page 28 of this paper. Listing these 58 cases with all variables contained in figure 2A, B, and C shows that in 36 of these cases inconsistent recall was identical with current vote (see also the last paragraphs of the previous note).

50 This kind of indicator of knowledge of 'what goes with what' captures part of what Converse (1964) refers to as constraint between cognitive elements.

50a We also used other methods of treating missing data in our analyses. The use of listwise deletion of missing data resulted in a great loss of cases, especially since there are 11 variables in the equation. Substitution of the mean for missing data keeps all respondents in the analysis, but it is not quite clear what effects it has on the substantial results of the analysis.

51 Literature on loglinear analysis (ordered from easy to advanced): Reynolds (1977), Everitt (1977), Bishop c.s. (1975).

52 We will use f to indicate observed frequencies and F for predicted frequencies.

53 Reynolds (1977) p. 57-58.

54 The fact that the parameters are estimated and not calculated exactly has to do with the intricacies of the algorithm used; see Dixon (1977).

55 Fienberg (1977), p. 37.

56 Reynolds (1977), p. 78.

57 See Dixon (1977).
58 See Brown (1976).

References

Abramson, P. R. (1976), Generational Change and The Decline of Party Identification in America: 1952-1974, *APSR,* 70, 469-478.

Achen, C. H. (1975), Mass Political Attitudes and The Survey Response, *American Political Science Review*, 69, 1218-1231.

Adamany, D., and Dubois, P. (1975), The 'Forgetful' Voter and An Underreported Vote, *Public Opinion Quarterly,* 39, 227-231.

Anastasi, A. (1958), *Differential Psychology,* New York: McMillan.

Arminger, G. (1976), Einführung in Loglineare Modelle zur Analyse von Nominal Skalierten Variabelen, *Format,* vol. 1.

Ascher, W., and Tarrow, S. (1975), The Stability Of Communist Electorates: Evidence From A Longitudinal Analysis Of French And Italian Data, *American Journal of Political Science,* 19, 475.

Asher, H. B. (1974), Some Consequences of Measurement Error in Survey Data, *American Journal of Political Science,* 18, 469-485.

Axelrod, R. (1973), Schema Theory: An Information Processing Model Of Perception and Cognition, *APSR,* 67, 1248-1267.

Bartlett, F. C. (1932), *Remembering,* Cambridge: Cambridge Press.

Bell, C. G., and Buchanan, W. (1966), Reliable and Unreliable Respondents: Party Registration and Prestige Pressure, *Western Political Quarterly,* 19, 37-43.

Benewick, R. J., Birch, A. H., Blumler, J. G., and Ewbank, A. (1969), The Floating Voter and The Liberal View of Representation, *Political Studies,* 17, 177-195.

Benham, T. W. (1965), Polling For A Presidential Candidate: Some Observations on the 1964 Campaign, *Public Opinion Quarterly,* vol. 29.

Benney, M., Gray, A. P., and Pear, R. H. (1956), *How People Vote,* London: Routledge and Kegan Paul.

Bijnen, E. J., and Hagenaars, J. A. P. (1971), De Tweede-Kamer verkiezingen, *Sociale Wetenschappen,* 14.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975), *Discrete Multivariate Analysis: Theory and Practice,* MIT Press, Cambridge.

Blondel, J. (1965), *Voters, Parties and Leaders,* Hammondsworth: Penquin.

Booth, A. (1970), The Recall of News Items, *Public Opinion Quarterly,* 34, 604-610.

Brown, M. B. (1976), Screening Effects in Multidimensional Contingency Tables, *Applied Statistics,* vol. 25.

Budge, I., Crewe, I., and Farlie, D. (1976), *Party Identification and Beyond. Representations of Voting and Party Competition,* London, Wiley.

Burnham, W. D. (1974), Theory and Voting Research: Some Reflections on Converse's 'Change in the American Electorate', *APSR,* 68, 1002.

Butler, D. E., and Stokes, D. (1969), *Political Change in Britain: Forces Shaping Electoral Choice,* London: McMillan.

Cahalan, D. (1968), Correlates of Respondent Accuracy in the Denver Validity Survey, *Public Opinion Quarterly,* 32, 607-621.

Cain, B. E. (1978), Dynamic and Static Components of Political Support in Britain, *American Journal of Political Science,* 22, 849-866.

Campbell, A., Converse, P. E., Miller, W. E., and Stokes, D. E. (1960), *The American Voter,* New York: Wiley.

Carlsson, G. (1965), Time and Continuity in Mass Attitude Change: The Case of Voting, *Public Opinion Quarterly,* vol. 32.

Clausen, A. R. (1968), Response Validity: Vote Report, *Public Opinion Quarterly,* 32, 588-606.

Converse, P. E. (1964), The Nature of Belief Systems in Mass Publics, in: Apter, D. (ed.), *Ideology and Discontent,* New York: Free Press.

Converse, P. E. (1966), Information Flow and the Stability of Partisan Attitudes, in: Campbell, A., e.a., *Elections and the Political Order,* New York: Wiley.

Converse, P. E. (1969), Of Time and Partisan Stability, *Comparative Political Studies,* 2, 139-171.

Converse, P. E. (1976), *The Dynamics of Party Support. Cohort Analyzing Party Identification,* Sage Library of Social Research 35.

Converse, P. E., Campbell, A., Miller, W. E., and Stokes, D. E. (1961), Stability and Change in 1960: A Reinstating Election, *APSR,* 55.

Cutler, N. E. (1970), Generation, Maturation and Party Affiliation: A Cohort Analysis, *Public Opinion Quarterly,* 33, 583-588.

Daalder, H. (1966), The Netherlands; Opposition in a Segmented Society, in: Dahl, R. A., *Political Opposition in Western Democracies,* Yale Un. Press, New Haven.

Daudt, H. (1961), *Floating Voters and The Floating Vote,* Leiden: Stenfert Kroese.

Daudt, H., Maesen, C. E. van der, and Mokken, R. J. (1968), Political Efficacy. A Further Exploration, *Acta Politica,* 3, 286-307.

Dixon, W. J., c.s. (1977), *BMDP: Biomedical Computer Programs,* P Series, University of California Press.

Dobson, D., and St. Angelo, D. (1975), Party Identification and The Floating Vote: Some Dynamics, *APSR,* 69, 481-490.

Dreyer, E. (1972), Media Use and Electoral Choice: Some Political Consequences of Information Exposure, *Public Opinion Quarterly,* 35, 544-553.

Eijk, C. v. d., and Niemöller, B. (1978), *Recall. On the hazards involved in using recall data to analyse voting dynamics,* Paper presented to the Dutch National Workgroup for Election Research.

Eijk, C. v. d., and Niemöller, B. (1979-a), *Dynamics of recalled vote, partisanship, partyidentification and voting behavior- or- Panta Rei,* Paper presented at Joint Sessions of Workshops, European Consortium for Political Research, Brussels, 1979.

Eijk, C. v. d., and Niemöller, B. (1979-b), *Dynamiek van partijpreferenties – een inventarisatie van frekwentie en omvang van veranderingen,* Paper gepresenteerd op het 'Politicologenetmaal', Amersfoort, 1979.

Elcock, H. (1976), *Political Behaviour,* London: Methuen.

Erikson, R. S. (1979), The SRC Panel Data and Mass Political Attitudes, *British Journal of Political Science,* 9, 89-114.

Everitt, B. S. (1977), *The Analysis of Contingency Tables,* Chapman and Hall, London.

Falter, J. W. (1977), Einmal Mehr: Läst sich dass Konzept der Parteiidentifikation auf Deutsche Verhältnisse übertragen? In: Kaase, M., *Wahlsoziologie Heute. Analysen aus Anlass der Bundestagswahl. Pol. Viert. Jahr.,* 18.

Fienberg, S. E. (1977), *The Analysis of Cross-Classified Categorical Data,* MIT Press, Cambridge.
Gamson, W. A. (1971), Political Trust and its Ramifications, in: Abcarian, G., and Soule, J. W. (eds.), *Social Psychology and Political Behavior: Problems and Prospects,* Columbus (Ohio): Merrill.
Glenn, N. D., and Hefner, T. (1972), Further Evidence on Aging and Party Identification, *Public Opinion Quarterly,* 36, 31-47.
Glenn, N. D. (1969), Aging, Disengagement, and Opiniation, *Public Opinion Quarterly,* 33, 17-33.
Glenn, N. D. (1972), Sources of the Shift to Political Independence: Some evidence from a Cohort Analysis, *Social Science Quarterly,* 494-519.
Hagenaars, J. (1973), Betrouwbaarheid van meting van achtergrondvariabele d.m.v. interviews: omvang, oorzaken en konsekwenties, *Sociale Wetenschappen,* 16.
Hedlund, R. D. (1977), Cross-over Voting in a 1976 Open Presidential Primary, *Public Opinion Quarterly,* 41, 498-514.
Hilgard, E. R., c.s. (1971), *Introduction to Psychology,* 5th edition, Harcourt, Brace, Jovanovitch.
Himmelweit, H. T., Jaeger Biberian, M., and Stockdale, J. (1978), Memory for Past Vote: Implications of A Study of Bias in Recall, *British Journal of Political Science,* 8, 365-384.
Iyengar, S. (1973), The Problem of Response Stability: Some correlates and Consequences, *Midwest Journal of Political Science,* vol. 17.
Iyengar, S. (1974), Magnifying Relationship Between Attitudinal Variables, *Public Opinion Quarterly,* 38, 90-95.
Kaase, M. (1967), *Wechsel von Parteipräferenzen. Eine Analyse am Beispiel der Bundestagswahl 1961,* Verlag Anton Hain.
Kamstra, O. W. M. (1971), *De dimensionaliteit van het geheugen,* Amsterdam: Doctoral Dissertation.
Kelley, S. Jr., and Mirer, T. W. (1974), The Simple Act of Voting, *APSR,* 68, 572-591.
Knoke, D., and Hout, M. (1974), Social and Demographic Factors in American Political Party Affiliations, 1952-1972, *American Sociological Review,* 39, 700-713.
Kritzer, H. M. (1978), An Introduction to Multivariate Contingency Table Analysis, *American Journal of Political Science,* vol. 22.
Luria, A. R. (1973), *The Working Brain. An Introduction to Neuropsychology,* London: Penguin.
Lijphart, A. (1974), The Netherlands: Continuity and Change in Voting Behavior, in: Rose, R. (ed.), *Electoral Behavior. A Comparative Handbook,* The Free Press.
Lijphart, A. (1975), *Verzuiling, pacifikatie en kentering in de Nederlandse politiek,* Amsterdam.
Macaluso, T. F. (1977), Political Information, Party Identification and Voting Defection, *Public Opinion Quarterly,* 41, 255-260.
Marsh, A. (1977), *Protest and Political Consciousness,* Sage Library of Social Research 49.
Mokken, R. J. (1971), *A Theory and Procedure of Scale Analysis. With Applications in Political Research,* Den Haag: Mouton.
Mughan, A. (1979), Party Identification, Voting Preference and Electoral Outcomes in Britain, *British Journal of Political Science,* 9, 115-128.
Muller, E. N. (1972), A Test of a Partial Theory of Potential for Political Violence, *APSR,* 66, 928-959.
Neuman, W. R. (1976), Patterns of Recall Among Television News Viewers, *Public Opinion Quarterly,* 40, 15-123.
Nie, N. H., Verba, S., and Petrocik, J. R. (1976), *The Changing American Voter,* Cambridge (Mass.): Harvard University Press.
Paige, J. M. (1971), Political Orientation and Riot Participation, *American Sociological Review,* 36, 810-820.
Parry, H. J., and Crossley, H. M. (1950), Validity of Responses to Survey Questions, *Public Opinion Quarterly,* 14, 61-80.
Pateman, C. (1970), *Participation and Democratic Theory,* Cambridge: Cambridge University Press.
Pedersen, M. N. (1979), The Dynamics of European Party Systems: Changing Patterns of Electoral Volatility, *European Journal of Political Research,* 7, 1-26.
Powers, E. A., Goudy, W. J., and Keith, P. M. (1978), Congruence Between Panel Data and Recall Data in Longitudinal Research, *Public Opinion Quarterly,* 42, 380-389.
Pulzer, P. G. J. (1967), *Political Representation and Elections in Britain,* London: Allen & Unwin.
Reynolds, H. T. (1977), *Analysis of Nominal Data,* Sage, Beverly Hills.
Riedel, J. A., and Dunne, J. R. (1969), When the Voter Decides, *Public Opinion Quarterly,* 33, 619-621.
Schreiber, E. M. (1975), Dirty Data in Britain and the USA: The Reliability of 'Invariant' Characteristics Reported in Surveys, *Public Opinion Quarterly,* 39, 493-506.
Schulman, M. A., and Pomper, G. M. (1975), Variability in Electoral Behavior: Longitudinal Perspectives from Causal Modeling, *American Journal of Political Science,* 19, 1-18.
Shively, W. P. (1972), Party Identification, Party Choice and Voting Stability: The Weimar Case, *APSR,* 66.
Sudman, S., and Bradburn, N. M. (1974), *Response Effects in Surveys,* Chicago: Aldine.
Taylor, D. G. (1976), The Accuracy of Respondent Coded Occupation, *Public Opinion Quarterly,* 40, 245-.
Thomassen, J. (1976), Party Identification as a Cross-National Concept: Its Meaning in the Netherlands, in: Budge, I, Crewe, I., and Farlie, D., *Party Identification and Beyond. Representations of Voting and Party Competition,* London: Wiley.
Thomassen, J. J. A. (1974), De betrouwbaarheid van vragen over kiesgedrag, *Acta Politica,* 9, 398-412.
Tingsten, H. (1937), *Political Behavior: Studies in Electoral Statistics,* London: King.
Tittle, C. R., and Hill, R. J. (1967), The Accuracy of Self-reported Data and Prediction of Political Activity, *Public Opinion Quarterly,* 31, 103-109.

Trenaman, J., and McQuail (1961), *Television and the Political Image*, London: Methuen.
Verba, S., and Nie, N. H. (1972), *Participation in America, Political Democray and Social Equality*, New York: Harper & Row.
Weir, B. T. (1975), The Distortion of Voter Recall, *American Journal of Political Science*, 19, 53-62.
Weiss, C. H. (1968), Validity of Welfare Mothers' Interview Response, *Public Opinion Quarterly*, 1968, 622-633.
Zukin, C. (1977), A Reconsideration of the Effects of Information on Partisan Stability, *Public Opinion Quarterly*, 41, 244-254.

De secretaris-generaal: politisering of verambtelijking? Een empirisch onderzoek naar sociale en politieke kenmerken van secretarissen-generaal 1945-1979

U. Rosenthal

1 Inleiding

Er is in Nederland weinig empirisch onderzoek gedaan naar sociologische en andere kenmerken van de top van het ambtelijk apparaat. Van Braam onderzocht de recrutering van hogere ambtenaren (referendarisschaal 130 en hoger) in 772 vacante posities (1968-1969).[1] Rosenthal en Van Schendelen verzamelden gegevens over de ambtelijke top vanaf het niveau van plaatsvervangend directeur (minimaal schaal 150). Hun onderzoek betrof 368 personen en had betrekking op de stand van zaken in 1975. Evenals Van Braam onderzochten zij de recruteringskanalen voor topposities in het ambtelijke apparaat. Zij konden globaal Van Braams constatering beamen dat naarmate een positie een hogere rang betrof, vaker intern gerecruteerd werd. Daarnaast presenteerden zij gegevens omtrent de periode van benoeming, de leeftijdsopbouw, de rangstructuur, en departementale overeenkomsten en verschillen.[2]

Een nog uitgelezener gezelschap topambtenaren werd door Kooiman, deels samen met Hubée-Boonzaaijer en Eldersveld, tot onderzoekspopulatie gekozen. Kooiman vergaarde gegevens over de sociale en politieke samenstelling van het corps directeuren en directeuren-generaal (1973). Zijn populatie betrof een gestructureerde steekproef van 75 functionarissen uit een totaal van meer dan 200 personen. Uit het onderzoek kwam naar voren dat 41% van de groep tussen de 47 en 57 jaar en 44% tussen de 57 en 66 jaar oud was. Slechts 15% had geen universitaire opleiding gehad. 38% had rechten gestudeerd, 23% economie en 21% natuurwetenschappen. Van de religieuze richtingen was de Nederlands Hervormde Kerk het best vertegenwoordigd: bijna éénderde, bij 20% Katholieken en 8% Gereformeerden. Eénderde van de respondenten had geen band met een religieuze richting of weigerde zich over zijn voorkeur uit te laten. Tenslotte wist Kooiman gegevens te vergaren aangaande de partijpolitieke voorkeuren. Ruim een kwart voelde zich het meest tot de VVD aangetrokken, en 6% tot DS'70. De drie grote confessionele partijen werden genoemd door 20% (13% KVP, 6% CHU, 1% ARP). De Partij van de Arbeid 'behaalde' 17%. Eén op de vier directeuren(-generaal) noemde geen voorkeur – om de een of andere reden.[3]