

The Meanings of ‘Yes’ and ‘No’. An Explanation for the Forbid/Allow Asymmetry

B. C. HOLLEMAN

Utrecht Institute of Linguistics OTS, Utrecht University, Trans 10, 3512 JK Utrecht, The Netherlands. E-mail: bregje.holleman@let.uu.nl

Abstract. Survey questions worded with the verb ‘forbid’ prove not to elicit opposite answers to equivalent questions worded with the verb ‘allow’ (Rugg 1941). Although ‘forbid’ and ‘allow’ are generally considered each other’s counterparts, respondents rather answer ‘no, not forbid’ than ‘yes, allow’. In order to find out which question is a more valid measure of the underlying attitude, this asymmetry in the answers has to be explained. Experiments show that the asymmetry arises because respondents translate similar attitudes differently into the answering options: forbid/allow questions are equally valid, but the way the attitudes are expressed on the answering scale differs due to the use of ‘forbid’ or ‘allow’. How does this translation process work? The leading hypothesis in forbid/allow research predicts that respondents holding moderate opinions feel that ‘yes forbid’ and ‘yes allow’ are very extreme, causing moderate respondents to prefer answering ‘not forbid’, or ‘not allow’. This article presents the results of 10 experiments investigating the meanings of the answering options to forbid/allow questions. Extreme connotations are shown to only provide part of the explanation for the occurrence of the forbid/allow asymmetry. In order to describe the answering process for forbid/allow questions, well-definedness of meanings proves to be an important additional factor. The meanings of answering options to allow questions are ill-defined compared of those to forbid questions, which causes allow questions to be less homogeneous measures of the underlying attitude than forbid questions.

Key words: question wording effect, question/answering process, nominal answering scales, interval scales, validity, reliability, survey research, questionnaire development, split-ballot design

1. Introduction

Research has repeatedly demonstrated that small changes in the wording of a question cause huge differences in the responses obtained (see Schuman and Presser, 1981/1996; Molenaar, 1982; Jobe and Mingay, 1991 for a review). This raises questions about the validity of survey questions: which particular question wording measures what the questionnaire designer intends to measure? The basic goal of research into wording effects is traditionally to generate practical advice for questionnaire design (Billiet, 1989), but equally important is the more fundamental goal to gain

insight into the cognitive processes underlying question answering, and in the variables that affect responses (Cicourel, 1982; Jobe and Mingay, 1991). Furthermore, this latter goal has to be fulfilled before questionnaire designers can be advised: in order to be able to advise on question wording, we have to know how wording variation causes differences in responses and in precisely what respect two seemingly equivalent questions measure something different.

A wording effect that has received a lot of attention for more than half a century of research is the forbid/allow asymmetry, identified by Rugg (1941). He found that respondents were more likely to support freedom of speech when the question was worded with the verb 'forbid', resulting in a difference of 21% between answers to two questions that seem logically equivalent (see Table I).

This finding was important, because many opinion polls inquire into moral issues such as abortion, or the use of marijuana – and most of such issues can be phrased in terms of whether a particular phenomenon should be 'forbidden' or 'allowed' (Schuman and Presser, 1981/1996). Furthermore, questionnaire designers are usually recommended to obtain balance in their questionnaires by varying positively and negatively phrased questions about the same issue. By applying this variation, the researcher avoids imposing only one perspective (either positive or negative) upon the respondents and can detect answering tendencies, such as yea-saying, at the same time. This recommendation, however, presupposes that the answers to negative questions are straightforward opposites to the answers obtained from their positive counterparts. Rugg's results show that this is not necessarily the case.

Since Rugg's finding, the forbid/allow effect has been replicated in the US (Schuman and Presser, 1981/1996; Bishop et al., 1988), Germany (Hippler and Schwarz, 1986, Bishop et al., 1988; Reuband, 2000), Belgium (Waterplas et al., 1988) and in the Netherlands (Holleman, 1999b, 2000). The experiments cover a wide range of issues, and use a variety of administration modes (phone, face-to-face or self-administered questionnaires). Very often a significant difference between responses to forbid/allow questions is found.

Table I. Forbid/allow experiment reported by Rugg (1941)

Do you think the United States should forbid public speeches against democracy?	
Yes, forbid	54%
No, not forbid	46%
Do you think the United States should allow public speeches against democracy?	
Yes, allow	25%
No, not allow	75%

The most important and general explanation for the forbid/allow asymmetry offered is the *connotations hypothesis* (Schuman and Presser, 1981/1996), which focuses on the extreme connotations of ‘forbid’ and ‘allow’. Schuman and Presser (1981/1996:280) write: “the former sounds harsher and may therefore be more difficult to endorse, whereas the latter in some contexts might seem to encourage a deviant behavior and therefore might invite opposition”.

The connotations hypothesis was tested once, in an experiment by Hippler and Schwarz (1986). They asked respondents to rate the extremeness of the fictitious opinion that “Mr X should forbid (not allow, allow, not forbid) peep shows” on a bipolar scale. ‘Allowing’ and ‘forbidding’ were judged to be more extreme than ‘not forbidding’ and ‘not allowing’ respectively. Differences were small however, and there was no doubt that ‘not forbidding’ is seen as a statement in favor of the issue, and ‘not allowing’ as a statement in opposition of it (Hippler and Schwarz, 1986:91).

In Hippler and Schwarz’s test, respondents were asked to rate the extremeness of someone else’s opinion, whereas it would be more insightful to know how respondents translate their *own* opinions into the bipolar scales to forbid/allow questions. The connotations hypothesis would predict that respondents holding moderate opinions against an issue, should logically answer ‘yes forbid’ but will answer ‘no, not forbid’ instead, because they feel the former option to be too extreme to reflect their moderateness. So a ‘yes’ to a forbid question would only reflect extreme con attitudes, whereas a ‘no’ can either reflect an extreme pro attitude or a moderate con attitude. Of course, similar predictions concerning the meanings of ‘yes’ and ‘no’ to allow questions would hold.

Obviously, these predictions about the way respondents translate their attitudes into response options cannot be tested by a comparison between the percentages of ‘yes’ and ‘no’ answers to forbid/allow questions. Somehow, respondents’ yes/no-answers have to be related to the true attitudes respondents hold. Before such an operation makes sense, however, a central assumption of the connotations hypothesis has to be checked – being that forbid/allow questions measure similar attitudes. This was done in two experiments (Holleman, 1999b).

1.1. FORBID/ALLOW QUESTIONS MEASURE SIMILAR ATTITUDES ...

According to general models of the question–answering process of attitude questions, the task of answering this type of question can be divided into four stages. First, respondents have to interpret the question, second they locate the relevant attitude structure, third they retrieve the attitude from long-term memory (or form an attitude), and, last, they map their judgement onto one of the precoded answering categories (Tourangeau and Rasinski,

1988; Sudman et al., 1996; Tourangeau et al., 2000). According to Krosnick and Schuman (1988), research on attitudes and response effects suggests a number of reasons why reported attitudes can be ‘distorted’. One possible approach is to view response effects as a result of slight changes in perceptions of the meanings of the answering options, as is done by the connotations hypothesis. In this case, the focus of the explanation lies on the last stage of the answering process. Answering a question on X-rated movies would mean that respondents retrieve or form a general attitude towards X-rated movies (similar for both the forbid and for the allow question). While mapping this evaluation onto the response options, the meaning of ‘no’ to ‘forbid’ comes to differ slightly from the meaning of ‘yes’ to ‘allow’ (and similarly for ‘no’ to ‘allow’ and ‘yes’ to ‘forbid’). Another approach is to explain the occurrence of response effects from the clarity and accessibility of attitudinal cues in memory. In this case, the focus is on the stages of attitude localization and retrieval (or formation). Respondents who are faced with ambiguous, conflicting or inaccessible internal cues, will be influenced in their perception of what their own attitude is by the wording of the question. In this case, respondents holding weak or uncrystallized attitudes might not only retrieve or form an attitude towards X-rated movies, but would also form or retrieve some general attitude towards forbidding (or allowing) things, due to the strong connotations of these words.

The connotations hypothesis assumes that respondents retrieve similar attitudes when answering forbid/allow questions – the forbid/allow asymmetry arises at the stage when they have to translate this attitude into a ‘yes’ or a ‘no’. In two experiments (Holleman 1999b) it was tested whether or not this assumption was appropriate. Correlational split-ballot designs were used in which analyses were conducted across sets of forbid questions and equivalent allow questions that intended to measure the same attitude. If two intelligence tests measure intelligence, their scores should correlate highly. The same reasoning was applied to attitude questions: if sets of forbid and allow questions measure the same attitude, the answers obtained to a set of forbid questions should correlate highly with the answers obtained to an equivalent set of allow questions.¹ The results of both experiments showed that forbid/allow questions measure similar attitudes, both for respondents holding strong attitudes as well as for those holding weaker attitudes. The answers to forbid questions differed from the answers obtained for equivalent allow questions though, confirming that something in the mapping stage causes the forbid/allow asymmetry.

1.2. ... BUT THEN WHY DO THE ANSWERS TO FORBID/ALLOW QUESTIONS DIFFER?

Unfortunately, these two experiments did not reveal equivocal patterns in respondents’ mapping behavior. Above the question level, at the level of

true scores, forbid/allow questions measure similar attitudes. On the level of questions, however, no consistent patterns could be detected as to how respondents translate their attitude into a 'yes' or 'no' when answering forbid questions or allow questions. This variation between questions is in line with earlier experiments: the asymmetry does not occur in each question, and seems to vary in size due to question contents or administration mode (Waterplas et al., 1988).² So working our way towards an explanation for the asymmetry, designs and analyses are needed that allow for generalizations above the level of specific question contents or administration modes.

Taking these design demands into account, this article focuses on two questions. The first, and main, goal is to gain insight into the mapping decisions respondents make in order to communicate similar attitudes. What do respondents mean when they answer 'yes forbid', 'not forbid', 'yes allow' or 'not allow'? Or in other words, how do these answers relate to respondents' true attitudes? It is difficult to give conclusive answers to these questions. However, it is possible to analyze what respondents do when answering forbid/allow questions that offer them more communicative scope to express their opinions. In the experiments that are reported here, insight will be gained into the meanings of 'yes' and 'no' by comparing the answers respondents give to two-points scale forbid/allow questions to the answers they give to equivalent seven-points scale questions. This way, the (assigned) meanings of the answering options to forbid/allow questions can be compared, as well as the variation in those meanings. Furthermore, this comparison provides the opportunity to test various predictions that follow from the connotations hypothesis.

A second goal of the research presented here, is to investigate whether the forbid/allow wording effect will disappear once respondents are offered more communicative freedom to express their attitude. As the asymmetry is probably caused by the extreme connotations of forbid and allow, affirmative answering options may also be regarded as extreme. The asymmetry may therefore disappear when more answering options are offered, allowing the expression of an intermediate position and more moderate positions on the answering scale.

2. Questionnaires and Experimental Design

Ten forbid/allow experiments were set up over a wide variety of issues and posed to respondents of various backgrounds. Each experiment had an identical split-ballot design and dealt with an issue that was a hot topic in public opinion in the Netherlands at the time of administration, such as new medical developments or road safety. Several different administration modes were applied in the experiments, although a written administration mode was the predominant method used. In several experiments, the

questions were posed during class at secondary schools. In these cases, the questionnaires were about topics relevant to secondary school students. In some cases, respondents were university students, while in others a general sample of people was drawn (e.g., from supermarket customers and passengers on public transport). For each experiment, the overall theme, and the number and type of respondents are given in Appendix A.

Each questionnaire developed for the experiments consisted of about 30 questions and of several forbid/allow questions. Each questionnaire began with a text in which a number of research goals were stated, intended to obscure the real goal of the experiments. This was followed by various questions on the subject of the questionnaire (see Appendix B for an example of the design and the forbid/allow questions included in the experiments). The experiments consisted of a number of forbid/allow questions, in most cases 6, with a dichotomous answering scale, separated by a variety of filler questions about related issues. In all but one experiment, the use of forbid/allow was varied over questions. Each questionnaire consisted of a second section, containing the same forbid/allow questions posed in the first section (in the same forbid or allow version) but accompanied by a seven-point scale. This second section was introduced to the respondents by stating that the researchers wanted to measure the opinions as accurately as possible and that some questions might therefore sound familiar.

Respondents were not given the opportunity to refer to the two-point-scale forbid/allow questions while answering the seven-point-scale questions, either because an interviewer administered the questionnaire, or because a researcher was present who asked respondents to hand in the first part of the questionnaire before starting with the second part. The order of the seven-points scale questions and the equivalent forbid/allow questions was not varied randomly in different questionnaires. This may have caused some question order effects due to memory: the answers to the seven-points scale questions may be influenced somewhat by the answers to the preceding forbid/allow questions. On the other hand, many questions were posed in between, and the data discussed later show no patterns that would indicate memory effects or some pressure for consistency (see Section 4).

Approximately 100 respondents were included in each experiment. Questionnaire versions were randomly distributed among the respondents, resulting in about 50 respondents per version. This design provided the opportunity for a comparison (a) between respondents, between the answers obtained for two-point-scale forbid questions and the answers obtained for two-point-scale allow questions, (b) between respondents, between the answers obtained for seven-point-scale forbid questions and the answers to seven-point-scale allow questions, and (c) within respondents, between the answers obtained for two-point-scale forbid questions and allow questions

with the answers from the same respondents to equivalent seven-point-scale forbid and allow questions. These comparisons were made for the entire sample of respondents ($N = 1054$) and for 61 forbid/allow questions with a two-point scale and 61 equivalent forbid/allow questions with a seven-point scale, on 10 different issues.

As noted earlier, previous forbid/allow research shows a large variation in the asymmetry size across questions and experiments. The design of the experiments presented here, provides opportunities to account for this variation. By the multilevel techniques used in the analyses, three sources of variation can be distinguished: variation between studies (due to different themes, different types of respondents and different administration modes), variation within studies between respondents (due to the different attitudes held by respondents) and a variation between questions (due to specific sub-issues addressed in each question, or to linguistic characteristics of a question). Variation in asymmetry sizes, or variation in attributed meanings to 'yes' and 'no', can be explained to some extent by differentiating it between the effects of experimental conditions and theme, the effects of different attitudes between persons, and the effects of question characteristics.

The next section will discuss the following analyses. First, a couple of analyses will be conducted in order to obtain a general impression of the data obtained, focusing on the randomization procedure of respondents across questionnaire versions within each experiment, and on the asymmetries found in each experiment separately and across experiments. It will be demonstrated that the key phenomenon, the asymmetry, is replicated in the current experiments, so it does make sense to explain the asymmetry based on these data. Second, it will be investigated whether the wording effect disappears once the answering scale offers more room to express moderate opinions. Third, the assigned meanings of 'yes' and 'no' to forbid/allow questions will be presented. The mean meanings (as expressed on the seven-point-scale forbid/allow questions), as well as the variation in these meanings (between respondents and between questions) will be discussed. The extent to which the connotations hypothesis can account for the results will prove to be limited, so additional explanations will be put forward. Last, but not least, the homogeneities (that is, the construct validity) of forbid/allow questions will be discussed. In each experiment, six forbid/allow questions were posed about one issue – so together the answers to this set of questions reflect some underlying attitude towards this issue. Now the focus of analysis is: which set of questions is a more reliable measure (in terms of their construct validity) of the underlying attitude: the set of three forbid questions about an issue, or the set of three equivalent allow questions about this issue? Differences in homogeneity will be related to differences in meanings of the answering options.

3. Results

Before the data for all experiments were analyzed simultaneously, a number of analyses were conducted for each experiment separately. These were performed to ensure that the randomization procedure assigning respondents to forbid/allow versions was adequate, to check the existence of the forbid/allow asymmetry via 'traditional' analysis and to obtain a general impression of the data obtained.

3.1. PRELIMINARY ANALYSES

Firstly, comparing the answers to the filler questions that were kept the same in both questionnaire versions checked the randomization procedure of respondents over versions in each experiment. It turned out that the groups of respondents who received different questionnaire versions did not differ in their answers to the filler questions, indicating that the differences found between forbid/allow answers could be interpreted as due to the manipulated question wording and not to differences in attitudes between the groups of respondents who answered the questions.

Secondly, χ^2 -square analyses (for each two-point-scale question) and *t*-tests (for each seven-point-scale question) were conducted in order to find out whether an asymmetry could be shown for the forbid/allow questions posed. In almost every experiment, 2 out of 6 dichotomous forbid/allow questions showed a significant wording effect.³ This is comparable to the proportion of significant asymmetries in the correlational experiments referred to earlier (Holleman, 1999b). The seven-point-scale questions showed about the same number of questions leading to a significant wording effect. Equivalent questions that displayed a significant wording effect accompanied by one type of answering scale did not necessarily show a significant difference when accompanied by the other type of answering scale, however.

Thirdly, the differences for forbid/allow questions in each experiment were computed and the ways in which these differences compared across different studies were assessed. Figure 1 shows that the answers obtained for forbid questions and for allow questions differ across experiments. This indicates that the answers differ due to the different themes in each experiment.

In Experiments 3, 7 and 8, there is a relatively large mean asymmetry for the two-point-scale forbid/allow questions, whereas in Experiment 6 a mean tendency towards more 'yes allow' than 'not forbid' is found (contrary to expectations). Experimental characteristics, such as time pressure (due to an oral versus a written administration mode) or the educational level of respondents do not seem to explain these variations. Experiment 8,

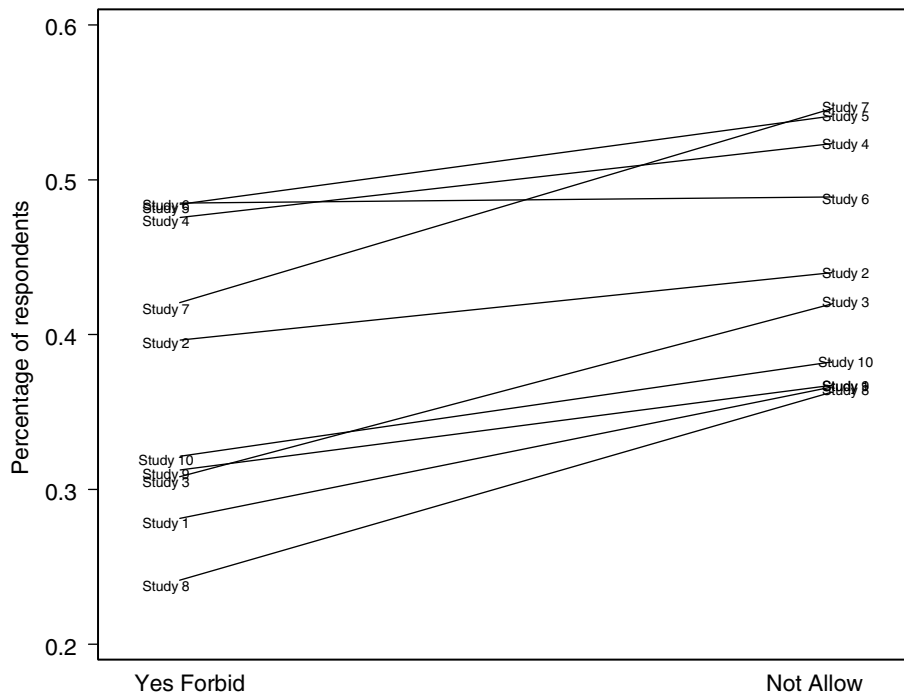


Figure 1. Differences between experiments in asymmetries obtained.

for example, shows a relatively large asymmetry, although it was administered by means of a written questionnaire to university students. Higher educational levels and written administration modes should be related to smaller asymmetry sizes, according to the literature (Schuman and Presser, 1981/1996; Bishop et al., 1988).

3.2. THE ASYMMETRY FOR YES/NO QUESTIONS

An analysis of the asymmetries per experiment shows that some questions displayed an asymmetry, while others did not. This stresses the importance of finding out whether there is an overall effect of the use of forbid/allow, as well as taking into account the variation in asymmetry size (or occurrence and non-occurrence of the asymmetry).

The mean answers to the two-point-scale forbid/allow questions of all 10 experiments show that there is a forbid/allow asymmetry: overall 44% of the respondents answered 'not allow', compared to 37% 'yes forbid' ($\chi^2 = 11.83$, $df = 1$, $p < 0.001$).⁴ This means that the asymmetry was replicated. Overall, the number of respondents answering 'not allow' is larger than the number of respondents answering 'yes forbid'. Accordingly, 'not forbid' is answered more often than 'yes allow'.

Table II. Mean asymmetry size and variances of the two-point-scale forbid/allow questions

	Forbid	Allow	χ^2	p
Overall percentage	Yes forbid: 37%	Not allow: 44%	11.8	< 0.001
Variances:				
Between studies	0.008 ($p < 0.001$)	0.005 ($p < 0.001$)	0.81	n.s.
Between persons	0.027 ($p < 0.001$)	0.023 ($p < 0.001$)	0.59	n.s.
Between questions	0.197 ($p < 0.001$)	0.217 ($p < 0.001$)	5.88	< 0.001

Note: χ^2 indicates the difference between the variance of ‘forbid’ and ‘allow’ at the different levels ($df = 1$), and p the significance of this difference.

How does the asymmetry vary across the different experiments? In Table II the variances of forbid and allow questions are provided for each level.

In the first row, the *variances between studies* are provided. The variance of forbid questions (0.008) is significant between studies, as is the variance of allow questions (0.005). Hence, the mean score for forbid questions differs across studies; as does the mean score of allow questions. This has already been illustrated in Figure 1. This significant variance for the forbid/allow questions is not surprising: the answers to forbid as well as allow questions vary according to the different themes (from road safety to new medical techniques) that were addressed in each experiment, due to different respondents (from rail travelers at Utrecht Central Station to secondary school pupils) who took part in the different experiments, and due to differences in the administration modes for each experiment.

For this study, it is more insightful to look at the *differences* in variances between forbid questions and the variance between allow questions than at each of the components separately. These do not differ significantly between studies ($\chi^2 = 0.81$, $df = 1$, n.s.). The wording effect does not differ much due to administration methods (oral versus written), types of respondents (e.g., secondary school pupils versus rail travelers) or the issues addressed (e.g., new medical technologies versus road safety). Hence, the differences in asymmetry sizes caused by experimental conditions (as depicted in Figure 1) cannot be shown statistically, which is probably partly due to the limited number of observations on the level of experiments.

As discussed earlier, differences in forbid/allow answers between persons over questions and differences within persons between questions can be distinguished within the differences in forbid/allow answers between studies. These are informative as well, because they indicate the extent to which forbid/allow answers vary due to respondent characteristics and due to question characteristics.

Table II shows that the *variance between persons* (within studies) is significant for forbid as well as allow questions. Due to differences between respondents, the answers to a set of forbid questions as well as to a set of allow questions differ from the mean answer to the forbid (or allow) questions for that experiment. This is not surprising, as one would expect the differences in attitudes respondents hold to be reflected in the answers they give.

The variance of forbid questions between persons does not differ significantly from the variance of allow questions, indicating that the differences between persons answering forbid questions are equal to the differences between persons who answered the same set of questions in an allow version. Assuming that respondents differ in their opinions, those differences are assessed equally well using forbid and allow questions.

The *variance between questions* (within persons and within studies) is significant for forbid questions as well as for allow questions (Table II). Due to differences in the specific sub-issues the questions address (and other characteristics specific to each question), the mean score of a specific forbid question differs from the mean of all forbid questions answered by a given respondent within an experiment and the same holds for allow questions. This variation for forbid as well as allow questions is not surprising either. It merely indicates that each question measures some specific issue or sub-issue compared to the other questions a respondent answered. It could indicate that the forbid/allow questions were not good representatives of the intended construct, i.e., that they did not measure one underlying construct or attitude, but it is rather premature to draw such conclusions at this point, as homogeneity analyses (see Section 5) can shed more light on this subject.

For now, it is important to note that the variations due to respondent and question characteristics are much larger than the variations due to experimental characteristics, implying that the latter are relatively unimportant when it comes to explaining or generalizing the forbid/allow asymmetry. In other words, the asymmetry size does vary over experiments, but the mean asymmetry size shows the wording effect can be generalized across experiments.

3.3. THE ASYMMETRY AND THE COMMUNICATIVE FREEDOM TO SAY MORE THAN JUST YES OR NO

A comparison of the mean answers to the seven-point-scale forbid/allow questions across all 10 experiments shows that a wording effect is present.⁵ For an asymmetry on the seven-point-scale questions, one would expect to find a tendency towards ‘disagreeing’ with forbid questions, rather than ‘agreeing’ with allow questions. This proved to be the case: the value

'1' corresponds with 'disagree strongly' with forbid questions and 'agree strongly' with allow questions (and the value '7' with 'agree strongly' with forbid questions, and 'disagree strongly' with allow questions). The overall mean answer to the forbid questions is 3.59, whereas the overall mean answer to allow questions turns out to be 3.94 ($\chi^2 = 10.55$, $df = 1$, $p < .001$).

More negative answers are given to seven-point-scale forbid questions than affirmative answers to allow questions of the same type. It can therefore be concluded that the wording effect is present for forbid/allow questions, even if answering scales that offer extended answering possibilities are used.

These results could be a reason to exclude the seven-point-scale questions from further analysis. However their inclusion will serve to provide an understanding of the mapping processes when answering forbid/allow questions.

4. The Meanings of 'Yes' and 'No'

How do respondents translate their true attitudes into a 'yes' or 'no' on the dichotomous answering scales? Do respondents who hold a moderate positive opinion towards an issue map their answer onto 'not forbid', as the connotations hypothesis would predict? The mapping process for forbid/allow questions and the mechanisms underlying the asymmetry will be analyzed by focusing on the assigned meanings of the answering options to forbid/allow questions. A post-hoc four-group design was constructed on the basis of the respondents' answers: respondents who answered 'yes forbid', respondents who answered 'not forbid', respondents who answered 'yes allow', and respondents who answered 'not allow'. The meanings of 'yes' and 'no' can be interpreted in terms of the seven-point-scale answers obtained, because the design of the experiments was such that each respondent answered several seven-point-scale questions equivalent to the two-point-scale questions. In other words, the seven-point-scale questions only differed from the two-point-scale questions in the number of scale points offered. Furthermore, preliminary analyses show that the test-retest reliability of forbid-2p and forbid-7p (as well as of allow-2p and allow-7p) is very high (0.99 and 0.87, respectively), indicating that respondents answered the questions far from randomly and that the questions measure the same attitude irrespective of the type of answering scale provided.⁶

Hence, a comparison of respondents' answers to the two-point-scale questions with their answers to equivalent seven-point-scale questions can provide an indication of what a 'yes' or 'no' to two-point-scale forbid/allow questions means. It indicates how the answering scales to forbid/allow

questions differ, and how the answers to forbid/allow questions should be interpreted.

Unfortunately, a straightforward comparison is not possible, due to the fact that the mapping problem did not disappear when a seven-point scale was offered. The wording effect found for the seven-point-scale forbid/allow questions shows that similar attitudes are translated differently into the seven-point agree–disagree scales due to the use of forbid/allow. The agree–disagree scale to forbid questions differs from the agree–disagree scale to allow questions.

A prerequisite for obtaining insight into the absolute meanings of ‘yes’ and ‘no’ to the two-point-scale questions is that the seven-point answering scales to forbid/allow questions are equivalent to each other. This proves not to be the case. Hence, an average meaning of ‘5’ for ‘yes allow’, as expressed on the seven-point allow scale, cannot be compared to an average meaning score of ‘2’ obtained for ‘not forbid’ as expressed on the different seven-point forbid scale. It would have been possible to compare these meanings in a straightforward manner using a different experimental set-up, in which the two-point-scale forbid/allow questions are followed by an equivalent seven-point scale that does not contain the verbs ‘forbid’ and ‘allow’. Since such a neutral scale was not available in this instance, some other way of obtaining comparable scales has to be found. Hence, the seven-point answering scales will have to be standardized.

Such a standardization can be obtained by transforming the observed scores to each seven-point-scale forbid/allow question to a standardized score.⁷ The answering scales to forbid/allow questions resulting from this standardization procedure can be compared because the midpoints of the scales become similar to each other. These midpoints will be referred to as ‘zero-points’, but in fact they reflect the mean opinion of respondents who answered forbid questions as well as the mean opinion of respondents who answered allow questions. As the forbid/allow questions were administered to two random subgroups of a random sample, and because forbid/allow questions were shown to measure similar attitudes, the mean opinion of respondents who answered allow questions can be assumed to be equal to the mean opinion of respondents who answered forbid questions. Hence, standardization of the answering scales provides the opportunity to compare the relative distances of the meanings of ‘yes’ and ‘no’ from the mean opinion expressed about the issues addressed in the questions posed in the current research. However, standardization of the answering scales does have some consequences for the variances around the mean scores for the meanings. For this reason, the analyses concerning the *variations* in meanings are referred to as an exploration.

The remainder of this section presents the analyses of the meanings of yes/no to forbid/allow questions. First, the meanings of the answers

to forbid/allow questions will be discussed. This will serve as a basis for explaining the asymmetry. But the mean meanings do not provide a full understanding of the mapping process underlying the asymmetry. Accordingly, the next subsection goes on to investigate the variation in the mean meanings. Then, the mean meanings and the variation in meanings will be interpreted, in an attempt to describe the mapping process underlying forbid/allow answers.

4.1. THE MEAN MEANINGS OF ‘YES’ AND ‘NO’

The mean meanings of the answers to forbid/allow questions are provided in Table III. This table shows that respondents answering ‘not forbid’ on the two-point scale also gave an answer indicating a considerable degree of disagreement to the seven-point-scale question (their mean score was 2.64 on the seven-points scale, on which ‘1’ meant ‘disagree with forbid’). The same holds for respondents who answered ‘yes allow’, as they gave answers indicating a considerable measure of agreement to the allow question on the seven-point scale.

Respondents who answer ‘not forbid’ or answer ‘yes allow’ express a mean opinion on the seven-point scale that also shows them to have a reasonably favorable attitude towards the issue in the question (i.e., against forbidding/in favor of allowing). A similar pattern is true for respondents who answered ‘not allow’ and ‘yes forbid’. They also express an opinion denoting considerable opposition to the issue in the question on the seven-point scale.

More importantly, Table IV shows that respondents who answered ‘not forbid’ to two-point-scale forbid questions obtain a mean score on the seven-point-scale forbid question that is -0.41 standard deviations away (i.e., to the left on the scale) from the mean answer to the seven-point-scale forbid questions. Respondents who answered ‘yes allow’ to a two-point-scale allow question also express an answer on the seven-point scale that is positioned to the left of the mean answer to seven-point-scale allow questions. However, compared to the meaning of ‘not forbid’, the meaning of ‘yes allow’ is positioned further to the left on the scale ($-.49$).

Table III. Distance of each answering option’s meaning from the midpoint of the scale in standard deviations

Answer on 2p	Not forbid	Yes allow	Not allow	Yes forbid
Distance from 7p midpoint	-0.41 (2.64)	-0.49 (2.76)	0.63 (5.36)	0.72 (5.26)

Note: Estimated mean scores for the meanings of the two-point-scale answering options in terms of the answers on the seven-point-scale are given between brackets.

Table IV. The variances between persons and questions (and their significance) in the meanings of the answers to the two-point-scale questions as expressed on the standardized seven-point scales

Answer 2 <i>p</i>	Not forbid	Yes allow	Not allow	Yes forbid
Variance between persons	0.19 ($p < .001$)	0.14 ($p < .001$)	0.14 ($p < .001$)	0.18 ($p < .001$)
Variance between questions	0.45 ($p < .001$)	0.51 ($p < .001$)	0.55 ($p < .001$)	0.60 ($p < .001$)

Statistical tests show that the meaning of ‘yes allow’ is more extreme than the meaning of ‘not forbid’, and that the meaning of ‘yes forbid’ is more extreme than the meaning of ‘not allow’ ($p < 0.05$). On a more conceptual level, this means that affirmative answers reflect more extreme opinions than negative answers to questions with the opposite wording: ‘yes allow’ is more extreme than ‘not forbid’, and ‘yes forbid’ is more extreme than ‘not allow’. This difference in the extremity for the meanings is in line with the connotations hypothesis, which states that forbidding and allowing both carry extreme connotations.

It is important to note that these differences in the extremity of ‘not forbid’ versus ‘yes allow’ and of ‘not allow’ versus ‘yes forbid’ explain the forbid/allow asymmetry. Due to the standardization of the seven-point-scale answers into *z*-scores, the answers to the seven-point-scale questions can be related directly to proportions (using a *z*-table). This makes it clear that the ‘no’ answers, which are less extreme than the ‘yes’ answers to the opposite versions, are likely to be chosen by more respondents: under the assumption of normal distributions, the mean opinion in the population and opinions close to that mean are likely to be held by more respondents than opinions further away from that mean (i.e., the midpoint of the scale). So, between versions, the number of respondents likely to choose ‘not forbid’ and ‘not allow’ is larger than the number of respondents likely to choose ‘yes allow’ or ‘yes forbid’, respectively. This is directly related to the key phenomenon to be explained in this study: the asymmetry in the percentages ‘yes’ and ‘no’ obtained to forbid/allow questions.

The difference in extremity of ‘not forbid’ compared to ‘yes allow’ and of ‘not allow’ compared to ‘yes forbid’ explains the higher percentage of respondents choosing that option, at least to some extent. However, it does not describe how the question–answering mechanism works within versions. If a respondent has to answer one question in either version, the respondent has to choose between ‘yes’ or ‘no’ and does not have the opportunity to compare forbid versions with allow versions. In order to obtain an insight into the question–answering process, the meanings of the answering options have to be compared *within* versions, instead of making a comparison of meanings between versions.

The connotations hypothesis describes forbidding and allowing as both carrying extreme connotations, causing moderate respondents to map their opinions onto 'no'. Within versions, it would therefore be interesting to know whether 'not forbid' is less extreme than 'yes forbid' and whether 'not allow' is less extreme in its meaning than 'yes allow'. The standardized meanings presented in Table III show that 'yes forbid' is indeed further away from the midpoint of the scale than 'not forbid' ($p < 0.05$). At the same time, however, 'not allow' is more extreme in its meaning than 'yes allow' ($p < 0.05$). Within versions, therefore, the mean meanings of the answering options to forbid questions do support the idea that a negative answer reflects more moderate opinions than the affirmative answer, but at the same time the mean meanings of the answering options to allow questions do not support this hypothesis.

One would expect the degree of extremity to be related to the amount of variance around the meaning: more extreme options are likely to vary less in their meanings, whereas a larger variance in meanings is expected to cause the less extreme meanings. The next section presents the variances in meanings of the answering options, which provides the opportunity to analyse whether this is the case.

4.2. THE VARIATION IN MEANINGS OF 'YES' AND 'NO'

Looking at the variation in meanings that respondents attribute to their forbid/allow answers provides a further understanding of the mapping process. Whereas the mean scores for the meanings show how the answer should be interpreted, the extent to which those meanings vary shows how well-defined these meanings are. The variances around the means will therefore be analyzed, whereas they give insight into the extent to which the meanings of 'yes' and 'no' vary within studies between persons, and within persons between questions.⁸

Table IV provides the variances around the mean meanings as expressed on the standardized scales.⁹ As can be seen in this table, the variances between persons are significant, showing that the mean meaning of 'not forbid' for a set of questions answered by one respondent differs from the mean meaning of 'not forbid' given by all respondents within a given experiment. The same holds for the mean meanings of the other three answering options compared to the mean meaning score for that answering option. This indicates that the meaning of a given answering option to a two-point-scale forbid/allow question is person dependent: for one person 'not forbid' may correspond with a mean score of '3' on the seven-point-scale forbid questions, for another person it may correspond with a mean score of '2' on the seven-point-scale questions.

The differences between the between-persons variances prove to be insignificant. The meanings of the two-point-scale answering options differ between persons, but the extent to which they differ is not related to the answering option: the meaning of ‘yes allow’ does not vary more between persons than the meaning of ‘not forbid’ or ‘not allow’.

The variances between questions within persons are significant for each of the answering options as well. This indicates that the meaning of a ‘not forbid’ answer to a specific question differs from the mean meaning of all questions to which a given respondent within an experiment answered ‘not forbid’. So the variances show that the meaning of each of the answering options to two-point-scale forbid/allow questions is partly dependent of the specific sub-issue addressed in a question.

Does the extent to which the meanings of the answering options to forbid/allow questions vary within persons over questions, differ systematically for one answering option compared to the other? The variance in meaning of ‘not allow’ between questions (0.55) proves to be similar to the variance in meaning of ‘yes allow’ (0.51), but the variance in meaning of ‘yes forbid’ between questions (0.60) is larger than the variance in meaning of ‘not forbid’ (0.45, $p < 0.05$). Between versions, the variance of ‘not allow’ (0.55) equals the variance in meaning of ‘yes forbid’ (0.60) between questions, but the variance in meaning of ‘yes allow’ (0.51) is larger than the variance in meaning of ‘not forbid’ (0.45, $p < 0.05$). The next subsection discusses to what extent the connotations hypothesis can account for the patterns found in the meanings of ‘yes’ and ‘no’ and the variances around those means.

4.3. THE CONNOTATIONS HYPOTHESIS

In the discussion of the mean meanings of the answering options to forbid/allow questions, the higher percentage for ‘not forbid’ compared to ‘yes allow’ (and for ‘not allow’ compared to ‘yes forbid’) was explained by the greater extremity of the meanings of ‘yes’. The meanings of ‘not forbid’ and ‘not allow’ are more moderate than the meaning of ‘yes allow’ and ‘yes forbid’, respectively, and those answering options are therefore likely to be chosen more often.

The next question is whether the relative moderateness of ‘no’ is caused by a larger variance in meaning of the two negative answering options, as the connotations hypothesis predicts. A comparison *between* versions shows that this is not the case, as the variance of ‘not forbid’ (0.45) is smaller than the variance of ‘yes allow’ (0.51), and the variance in meanings of ‘not allow’ (0.55) does not differ from the variance in meanings of ‘yes forbid’ (0.60). It can therefore be said that between versions, differences in the extremity of the answering options cannot be explained by a larger

variation in meanings of ‘not forbid’ compared to ‘yes allow’ and of ‘not allow’ compared to ‘yes forbid’.

A similar pattern is shown when the variances *within* versions are compared. The connotations hypothesis would predict that respondents holding moderate opinions prefer ‘not forbid’ to ‘yes forbid’ because forbidding is seen to be relatively harsh. Likewise, moderate respondents would prefer ‘not allow’ to ‘yes allow’, because allowing carries an extreme connotation of support for a behavior. This leads to the prediction that the negative answers to either version not only have a more moderate mean meaning than the affirmative answers to the same question version, but also that the variance in meanings of ‘no’ will be larger than the variance in meanings of ‘yes’ to the same version. Two kinds of respondents will answer ‘no’: those who really mean ‘no!’ (not forbid/not allow), and those who merely want to express ‘not yes’ (forbid/allow).

Within versions the differences in extremity cannot be explained by differences in the variances, however: for forbid questions, the mean meaning of ‘not forbid’ is less extreme than the mean meaning of ‘yes forbid’, but the variance in meanings of ‘yes forbid’ is larger. For allow questions, the mean meaning of ‘yes allow’ is more moderate, but the variances in meaning of ‘yes allow’ and ‘not allow’ do not differ. Hence, straightforward predictions based on the connotations hypothesis do not seem to explain the patterns in the variances found for the meanings of yes/no to forbid/allow questions. Another line of reasoning therefore needs to be established to explain why one of the less extreme answering options in terms of mean meaning (i.e., ‘not forbid’) also shows the smallest variation in meaning.

4.4. AN ALTERNATIVE DESCRIPTION OF THE MAPPING PROCESS

Why is the moderateness in the mean meaning of ‘not forbid’ not related to a larger variance in meanings of this answering option? The connotations hypothesis would predict that respondents who hold a moderate opinion against the issue in the question, will answer ‘not forbid’, finding ‘yes forbid’ too extreme and too harsh to cover their ‘true attitude’. An alternative explanation for the variances and mean meanings should be able to account for why the answering options to forbid questions show a pattern in variances and means that differs markedly from the patterns found for allow questions: for allow questions as opposed to forbid questions, the mean meaning of ‘yes (allow)’ is less extreme than the meaning of ‘no (not allow)’, and the variance in meanings of ‘yes allow’ equals the variance in meanings of ‘not allow’. In this subsection, it will be argued that the concept of extremity can be used to explain these patterns in variances, but that it must be applied differently than in the connotations hypothesis.

The larger variance in the more extreme answering options (as opposed to options that are moderate in their mean meaning scores) may be explained by the fact that respondents who map their opinion onto an answering option with an extreme meaning, feel a need to modify this extremity by choosing a moderate answering option on the seven-point scale. The extremity of the answering option 'yes forbid' is accordingly modified on the seven-point scale, causing a large variance in the meanings of 'yes forbid'. Respondents mapping their opinion onto the more moderate 'not forbid' do not need to modify their answer, because it is already relatively moderate. This could also explain the difference in variances between versions: 'yes allow' is more extreme than 'not forbid', so if respondents mapped their opinion onto 'not forbid', they feel less need to modify this moderateness when given the chance on a seven-point scale than respondents who mapped their opinion onto the relatively extreme 'yes allow'.

This gives rise to the question of whether the small variance in meanings of 'not forbid' compared to 'yes allow' should be interpreted as a positive or as a negative feature of forbid questions. The negative interpretation would be that respondents who map their opinion onto 'not forbid' assign a larger variety of meanings to the relatively small part of the seven-point scale they map their opinion onto than is reflected by the relatively homogeneous answers they give. In other words, respondents answering 'not forbid' may answer '2' relatively homogeneously on the seven-point-scale forbid question, but this score of '2' in fact reflects a variety of moderate and more extreme opinions against forbidding. In this interpretation, the seven-point forbid answering scale is less able to show subtle variation in opinions and differences between respondents, than the allow answering scale. In contrast to this negative interpretation for forbid questions, the positive interpretation would be that extremity in the meaning of a yes/no answering option is not in fact preferable, because respondents modify the extreme answering option 'yes allow' once they get the chance when offered an answering scale containing more scale points. This would imply that the more moderate option 'not forbid' reflects respondents' true opinions more consistently than the extreme 'yes allow', which reflects a variety of moderate and extreme opinions in favor of allowing.

The wording effect found for the seven-point-scale questions means there is no way to decide whether the negative or positive interpretation for forbid questions is true. However, there are a number of reasons to prefer the positive interpretation for forbid questions. The most important reason being that forbid questions prove to be more reliable, i.e., homogeneous, measures of the underlying attitude than allow questions. The next subsection will discuss this in more detail. First, the interpretation of the patterns in variances and mean meanings will be continued.

If the interpretation that favors forbid questions is followed, the answering option 'not forbid' reflects respondents' true attitudes more consistently than the answering option 'yes allow'. For the cognitive processes underlying the answering of forbid/allow questions, this means that the connotations hypothesis is correct in stating that 'forbid' and 'allow' carry extreme connotations, causing 'yes forbid' to be more extreme than 'not allow', and 'yes allow' to be more extreme than 'not forbid' (which, in turn, means that the former answering options are chosen by more respondents than the latter). Within versions, however, an additional cognitive mechanism seems to be at work, which is related to extremity as well as to the concept of 'well-definedness'.

For forbid questions, the answering option 'yes forbid' is relatively extreme, causing the meaning of 'not forbid' to be more moderate and better defined, as it is not necessary to modify a moderate answer. For allow questions, the question-answering process works according to a similar mechanism, but operates differently due to a difference in the extremity of the answering options to forbid and allow questions.

An explanation of the relative extremity of 'yes forbid' may be derived from the semantic properties of 'forbid' and 'allow'. The answering option 'yes forbid' is extreme, as it unambiguously refers to an act of inserting a barrier, to a restrictive action.¹⁰ The answering option 'not allow' refers to a similar action. Therefore, 'not allow' is extreme, and needs to be modified on the seven-point answering scale (which is indicated by a large variance). The answering option 'yes allow' is more ambiguous in its meaning than 'yes forbid', however. It can refer to an action of removing a barrier, which might in some contexts seem to imply that deviant behavior is encouraged, according to Schuman and Presser (1981/1996), but it can also refer to the non-action of not inserting a barrier (similar to the meaning of 'not forbid'). The large variance in meanings of 'yes allow' might therefore reflect a tendency of respondents to modify extreme answering behavior ('yes allow' in its meaning of remove a barrier), as well as an ambiguity in meanings of 'yes allow' (which also reflects the more moderate 'not inserting a barrier').

The variance in meanings of 'not allow' equals the variance in meanings of 'yes allow', due to the ambiguity of 'yes allow'. The ambiguity in one of the answering options to allow questions makes the other answering option ambiguous as well: 'not allow' is extreme for it refers to inserting a barrier and therefore requires modification, but at the same time 'not allow' can imply not remove a barrier. The fact that 'yes allow' is not very well-defined in its meaning, means that 'not allow' is not very well-defined either, as shown by equal variances in meanings of 'yes allow' and 'not allow'. The meanings of both answering options to allow questions therefore seem less well-defined than the meanings of the answering options to forbid questions.

Summarizing, this alternative explanation of the variances and mean meanings, suggests that the process by which respondents translate their true attitudes into the yes/no options, is an interaction between the extremity of respondents' true attitudes, and the extremity and the well-definedness of the answering options' meanings. The connotations hypothesis correctly explains why the extreme affirmative answering option to forbid questions is chosen less often than the more moderate negative answering option to allow questions. However, the concept of 'extremity' alone, does not sufficiently explain why the variation in meanings of 'not forbid' is smaller than the variation in 'yes forbid', and why the variation in meanings of 'yes allow' equals the variation in meanings of 'not allow'. The mapping decisions respondents employ when answering a forbid question or an allow question, seem to be described better if the notion of well-definedness is added.

The plausibility of this, post-hoc, explanation can be verified by an analysis of differences in the homogeneity of the forbid/allow questions. Is it a reasonable assumption that the more moderate option 'not forbid' reflects respondents' true opinions better than the extreme 'yes allow', which reflects a variety of moderate and extreme opinions in favor of allowing? In other words, is the decision to prefer the explanation that is favorable to forbid questions correctly made? If it is, one would expect that sets of forbid questions about one and the same issue are more homogeneous measures of the underlying attitude than sets of allow questions. If an attitude towards an issue is measured with several forbid questions tapping this construct, one would expect that a positive overall attitude would manifest itself by several 'not forbid' (or 'disagree to forbid') answers and a negative attitude by several 'yes forbid' (or 'agree to forbid') answers to this set of questions. If this homogeneity tendency can be found for forbid questions, but is less the case for allow questions, it would be reasonable to conclude that forbid questions are a better reflection of respondents' true attitudes than allow questions. It would mean that differences between respondents' true underlying attitudes can be assessed better by sets of forbid questions than by sets of allow questions, which would imply that the larger variation in meanings of the answers to allow questions are not reflecting variations in true attitudes, but are in fact artifacts of the allow measurement.

5. Relating the Mapping Process to the Reliabilities of Forbid/Allow Questions

The attitude respondents retrieve does not depend of whether a forbid question or an allow question was posed. This being the case, how well do clusters of forbid questions, or clusters of allow questions, reflect this underlying attitude? Or in other words, does the homogeneity of forbid questions and allow

questions differ? If it is true that the meaning of ‘forbid’ is more well-defined and more extreme, one would expect that the answers to forbid questions are a more homogeneous reflection of the underlying attitude than the answers to the ill-defined and moderate allow questions.

Table V shows that neither the set of forbid questions nor the set of allow questions are very homogenous measures: neither provides a particularly reliable measure of the underlying attitude (towards road safety, for example). Note, however, that this reliability index depends on the number of questions that were part of the scale. In this research, each scale that was supposed to measure one underlying attitude consisted of six questions, three of which were in an allow format, and three in a forbid wording. If 10, instead of 3, forbid questions had been posed to each respondent, the homogeneity would have gone up to 0.6, which is reasonable. However, instead of the similarities between forbid/allow questions, the differences between forbid and allow questions and between seven-point-scale and two-point-scale questions are more interesting for the current research.

Table V shows that seven-point-scale questions measure the attitude more reliably than two-point-scale questions, both in the forbid and allow version. This was to be expected, because by using a seven-point scale variances increase. In addition it can also be seen that both for two-point-scale questions and for seven-point-scale questions, forbid questions are a more reliable measure than allow questions. So forbid/allow questions measure the same construct, independent of question wording. But in terms of homogeneity, or cluster reliability, forbid questions lead to a more reliable measure of an underlying attitude than allow questions, both for seven-point-scale and two-point-scale questions.

When answering a set of two-points scale allow questions about an issue, a respondent is likely to answer some questions by circling ‘yes allow’ and some by indication ‘no, not allow’. In contrast, when answering a set of two-points scale forbid questions about the same issue, respondents are less likely to show this switching behavior. A similar difference in switching behavior shows for the seven-points scale questions, as for these

Table V. The homogeneity of forbid/allow questions

	Two-point-scale questions		Seven-point-scale questions	
	Forbid	Allow	Forbid	Allow
Homogeneity	0.31	0.26	0.42	0.31

Note: The differences between 0.31 and 0.26 is significant on 0.02, for the difference between 0.42 and 0.31 the p -value is 0.004

seven-points scales the homogeneity of allow questions is lower than the homogeneity of forbid questions as well.

Hence, it seems that the small variance in meanings of ‘not forbid’ (as expressed on the seven-points forbid scale) compared to the large variance in meanings of ‘yes allow’, noted in the previous section, should be interpreted as a positive feature: true attitudes are expressed more homogeneously when answering forbid questions than when answering allow questions. The reason for this conclusion is twofold.

First, an analysis of the answering behavior for seven-points scale forbid/allow questions showed that respondents use more scale points when answering seven-point-scale forbid/allow questions than when answering two-point-scale questions. It does not seem very plausible to assume that this is the case for all respondents, except for those respondents who answered ‘not forbid’ on the two-point-scale questions. There is no reason to assume that those respondents should all of a sudden choose a relatively small number of scale points on the seven-point scale to express an equally large amount of variation in true opinions.

Second, the reliability (i.e., homogeneity) of allow questions is consistently lower than the homogeneity of forbid questions. It seems that respondents may find it easier to answer allow questions: the meaning of ‘to allow’ is relatively ill-defined, causing a small cognitive distance between ‘yes allow’ and ‘not allow’. When answering allow questions respondents do not seem to feel very restricted by the meanings of ‘allow’: they can switch between ‘yes allow’ and ‘not allow’ when answering sets of allow questions about the same issue and switch in the meaning they attach to ‘yes allow’. By doing this, they can express the moderateness (or even non-existence) of their true attitude. When answering forbid questions their mapping problem becomes more evident, as the meaning of ‘yes forbid’ is relatively extreme and therefore well-defined. Respondents modify this extremity by moderating their answers on the seven-point scale. The fact that the meaning of ‘yes forbid’ is better defined than all of the other answering options, implies that the meaning of ‘not forbid’ is also well-defined. Because of its reference to a non-action, its meaning seems to be rather moderate, which might cause the ‘disagree to forbid’-side of the seven-point forbid scale to entail a rather moderate meaning as well. Respondents probably tend not to distinguish many subtleties within this moderate area on the seven-point scale, resulting in a relatively homogeneous moderate meaning of ‘not forbid’.¹¹

6. Conclusion and Discussion

Questions worded differently (i.e., with forbid or allow) do not cause different attitudes to be retrieved. This shows that question answering is a

purposeful act: respondents do not process the question text as a whole before they start answering the question. Rather, when answering ‘Do you think the government should forbid (allow) abortion’, they seem to extract the core of the question, ‘what do I think of abortion’, locate or form their attitude on abortion (as expressed on a psychological pro/con scale) and then decide how to translate this psychological evaluation into an answering option. The outcomes of this translation process are dependent of the qualifier in the survey question posed: a moderate pro opinion is mapped differently onto the yes/no answering scale if the verb ‘forbid’ was used, compared to the situation in which the question was worded with the verb ‘allow’. Hence, ‘yes’ and ‘no’ do not reflect a pro/con evaluation of ‘abortion’, but of ‘forbidding abortion’ or ‘allowing abortion’: ‘yes’ does not mean ‘yes’, but means yes + forbid, or yes + allow. This could be called the *Yes/no⁺ Theory*: the answering options ‘yes’ and ‘no’ do not have a nominal constant meaning, but are the extremes of an interval scale and acquire their relative meaning from the qualifier that was used in the question.

For the forbid/allow asymmetry, the explanation that emerges is that the asymmetry is caused by the well-defined nature and extremity of ‘yes forbid’ compared to the moderateness of ‘not allow’, as well as by the moderateness of ‘not forbid’ compared to the extremity of ‘yes allow’. In a given random population, less respondents are likely to hold extreme opinions – compared to the number of respondents holding moderate opinions – and that is why the more moderate answering options are likely to be chosen more often. This explains the higher number of respondents mapping their judgments onto ‘no’ between versions.

The smaller degree of extremity in the meaning of ‘no’ to both question wordings is caused by a collective assignment of moderate meanings, and not by a relatively large variance the meanings of moderate answers. So the connotations hypothesis is right in describing ‘forbid’ and ‘allow’, i.e., the answering options ‘yes forbid’ and ‘yes allow’, to be extreme, but the prediction following from the connotations hypothesis that moderate respondents map their opinion onto ‘no’ when answering forbid/allow questions is not confirmed by the results reported here.

Besides differences in extremity, another cause of the asymmetry is a difference in well-definedness of meanings of ‘forbid’ and ‘allow’. Due to the ambiguity of ‘allow’ in its reference to not inserting a barrier or removing a barrier (or in doing nothing with respect to an issue as apposed to encouraging it), the answering option ‘yes allow’ is not very well-defined compared to ‘not forbid’ (which only refers to not inserting a barrier). This ill-defined meaning of ‘yes allow’ also has a negative influence on the definition of ‘not allow’: its mean meaning is extreme, as it refers to an act of inserting a barrier, but its variance in meanings is large, as it also refers to an act of not removing a barrier. This ambiguity also seems related to

the fact that respondents can switch between 'yes' and 'no' when answering allow questions. Sets of allow questions are less reliable measures of the underlying attitude than forbid questions, due to less well-defined meanings of 'yes allow' and 'not allow' (compare to 'yes forbid' and 'not forbid'). The answering option 'yes allow' can refer to causing as well as to letting. 'Yes forbid', on the other hand, only refers to causing, to inserting a barrier. Due to the ill-defined meaning of 'yes allow', 'not allow' is also ambiguous in its meaning. This gives respondents more reason, as well as more scope for switching between 'yes' and 'no' when answering allow questions. The meaning of 'yes forbid' and 'not forbid' is well-defined, causing respondents to be relatively consistent in mapping their attitudes onto the answering scales when answering a set of forbid questions. This might cause respondents to encounter more of a mapping problem when answering forbid questions, but it also causes the answers to forbid questions to reflect the underlying attitude more consistently, thereby providing the possibilities to make a better distinction between the attitudes of different respondents. In addition, it makes the answering option 'not forbid' relatively easy to interpret for the researcher. The group of respondents answering 'not forbid' expresses a relatively moderate opinion by choosing that answer, but is fairly homogenous in its moderateness. What is more, respondents choosing 'not forbid' are relatively consistent when answering several forbid questions about the same issue, which accounts for the greater reliability of forbid questions.

Hence, survey designer should prefer forbid questions to allow questions, because of their greater reliability as well as because the answers obtained can be interpreted more easily and homogeneously. Furthermore, this research shows that seven-points scales should be preferred to two-points scales: they do not offer enough communicative freedom to solve respondents' mapping problems, but they are more reliable measures of the underlying attitude.

For future research, it is necessary to find out whether and to what extent the results presented here can be generalized to other question wordings, as this will provide a theoretical basis to the interpretation of wording effects. It seems plausible to assume that many question-wording effects are caused by mapping differences, and not by the retrieval of different attitudes. Usually, dichotomous yes/no answering scales are in fact interval scales, similar to Likert-type scales. The prediction following from this research would be that questions in which an extreme and well-defined qualifier is used, such as forbid questions, cause the meanings of 'yes' and 'no' to be extreme and well-defined also, which will result in more reliable measurements. This could be tested by checking whether the meaning of the answering options to questions containing other contrastive word pairs, such as 'break off' versus 'continue', show similar patterns compared

to those of ‘forbid’ and ‘allow’. Furthermore it could be looked into how the (Likert-type) answering scales’ meanings change once *contrary* questions are formulated, such as questions containing the words ‘good’ versus ‘bad’.

In the mean time, the implications of the results presented here are two-fold. For survey design, the Yes/No⁺ Theory suggests it is sensible to use several questions that vary in the qualifying dimension that is used to tap one underlying attitude, instead of a single question. By doing that, the meaning of the yes/no, or agree/disagree, answers is at least dependent of several different qualifiers, instead of only just one. Should an attitude or opinion be assessed using one question only, it is sensible to choose the question wording that is most extreme and well-defined in its meaning. This may cause larger mapping problems for respondents, but does offer optimal opportunities to interpret the answers back into some ‘true value’. Furthermore, it’s important to bear in mind not to interpret the answers beyond the actual question posed: respondents who answered that ‘X-rated movies should not be forbidden’ do not necessarily think that ‘X-rated movies should be allowed’ or that ‘the showing of X-rated movies is a good thing’.

Acknowledgements

The author wishes to thank dr. Huub van den Bergh for his thinking along and his help with the statistical analyses.

Appendix A: Description of the 10 Experiments

• Experiment 1

Theme: youth culture. *Subjects:* 100 secondary school pupils (HAVO – higher vocational education /VWO – pre-university education) in 4 different classes. *Procedure:* written questionnaires administered during class, versions randomized over students (but students sitting next to each other received the same version). Part two was completed after filling in part one of the questionnaire and turning it face down. *Total number of questions:* 32. *Number of forbid/allow questions:* 7.

• Experiment 2

Theme: medical choices. *Subjects:* 100 university students. *Procedure:* a written questionnaire administered in several university canteens. *Total number of questions:* 22. *Number of forbid/allow questions:* 6.

- **Experiment 3**

Theme: hooligans. *Subjects:* 100 secondary school pupils (HAVO – higher vocational education/VWO – pre-university education) in 4 different classes. *Procedure:* a written questionnaire administered during classes. *Total number of questions:* 32. *Number of forbid/allow questions:* 6.

- **Experiment 4**

Theme: supermarket policies. *Subjects:* 100 supermarket customers. *Procedure:* oral questionnaire administered to people entering a supermarket. *Total number of questions:* 35. *Number of forbid/allow questions:* 6.

- **Experiment 5**

Theme: euthanasia and organ donation. *Subjects:* 118 secondary school pupils (VWO – pre-university education). *Procedure:* written questionnaires administered during class, versions randomized over students (but students sitting next to each other received the same version). *Total number of questions:* 27. *Number of forbid/allow questions:* 6.

- **Experiment 6**

Theme: road safety. *Subjects:* 100 subjects at a station. *Procedure:* a written questionnaire administered at Utrecht Central Station. *Total number of questions:* 36. *Number of forbid/allow questions:* 6.

- **Experiment 7**

Theme: young people and television. *Subjects:* 100 workers in a hospital. *Procedure:* the written questionnaire was filled in during lunch break in the staff canteen. Persons sharing a table received similar versions. *Total number of questions:* 34. *Number of forbid/allow questions:* 6.

- **Experiment 8**

Theme: commercials. *Subjects:* 100 students and teachers at a university. *Procedure:* a written questionnaire administered in university canteens. *Total number of questions:* 42. *Number of forbid/allow questions:* 6.

- **Experiment 9**

Theme: intake restrictions and restricted number quota [‘plaatsingsfixus’] for university studies. *Subjects:* 118 pupils in the 5th and 6th grade of

secondary school (VWO – pre-university education). *Procedure*: written questionnaires administered during class, versions randomized over students (but students sitting next to each other received the same version). *Total number of questions*: 30. *Number of forbid/allow questions*: 6.

• Experiment 10

Theme: right-wing political parties. *Subjects*: 118 pupils from 5 secondary school classes. *Procedure*: written administration during class. *Total number of questions*: 30. *Number of forbid/allow questions*: 6.

Appendix B: Example of Question Wordings and Experimental Design

Example of question wordings and experimental design: the experiment on road safety (experiment 6). Below, version 1 of the questionnaire (administered to a random half of the sample of respondents) is given. Version 2 is similar to version 1, but ‘forbid’ is replaced by ‘allow’ and vice versa.

– Introduction

Section 1:

- general question on road safety
- f/a question 1: Do you think the government should forbid the use of mobile phones while driving? Yes/no
- filler questions on road safety
- f/a question 2: should allow billboards alongside motor ways? Yes/no
- Filler questions on road safety
- f/a question 3: ... should allow unprotected level crossings? Yes/no
- Filler questions on road safety
- F/a question 4: ... should forbid trucks to pass cars on motor ways? Yes/no
- Filler questions on road safety
- F/a question 5: ... should allow riding on a motorized bicycle without safety helmet? Yes/no
- Filler questions on road safety
- F/a question 6: ... should forbid skeelering in pedestrian areas?
- Filler questions on road safety? Yes/no

Section 2:

- Introduction
- F/a question 1: The government should forbid the use of mobile phones while driving. Totally agree-totally disagree

- Filler question
- F/a question 2: The government should allow billboards alongside motor ways. Totally agree-totally disagree

Etc.

- Thank you for your cooperation, etc.

Appendix C: The Question Wordings Used

Experiment 1

Theme: youth culture

Subjects and procedure: 100 secondary school pupils (Havo/VWO) in four different classes, written questionnaire administered during class.

Questions (version 1):

1. Vind je dat de regering de verkoop van softdrugs ook aan mensen jonger dan 18 jaar moet toelaten? *Do you think the government should allow the sale of soft drugs also to people under 18?*
2. Vind je dat de overheid de verkoop van ECO-drugs (bijvoorbeeld paddo's) moet verbieden? *Do you think the government should forbid the sale of ecological drugs (e.g. magic mushrooms)?*
3. Vind je dat het dragen van petjes in de klas moet worden toegelaten? *Do you think wearing baseball caps should be allowed in class?*
4. Vind je dat het op school dragen van bomberjacks met Nederlandse vlaggetjes moet worden verboden? *Do you think wearing bomberjackets with Dutch flags on them should be forbidden?*
5. Vind je dat het dragen van hoofddoekjes in de klas moet worden toegelaten? *Do you think wearing head scarves in class should be allowed?*
6. Vind je dat de regering het uitzenden van geweldsfilms op de televisie voor 20.00 moet verbieden? *Do you think the government should forbid the broadcasting of violent films before 8 p.m.?*
7. Vind je dat het verhuren van geweldsvideo's aan jongeren onder de 18 moet worden toegelaten? *Do you think renting violent videos to people under 18 should be allowed?*

Experiment 2

Theme: medical choices

Subjects and procedure: 100 university students in university canteens, written questionnaire

Questions (version 1):

8. Vind je dat de overheid het klonen van dieren moet verbieden? *Do you think the government should forbid the cloning of animals?*

9. Vind je dat de overheid adoptie van kinderen door mannelijke homoparen moet toelaten? *Do you think the government should allow gay male couples to adopt children?*
10. Vind je dat de overheid kunstmatige inseminatie bij alleenstaande vrouwen moet verbieden? *Do you think the government should forbid artificial insemination for single women?*
11. Vind je dat de overheid dierproeven voor de ontwikkeling van cosmetische artikelen moet verbieden? *Do you think the government should forbid the testing of cosmetic products on animals?*
12. Vind je dat de overheid moet toelaten dat gynaecologen eicellen in de baarmoeder plaatsen om de kans op kinderen te vergroten? *Do you think the government should allow gynaecologists to implant female germ cells in order to increase the likelihood of pregnancy?*
13. Vind je dat de overheid plastische chirurgie, als dat vanwege schoonheidsredenen wordt gedaan, moet toelaten? *Do you think the government should allow plastic surgery if this is done for cosmetic reasons?*

Experiment 3

Theme: hooligans

Subjects and procedure: 100 secondary school pupils (4-HAVO/VWO) in four different classes, written questionnaire during classes

Questions (version 1):

14. Vind je dat het verboden moet zijn voor een burgemeester om een risicowedstrijd af te gelasten? *Do you think council mayors should be forbidden to cancel high-risk soccer games?*
15. Vind je dat de overheid een nadrukkelijke aanwezigheid van de ME tijdens risicowedstrijden moet verbieden? *Do you think the government should forbid the visible presence of special police units during high-risk soccer games?*
16. Vind je dat het toegelaten moet zijn voor een scheidsrechter om de wedstrijd te staken wanneer er oerwoudgeluiden komen van de tribune? *Do you think the referee should be allowed to stop a soccer game when the audience is insulting players by making 'jungle noises' from the stands?*
17. Vind je dat de overheid bij risicowedstrijden supporters van de uit spelende club moet verbieden de wedstrijd bij de wonen? *Do you think the government should forbid supporters of the visiting club to watch the game from the stadium if it is a high-risk game?*
18. Stel dat de supporters van een club zich misdragen. Vind je dat het dan toegelaten moet zijn voor de Voetbalbond om die club te straffen voor het gedrag van haar supporters? *Suppose supporters of a*

club misbehave. Do you think the Soccer Union should be allowed to punish the club for the behaviour of its supporters?

19. Vind je dat het toegelaten moet worden dat de politie supporters die bekend staan als voetbalvandalen uit voorzorg in de cel zet op de dag dat hun club een wedstrijd speelt? *Do you think the police should be allowed to detain supporters who are known hooligans on the days their club is playing as a precautionary measure?*

Experiment 4

Theme: supermarket policies

Subjects and procedure: 100 supermarket customers, oral questionnaire

Questions (version 1):

20. Vindt u dat de overheid moet verbieden dat supermarkten 's avonds na 18.00 open zijn? *Do you think the government should forbid supermarkets to be open after 6 p.m.?*
21. Vindt u dat de overheid moet verbieden dat supermarkten op zondag open zijn? *Do you think the government should forbid supermarkets to be open on Sundays?*
22. Vindt u dat de overheid de verkoop van vloeibare zuivelproducten, zoals melk, karnemelk en yoghurt, in niet-herbruikbare kartonnen verpakkingen, moet verbieden? *Do you think the government should forbid the sale of liquid dairy products (such as milk and yoghurt) in non-recyclable packaging?*
23. Vindt u dat de overheid het verstrekken van gratis plastic tasje in supermarkten moet verbieden? *Do you think the government should forbid the handing out of free plastic bags in supermarkets?*
24. Vindt u dat de overheid de verkoop van wijn in flessen zonder stategie moet verbieden? *Do you think the government should forbid the sale of wine in non-returnable bottles?*
25. Vindt u dat de overheid de verkoop van rookwaren in supermarkten moet verbieden? *Do you think the government should forbid the sale of cigarettes in supermarkets?*

Experiment 5

Theme: euthanasia and organ donation

Subjects and procedure: 118 secondary school pupils (VWO), written questionnaire during classes

Questions (version 1):

26. Een patiënt lijdt aan een ongeneeslijke ziekte en heeft veel pijn. De patiënt heeft zelf om voortijdige levensbeëindiging gevraagd. Vind je

dat euthanasie in deze situatie toegelaten moet worden? *A patient is suffering from an incurable disease and is in a lot of pain. The patient himself has asked for his life to be terminated. Do you think euthanasia should be allowed in that situation?*

27. Een patiënt ligt langdurig in coma. Hij heeft een aantal jaren daarvoor aangegeven dat, wanneer hij ooit in een langdurige coma zou raken, zijn leven beëindigd mag worden. Vind je dat euthanasie in deze situatie toegelaten moet worden? *A patient is in a long-lasting coma. A number of years earlier he has stated that if he should ever be in a state of long-lasting coma, his life may be terminated. Do you think euthanasia should be allowed in that situation?*
28. Een zwaar depressieve vrouw ziet al jaren geen uitweg meer. Zij verzoekt haar arts om haar leven te beëindigen. Vind je dat euthanasie in deze situatie toegelaten moet worden? *A badly depressive woman has lived without hope for many years. She asks her family doctor to end her life. Do you think euthanasia should be allowed in that situation?*
29. Een patiënt ligt in coma. Zijn familie geeft toestemming om zijn leven te beëindigen. De patiënt heeft hierover zelf nooit een uitspraak gedaan. Vind je dat euthanasie in deze situatie toegelaten moet worden? *A patient is in a coma. His family give their consent to terminate his life. The patient himself has never expressed his opinion on this point. Do you think euthanasia should be allowed in that situation?*
30. Een patiënt sterft. Het ziekenhuis wil zijn organen voor transplantatie gebruiken. De patiënt heeft hierover zelf nooit een uitspraak gedaan. Vind je dat donortransplantatie in deze situatie toegelaten moet worden? *A patient dies. The hospital wishes to use his organs for transplantation. The patient himself has never expressed his opinion on this point. Do you think donor transplantation should be allowed in that situation?*
31. Een patiënt sterft. Zijn longen zijn nog in goede staat. Een ander patiënt heeft om te blijven leven binnen zeer korte tijd een long nodig. Noch door de eerste patiënt, noch door zijn familie is toestemming gegeven voor donortransplantatie. Vind je dat donortransplantatie in deze situatie toegelaten moet worden? *A patient dies. His lungs are still in good condition. Another patient urgently needs a lung in order to stay alive. Neither the first patient nor his family have given permission for donor transplantation. Do you think that donor transplantation should be allowed in that situation?*

Experiment 6

Theme: safety in traffic

Subjects and procedure: 100 subjects in the central hall of a railway station, written questionnaire

Questions (version 1):

32. Vindt u dat de overheid telefoneren tijdens het autorijden moet verbieden? *Do you think the government should forbid the use of mobile phones while driving?*
33. Vindt u dat de overheid reclame langs de snelweg moet toelaten? *Do you think the government should allow billboards alongside motor ways?*
34. Vindt u dat de overheid onbeveiligde spoorwegovergangen (spoorwegovergangen zonder spoorbomen) moet toelaten? *Do you think the government should allow unprotected level crossings (without barriers)?*
35. Vindt u dat de overheid het inhalen door vrachtwagens op snelwegen moet verbieden? *Do you think the government should forbid trucks to pass cars on motor ways?*
36. Vindt u dat de overheid het rijden zonder helm op een motorfiets moet toelaten? *Do you think the government should allow riding a motorised bicycle without safety helmet?*
37. Vindt u dat de overheid skeeleren in voetgangersgebieden moet verbieden? *Do you think the government should forbid skeelering in pedestrian areas?*

Experiment 7

Theme: young people and television

Subjects and procedure: 100 workers in a hospital, written questionnaire administered in canteens during lunch break

Questions (version 1):

38. Vindt u dat de overheid moet verbieden dat jongeren roken in Nederlandse televisieprogramma's en films? *Do you think the government should forbid youngsters to smoke in Dutch television programs and films?*
39. Vindt u dat de overheid moet toelaten dat jongeren alcohol drinken in Nederlandse televisieprogramma's en films? *Do you think the government should allow youngsters to drink alcohol in Dutch television programs and films?*
40. Vindt u dat de overheid moet verbieden dat er gevloekt wordt in Nederlandse televisieseries? *Do you think the government should forbid swearing in Dutch television series?*

41. Vindt u dat de overheid moet toelaten dat speelfilms op televisie uitgezonden worden zonder vermelding van een minimumleeftijd? *Do you think the government should allow the broadcast of feature films on television without stating the minimum age?*
42. Vindt u dat de overheid moet toelaten dat gewelddadige speelfilms voor 22.00 uur op televisie uitgezonden worden? *Do you think the government should allow the broadcasting of violent movies on television before 10 p.m.?*
43. Vindt u dat de overheid moet verbieden dat schokkende beelden getoond worden in de vroege journaals? *Do you think the government should forbid showing shocking items during early editions of the news on television?*

Experiment 8

Theme: commercials

Subjects and procedure: 100 students and teachers of university, written questionnaire in canteens

Questions (version 1):

44. Vindt u dat de overheid ethisch-geladen reclames, zoals de Benetton-reclames, moet verbieden? *Do you think the government should forbid ethically loaded commercials and advertisements, such as the Benetton advertisements?*
45. Vindt u dat de overheid reclame voor alcohol moet toelaten? *Do you think the government should allow commercials for alcoholic drinks?*
46. Vindt u dat de overheid sluikreclame in televisieprogramma's moet verbieden? *Do you think the government should forbid plugged (hidden) advertising in television programs?*
47. Vindt u dat de overheid reclame voor 06-sexlijnen op de televisie moet toelaten? *Do you think the government should allow advertisements for telephone sex on television?*
48. Vindt u dat de overheid het verspreiden van reclamefolders op straat moet verbieden? *Do you think the government should forbid the distribution of commercial flyers to people in the street?*
49. Vindt u dat de overheid reclame door middel van vliegtuigen moet toelaten? *Do you think the government should allow advertising by means of airplanes?*

Experiment 9

Theme: fixed quota [numerus fixus] and restricted number quota [plaatsingsfixus] for university studies

Subjects and procedure: 118 pupils in the fifth and sixth grade of secondary school (VWO), written questionnaire during classes

Questions (version 1):

50. Vind je dat de overheid scholieren die geen toegespitst vakkenpakket hebben moet verbieden mee te loten voor een studie met numerus fixus? *Do you think the government should forbid pupils without the required combination of school-leaving certificate subjects to draw lots for fixed-quota disciplines?*
51. Vind je dat de overheid universiteiten moet toelaten om eigen lotingscriteria te formuleren voor numerus fixus studies (bijvoorbeeld door mensen met een bepaald cijfergemiddelde direct in te loten)? *Do you think the government should allow universities to formulate their own lottery criteria for fixed-quota disciplines (e.g. by allowing people with a set grade average in outright)?*
52. Vind je dat de overheid scholieren moet verbieden mee te loten voor meerder numerus fixus studies? *Do you think the government should forbid pupils to draw lots for more than one fixed-quota discipline?*
53. Vind je dat de overheid universiteiten moet toelaten om toelatingsexamens voor numerus fixus studies te hanteren? *Do you think the government should allow universities to set entrance examinations for fixed-quota disciplines?*
54. Vind je dat de overheid scholieren moet verbieden om mee te loten voor meerdere studies met een plaatsingsfixus? *Do you think the government should forbid pupils to draw lots for more than one restricted-number course?*
55. Vind je dat de overheid universiteiten moet verbieden eigen lotingscriteria te formuleren ten aanzien van de plaatsingsfixus? *Do you think the government should forbid universities to formulate their own lottery criteria for restricted-number courses?*

Experiment 10

Theme: right-wing political parties

Subjects and procedure: 118 pupils from five secondary school classes, written questionnaire administered during classes

Questions (version 1):

56. Vind je dat de overheid uitzendingen op de televisie voor extreemrechtse partijen moet verbieden? *Do you think the government should forbid broadcasts for extreme right-wing political parties?*

57. Vind je dat de overheid demonstraties van extreem rechts moet verbieden? *Do you think the government should forbid demonstrations by the extreme right?*
58. Vind je dat de overheid bijeenkomsten van extreem-rechts moet verbieden? *Do you think the government should forbid meetings of the extreme right?*
59. Vind je dat de overheid openbare toespraken van extreem-rechtse partijen moet verbieden? *Do you think the government should forbid public speeches of extreme right-wing parties?*
60. Vind je dat de overheid berichten in de media over extreem-rechtse partijen moet verbieden? *Do you think the government should forbid media coverage of extreme right-wing parties?*
61. Vind je dat de overheid extreem-rechtse partijen moet verbieden? *Do you think the government should forbid extreme right-wing political parties?*

Notes

1. In fact, the design of the experiments was much more complicated. A four-groups design was used and measurements were repeated across time – in order to avoid different sorts of distortions, such as memory effects. A structural modelling approach was used to assess the correlations between forbid/allow questions within each time of measurement and between different times of measurement. Analyses focused on Jöreskog's (1971) concept of 'congenericity': two tests (sets of questions) measure the same construct if they correlate unity (1.0), irrespective of measurement errors. For details, readers are referred to Holleman (1999b).
2. A meta-analysis of previous forbid/allow experiments shows a mean asymmetry size of 14%: 'not forbid' was answered 14% more than 'yes allow'. The variance in the asymmetry size turned out to be large (97), the standard deviation was 9.85 (Holleman, 1999a, 2000).
3. Seventeen out of 61 two-point-scale questions showed a significant wording effect ($p < .05$), as well as 17 out of 61 seven-point-scale questions. This finding of 17 wording effects out of 61 observations is significant ($p < .001$).
4. The dependent variable is dichotomous, which means the variance is known once the mean score is known ($s^2 = (p^*(1 - p))$). For example, if the mean score to a question is 1, the variance is zero. A logit model was used [$\log(p/(1 - p))$] in analyzing these data but only the proportions (the percentages of yes/no answers obtained) will be reported. In this type of regression models the difference between means as well as between variances can be tested by a chisquare distributed test statistic (Houtkoop & van den Bergh, 2000).
5. Due to the fact that the analyses conducted here assume normality of the distributions, the distributions were normalized so as not to contradict model assumptions. In order to facilitate interpretation, the results were again transformed to a scale with the same mean and variance of the observed scores.
6. One could counter-argue that these high reliabilities mainly show that there was an order effect, some pressure for consistency, or a memory effect. This might be the case. On the other hand, there is no good reason why this pressure or memory effect should be differ-

ent for forbid questions than for allow questions. Hence, it remains a rewarding enterprise to compare the relative meanings of ‘yes’ and ‘no’ to forbid/allow questions.

7. This transformation is acquired by subtracting the mean answer to each forbid question from each forbid question’s score, and dividing the result by the standard deviation obtained for that forbid question. This is done for each forbid and for each allow question. The result of this computation is that each answer to a seven-point-scale forbid question and each answer to a seven-point-scale allow question can be expressed in terms of its standard deviation from a fixed scale midpoint.
8. The variances between studies prove to be insignificant, so without loss of fit these can be fixed at zero. This means that the variance of the mean answer on forbid-7p that was given by respondents who answered ‘yes’ to forbid-2p does not differ significantly between studies. The same holds for the mean answer expressed to forbid-7p that was given by respondents who answered ‘no’ to forbid-2p, and similarly for the allow questions. Hence, only the variances between persons and between questions need to be discussed.
9. The standardized variances differ from the observed variances due to the fact that standardization equals the variances of the forbid seven-point scale to the variances of the seven-point-scale allow questions (both becoming 1.0), although the observed variances were different from each other. As the focus in this section is on an interpretation of the *differences* in meanings of the answering options to forbid/allow questions between versions, the differences in standardized variances can be better interpreted than the differences in observed variances in this exploration.
10. This analysis is based on Talmy’s (1988) semantic framework of Force Dynamics.
11. Note that this is a modification of the hypothesis formulated in the introduction, which assumed that the connotations of ‘forbid’ might be more extreme than the connotations of ‘allow’. The findings discussed here suggest that ‘yes forbid’ is more extreme than all of the other answering options. If the connotations of forbidding in general are more extreme, one would expect the extremity of ‘not forbid’ to be more pronounced as well.

References

- Billiet, J. (1989). Wat te doen? Beschouwingen over het nut van pasklare voorschriften voor het ontwerpen van survey-vragen [What to do? Reflections on the use of ready-made advice on the design of survey questions]. In: J. Van der Zouwen and W. Dijkstra (eds.), *Sociaal wetenschappelijk onderzoek met vragenlijsten. Methoden, knelpunten, oplossingen*. Amsterdam: VU-Uitgeverij, pp. 35–52.
- Bishop, G. F., Hippler, H.-J., Schwarz, N. & Strack, F. (1988). A comparison of response effects in self-administered and telephone surveys. In: R. M. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls J. Waksberg (eds.), *Telephone Survey Methodology*. New York: Wiley, pp. 273–282.
- Cicourel, A. V. (1982). Interviews, surveys and the problem of ecological validity. *The American Sociologist* 17: 11–20.
- Hippler, H.-J. & Schwarz, N. (1986). Not forbidding isn’t allowing: The cognitive basis of the forbid/allow asymmetry. *Public Opinion Quarterly* 50: 87–96.
- Holleman, B. C. (1999a). Wording effects in survey research. Using meta-analysis to explain the forbid/allow asymmetry. *Journal of Quantitative Linguistics* 6: 29–40.
- Holleman, B. C. (1999b). The nature of the forbid/allow asymmetry. Two correlational studies. *Sociological Methods & Research* 28/2:209–244.
- Holleman, B. C. (2000). *The Forbid/Allow Asymmetry. On the Cognitive Mechanisms Underlying Wording Effects in Surveys*. Amsterdam/Atlanta: Rodopi.

- Houtkoop, H. & van den Bergh, H. (2000). Effects of introductions in large-scale telephone survey interviews. *Sociological Methods & Research* 28/3: 281–300.
- Jobe, J. B. & Mingay, D. J. (1991). Cognition and survey measurement: History and overview. *Applied Cognitive Psychology* 5/1: 175–192.
- Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika* 36: 109–133.
- Krosnick, J. A. & Schuman, H. (1988). Attitude intensity, importance, and certainty and susceptibility to response effects. *Journal of Personality and Social Psychology* 54/6: 940–952.
- Molenaar, N. (1982). Response effects of “formal” characteristics of questions. In: W. Dijkstra and J. Van der Zouwen (eds.), *Response Behaviour in the Survey Interview*. London: Academic Press, pp. 49–89.
- Reuband, K.-H. (2000). “Erlauben” vs. “nicht erlauben” oder “verbieten”? wie sich unterschiedliche Frage-Alternativen auf das Antwortverhalten auswirken. *ZUMA-Information* 48: 42–55.
- Rugg, D. (1941). Experiments in wording questions II. *Public Opinion Quarterly* 5: 91–92.
- Schuman, H. & Presser, S. (1981/1996). *Questions and Answers in Attitude Surveys. Experiments on Question Form, Wording, and Context*. Thousand Oaks, CA: Sage Publications (Originally published in New York: Academic Press).
- Sudman, S., Bradburn, N. M. & Schwarz, N. (1996). *Thinking About Answers. The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass Publishers.
- Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science* 12: 49–100.
- Tourangeau, R., Rips, L. J. & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge University Press.
- Tourangeau, R. & Rasinski, K. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin* 103: 299–314.
- Waterplas, L., Billiet, J. & Loosveldt, G. (1988). De verbieden versus niet toelaten asymmetrie. Een stabiel formuleringseffect in survey-onderzoek? [The forbid/allow asymmetry. A stable wording effect in survey research?] *Mens en Maatschappij* 63: 399–415.