

**Monitoring Attitudes Over Time –  
Real Change or the Result of Repeated Interviewing?**

Fabienne Kraemer<sup>1</sup>, Peter Lugtig<sup>2</sup>, Bella Struminskaya<sup>2</sup>, Henning Silber<sup>1</sup>, Bernd Weiß<sup>1</sup>, and  
Michael Bosnjak<sup>3</sup>

<sup>1</sup> Department of Survey Design and Methodology, GESIS – Leibniz Institute for the Social  
Sciences, Mannheim, Germany

<sup>2</sup>Department of Methodology and Statistics, Utrecht University, Utrecht, Netherlands

<sup>3</sup>Department of Psychological Research Methods, Trier University, Trier, Germany

**Author Note**

This work was supported by the “German Research Foundation” (DFG) [Grant Number: 418316085].

This manuscript is currently under consideration for publication at Sociological Methods and Research.

Recommended citation of this preprint:

Kraemer, F., Lugtig, P., Struminskaya, B., Silber, H., Weiß, B., and Bosnjak, M. (2024)

Monitoring Attitudes Over Time: Real Change or the Result of Repeated Interviewing?

SocArXiv, Available at:

<https://osf.io/preprints/socarxiv/jcyd9>

**Abstract**

Researchers often use panel data to study change and stability of social patterns. However, repeated interviewing can affect respondents' attitudes in a panel study by raising awareness and triggering reflection processes on surveyed topics (*cognitive stimulus hypothesis*). We investigated change in respondents' attitudes about abortion and the underlying mechanisms of attitude change across six waves using data from a survey experiment administered within a probability-based and a non-probability panel in Germany. We manipulated the frequency of receiving identical attitude questions on the same issue. We estimated multiple-group and longitudinal structural equation models to differentiate change in the measurement of reported attitudes from "real" attitude change. We show that repeatedly asking about abortion increases the reliability of respondents' reported abortion attitudes, providing support for the cognitive stimulus hypothesis. Our results also suggest that improved response behavior due to general survey experience further enhances attitude reliability when answering identical attitude questions repeatedly.

*Keywords:* panel conditioning, cognitive stimulus hypothesis, attitude change, structural equation modeling, panel studies

### Introduction

Panel surveys are widely used due to their possibility to track attitudes and subjective beliefs of individuals over time and thus to study change and stability of social patterns. However, previous research indicates that shifts in respondents' attitudes over the course of a longitudinal study do not solely reflect individual or social change but instead, might result from repeated interviewing (Bergmann & Barth, 2018; Sturgis et al., 2009; Waterton & Lievesley, 1989). This potential bias inherent to panel studies is generally referred to as *panel conditioning* and describes the impact of repeated interviews on respondents' behaviors and attitudes or the way in which they are reported in subsequent waves of the study (Kalton et al., 1989; Lazarsfeld, 1940).

Recent studies have explored the existence and magnitude of panel conditioning (e.g., Cornesse et al., 2023; Eckman & Bach, 2021; Kraemer et al., 2023; Silber et al., 2019). Research that specifically investigate the effects of panel conditioning on respondents' attitudes over time suggest that repeated surveying on the same issue raises respondents' awareness and stimulates further reflection processes on surveyed topics – a process which has been referred to as the *cognitive stimulus model* (Sturgis et al., 2009). Studies have provided some evidence on the assumptions of the cognitive stimulus model by documenting more stable and reliable survey responses and an increased opinionation, that is, reporting attitudes instead of saying 'don't know,' among experienced panelists (Bergmann & Barth, 2018; Binswanger et al., 2013; Kroh et al., 2016; Sturgis et al., 2009).

Previous studies typically used non-experimental designs (Bergmann & Barth, 2018; Binswanger et al., 2013; Sturgis et al., 2009) which are susceptible to other types of bias affecting change rates, such as maturation effects or selective dropout from panel studies. Such effects are very difficult to control for in observational data and might induce systematic bias of unknown magnitude into change estimates. But, more importantly, most studies have

not investigated the reasons why attitudinal responses in panel studies become more stable and reliable over time. Are these panel conditioning effects the result of increased reflection by the respondent on the survey topics or do panel respondents simply become more consistent in their responses over time due to an increased experience with answering surveys in general, that is, gain experience with typical response tasks?

In this paper, we aim to address these research gaps by examining the following research questions: (1) Does repeatedly asking the same attitude question lead to panel conditioning in form of more consistent answers and an increased opinionation over time? (2) Can conditioning effects best be explained as a result of increased reflection by the respondent on the topic of the question, or simply by the fact they become a ‘better respondent’ due to increases in general survey experience?

Our study is unique as it uses data of a longitudinal survey experiment comprising six survey waves carried out both within a probability-based and a non-probability panel. The experimental design manipulates the frequency of receiving identical attitude questions over the course of the study and allows us to investigate the impact of different levels of exposure to identical attitude questions on the existence and magnitude of attitude changes.

Additionally, this design enables us to study attitude change not only between different respondent groups but also across the six survey waves to further examine the underlying mechanisms of changes in respondents’ attitudes. Additionally, we included measures of attitude certainty and knowledgeability to gain further insights on the mechanisms and relevance of reflection processes for changes in respondents’ (reported) attitudes.

### **Background**

Research addressing the effects of panel conditioning on attitudes and subjective beliefs is scarce. While most panel conditioning studies have focused on identifying and estimating bias in overall response quality or behavioral outcomes such as voter turnout (e.g.,

## ATTITUDES OVER TIME

Battaglia et al., 1996; Clausen, 1968; Halpern-Manners et al., 2017; Sun et al., 2019; Yalch, 1976), only few studies have investigated how repeated interviewing changes respondents' attitudes over the course of a panel study (e.g., Binswanger et al., 2013; Jagodzinski et al., 1987; Sturgis et al., 2009).

Despite the variety in research designs, the existing studies mainly provide consistent findings regarding attitude change in panel studies and document more stable, reliable answers to attitudinal questions and an increased opinionation on surveyed topics across waves.

In one of the earlier studies that investigated attitudinal responses over the course of a panel study, Jagodzinski et al. (1987) differentiated between changes in the latent attitudes and changes in the reported attitudes and showed that the reliability of attitude reports regarding immigrant workers in Germany increased between the first and second survey wave. Waterton and Lievesley (1989) compared the political attitudes of members of the British Social Attitudes Panel over the course of three survey waves to fresh cross-sections and found an increase in politization due to repeated interviewing with panelists reporting more often to support a political party. Similarly, Sturgis and colleagues (2009) found that panelists of the British Household Panel Study provided fewer “don’t know” responses to a left-right political orientation question over the course of 11 panel waves – a finding which they attributed to an increase in (political) opinionation. Additionally, Sturgis et al.’s study found that respondents’ answers to attitudinal questions became more reliable and stable over time.

Subsequent studies have provided further evidence on an increase in stability and reliability of attitudinal survey responses over time. Kruse (2009) showed an increase in inter-wave stabilities of environmental attitudes. Kroh et al. (2016) tested the consistency of individual response patterns in 17 multi-item scales which were administered in 30 subsequent panel waves within the German Socio-Economic Panel (GSOEP) and found that

## ATTITUDES OVER TIME

repeated interviewing increased the reliability of individual survey responses over time. Research by Sun et al. (2019) who investigated the effect of panel conditioning on overall response quality using several response quality indicators documented an increase in reliability of selected attitude scales over time for members of two probability-based panel studies in the USA providing further evidence on an increase of reliability of attitudinal responses over the course of a panel study. Previous findings that hint at an increased opinionation of experienced panelists on surveyed topics were supported by Bergmann and Barth (2018), who used data from the German Longitudinal Election Study (GLES) and showed that repeated interviewing decreased respondents' political indecision by affecting attitude strength. This effect was especially prominent for respondents who initially held weak attitudes regarding party vote intentions.

Most of the existing studies on panel conditioning and its effect on attitudinal responses explain their findings based on the notion that repeated interviewing increases awareness and prompts respondents to reflect intensively on the surveyed topics which leads them to form opinions where none existed prior to the survey or to change existing attitudes. The assumption of triggered reflection processes due to the repeated interviewing on specific topics was extended to a general theoretical framework – the *cognitive stimulus model* (Sturgis et al., 2009).

The cognitive stimulus model assumes that respondents – instead of forming opinions “on the spot,” base their opinions on information and intensive cognitive engagement with an issue at hand: Being repeatedly exposed to and asked about the same issues might stimulate respondents to intensively reflect on the surveyed topics and their stance towards them. Respondents might even discuss the issues with others or obtain further information, for example, by paying closer attention to the news (Sturgis et al., 2009). Sturgis et al. (2005) argue that even if respondents do not change their existing views, the triggered reflection processes lead them to hold more consistent attitudes and to show an increased attitude

strength (*attitude crystallization*) than without any intensified deliberation due to repeated interviewing. The three main empirical implications of the cognitive stimulus model are: (a) attitudes become more *reliable* over time, (b) attitudes become more *stable* over time, (c) and *opinionation* on surveyed issues increases.

Following the assumptions of the cognitive stimulus model (Sturgis et al., 2009), respondents' learning effects over the course of a panel study are topic- or question-specific. Changes in respondents' answers are due to a repeated exposure to similar or identical question content and the thereby prompted reflection processes. Waterton and Lievesley (1989) have provided initial evidence for the mechanism of content-specific reflection processes by conducting follow-up in-depth interviews with panel members. They showed that the survey evoked discussions among the panelists and their friends and family (Waterton & Lievesley, 1989). On the contrary, Clinton (2001) found that long-term members of an online panel reported lower levels of news consumption compared to respondents who had been in the panel for a shorter period of time – an assumed consequence of topic-specific reflection besides discussions with others. In any case, the present evidence regarding the mechanisms underlying observed attitude crystallization is at best suggestive.

In contrast to the cognitive stimulus model, some studies have argued that observed attitude change and increases in the reliability of attitudinal responses over time might simply be the result of an increased experience with answering surveys in general and, consequently, “better” response behavior leading to more accurate self-reports (Kroh et al., 2016; Warren & Halpern-Manners, 2012; Waterton & Lievesley, 1989). It is argued that repeatedly answering surveys may increase respondents' familiarity with the overall survey procedure and improve their understanding of the rules and requirements of an interview (Struminskaya, 2016; Waterton & Lievesley, 1989). Consequently, respondents might learn about the relevance of reporting their true opinions and behaviors to the researcher and learn to use different survey instruments as well as to map their answers to different response scales (Nancarrow &

Cartwright, 2007; Waterton & Lievesley, 1989). In fact, research suggests respondents becoming familiar with the response task and their experience with different survey instruments might decrease initial complexities of question comprehension tasks (Waterton & Lievesley, 1989), facilitate less error-prone retrieval of relevant information (Bailar, 1989), and might decrease error in mapping the retrieved information to the response scale as respondents' familiarity in dealing with different response scales increases (Basso et al., 2001).

Respondents might even prepare for subsequent interviews by looking up relevant information beforehand: studies using external validation data showed increases in the accuracy of respondents' income reports and reports regarding the receipt of unemployment benefits with higher panel experience (Frick et al., 2006; Yan & Eckman, 2012). We refer to this mechanism as the *"better respondents" model* (Yan & Eckman, 2012).

Findings from the German GESIS Panel provide further support for the "better respondents" model by documenting shorter response times for more experienced panelists which cannot be attributed to satisficing response behavior (Kartsounidou et al., 2023; Kraemer et al., 2023). Similarly, it has been observed that experienced members of the LISS Panel in the Netherlands more often did their task "on time" compared to novice respondents (Scherpenzeel & Zandvliet, 2011), indicating "better" response behavior with an increased survey experience.

Both models, the cognitive stimulus and the "better respondents" model can lead to the same outcome but are based on different mechanisms. While Sturgis et al. (2009) attribute fewer "don't know" responses of more experienced panelists to an increased opinionation on the subject matter, Waterton and Lievesley (1989) explain their finding of a decrease in "don't know" responses with respondents learning the "rules that govern the interview process," arguing that respondents learn about the importance of their survey responses to researchers and therefore, provide more often substantial answers.



Altogether, the existing empirical findings are only suggestive of the mechanisms underlying observed changes in attitude reports and there is no direct evidence on the role of content-specific reflection processes versus familiarity with the response task. To contribute to the literature, we aim to investigate the mechanisms of observed attitude changes in repeated measurement settings by testing the following four hypotheses. Two of the expected findings (H1 and H2) could be due to either the cognitive stimulus or the “better respondents” model, while Hypothesis 3 is based on the assumptions of the cognitive stimulus model, and Hypothesis 4 is based on the “better respondents” model.

According to the cognitive stimulus model (Sturgis et al., 2009), repeatedly being asked about a specific issue increases awareness and prompts respondents to intensively reflect on the subject matter. In the case of non-existent opinions toward the issue, this might lead to the formation of new opinions and, therefore, an increase in opinionation reflected in fewer “don’t know” answers. The “better respondents” model (Yan & Eckman, 2012) however, would assume that fewer “don’t know” answers simply indicate an increased familiarity with the response task and the survey procedure increasing respondents’ willingness to report their attitudes. Based on both models, we hypothesize:

*Hypothesis 1: Respondents who repeatedly answer identical attitude questions will give fewer “don’t know” answers compared to respondents who do not repeatedly receive identical attitude questions.*

Similarly, attitude crystallization due to repeated interviewing is the core assumption of the cognitive stimulus model suggesting that the triggered increase in cognitive engagement with the surveyed topics leads to more consistent and crystallized attitudes which are less susceptible to change. Following the “better respondents” model on the other hand, more consistent and reliable attitudes would not reflect internal attitude change but

## ATTITUDES OVER TIME

respondents simply becoming better at answering (multi-item) scales, leading to more accurate attitude reports in which “noise” and random errors have been canceled out due to experience with the response tasks. Using either model, we expect:

*Hypothesis 2: Respondents who repeatedly answer identical attitude questions show a crystallization of attitudes (i.e., increases in attitude reliability and/or stability) compared to respondents who do not repeatedly receive identical attitude questions.*

The next hypothesis is only in line with the assumptions of the cognitive stimulus model and not the “better respondent” model as it addresses the underlying mechanism of the cognitive stimulus model:

*Hypothesis 3: Respondents who are repeatedly exposed to identical attitude questions show an increased attitude strength and knowledgeability on the subject matter compared to respondents who do not repeatedly receive identical attitude questions.*

In addition, we derived one expectation that can be explained by the “better respondents” model, but not the cognitive stimulus model. Following Kroh et al. (2016) who discuss the possibility that the increased familiarity with a response task could have an impact on increases in reliability of individual survey responses over time, we hypothesize:

*Hypothesis 4: Respondents, who have high levels of general survey experience, show more reliable attitudes than respondents with low levels of general survey experience.*

## Methods

### Experimental Design

To test our hypotheses, we use data collected in a longitudinal survey experiment which was carried out both within a probability-based mixed-mode panel study and a non-probability online access panel in Germany. The survey experiment comprised six panel waves and manipulated the frequency of answering identical questions (i.e., target questions) over six consecutive waves. Respondents were randomly assigned to one of three experimental groups: one experimental group received the identical target questions in each of the six waves (i.e., fully conditioned), another experimental group received the target question only from Wave 4 onwards (i.e., medium conditioned), and the control group received the target questions only once in the final wave (i.e., unconditioned)<sup>1</sup> (for an illustration of the experimental design, see Figure 1). In the non-probability panel study, respondents were randomly assigned to three groups of equal sample size ( $n_{\text{fully conditioned}} = 647$ ;  $n_{\text{medium conditioned}} = 651$ ;  $n_{\text{unconditioned}} = 648$ ). Respondents of the probability-based panel study have been randomly assigned to three experimental groups of unequal sample size with 50% of the respondents being assigned to the fully conditioned group, whereas 25% of respondents were assigned to the medium conditioned and the unconditioned groups, respectively ( $n_{\text{fully conditioned}} = 2,337$ ;  $n_{\text{medium conditioned}} = 1,165$ ;  $n_{\text{unconditioned}} = 1,158$ )<sup>2</sup>.

---

<sup>1</sup> We refer to the three experimental groups as “fully conditioned”, “medium conditioned”, and “unconditioned” group dependent on their frequency of exposure to identical attitude questions. The used terminology, however, does not indicate whether respondent conditioning has actually happened due to our experimental setup.

<sup>2</sup> The unequal sample size of the three experimental groups resulted from subsequent changes in the survey design during the fieldwork period. The unconditioned control group was equally split to introduce an additional manipulation of conditioning frequency in which respondents received the attitudinal (target) questions three out of six times (i.e., medium conditioned).

## ATTITUDES OVER TIME

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6
Fully conditioned group	Target questions	Target questions	Target questions	Target questions	Target questions	Target questions
Medium conditioned group	Questions on other topics	Questions on other topics	Questions on other topics	Target questions	Target questions	Target questions
Unconditioned group	Questions on other topics	Questions on other topics	Questions on other topics	Questions on other topics	Questions on other topics	Target questions

**Figure 1.** Experimental design: Manipulation of exposure to identical target questions

### Data

*Non-probability panel study.* The experimental study conducted within the non-probability online access panel was fielded between October 2020 and December 2021. The recruitment of respondents targeted persons aged 18 years and older who were at the time of the study residing in Germany and were registered as a member of a commercial online access panel, which was used as the sampling frame. For the implementation of the study, a quota sample was drawn based on the demographic distribution of the German Microcensus. Sample members were selected using cross-quotas on gender (female; male), age (18-35 years; 36-65 years; 66 years and older), and level of education measured as the highest general school leaving certificate (below secondary school certificate; secondary school certificate; higher education entrance qualification). Data was collected exclusively online via web-based surveys with each survey wave taking about 15 minutes to complete, except for the first wave that took about 35 minutes due to the inclusion of additional background measures on socio-demographics, political attitudes, and personality.

Selected panel members were explicitly invited to participate in the study via a personalized email by the panel provider. However, every panel member could also participate via the panel's online user interface. Only respondents who completed the first

## ATTITUDES OVER TIME

wave were able to participate in the following survey wave. The invited panel members received up to four email reminders. Each reminder was sent after three to four days following the invitation or the preceding reminder email. Respondents received a postpaid incentive after each completed survey wave. The amount of the paid incentive varied by wave from €0.60 to €2.50, depending on the average survey duration. Furthermore, respondents received an additional bonus payment of €2 if they successfully participated in all six panel waves.

Across the six waves of the panel study, the completion rates (COMR) (Callegaro & DiSogra, 2008) varied from 93.9% in the first wave to 92.9% in the last wave of the study. Up to the last wave of the study, 42.0% of the respondents dropped out of the study (see Online Appendix B, Tables B.1 and B.2 for detailed completion and attrition rates for each wave separated by the different respondent groups; Callegaro & DiSogra, 2008).

*Probability-based panel study.* The longitudinal survey experiment was additionally fielded within the GESIS Panel – a German probability-based mixed-mode access panel comprising about 5,200 panelists that are surveyed on topics such as subjective well-being, personality and personal values, media usage, and work and leisure (GESIS, 2023). The GESIS Panel was initially recruited in 2013, based on a random sample drawn from municipalities' population registers targeting German-speaking persons aged 18-70 years who were permanently residing in private households in Germany. The recruitment process followed a multi-stage procedure which initially included a face-to-face interview, whose response rate (AAPOR RR1) was 35.5% (Bosnjak et al., 2018). In 2016 and 2018, two refreshment samples ( $n = 1,710$ ;  $n = 1,607$ ) were recruited to counter the effects of panel attrition.

Data collection in the GESIS Panel is administered in two modes, via web-based surveys (online mode) or mailed paper questionnaires (offline mode). By 2021, a majority of respondents (about 75%) complete the surveys online, whereas the remaining 25% of

panelists participate in the offline mode. Irrespective of the participation mode, each panelist receives a survey invitation by mail that contains a prepaid cash incentive of €5. Panelists participating in the online mode additionally receive an email survey invitation and up to two email reminders: one week and two weeks after the invitation. Offline respondents do not receive reminders.

In this paper, we use the data of six consecutive survey waves of the GESIS Panel (waves 41 to 46), which were fielded between October 2020 and January 2022. Since 2021, survey waves of the GESIS Panel are administered every three months but until 2020, data collection was conducted bimonthly. Accordingly, while the experimental study was fielded, the interval between survey waves changed from two to three months. The questionnaire of the experimental study took up around 5 minutes to complete.

In the first wave of the experimental study, completion rates (COMR) varied from 93.2% for the first panel cohort to 93.7% and 89.7% for the second and third cohort, respectively<sup>3</sup>. In the last wave, they ranged from 92.2% for the first and second panel cohort to 90.3% for the third panel cohort. Cumulative response rates (CUMR) in the first wave of the experimental study ranged from 11.5% for the first panel cohort to 9.7% for the second and 9.6% for the third panel cohort. In the last wave of the experimental study, cumulative response rates ranged from 10.7% for the first panel cohort to 8.8% and 8.8% for the second and third panel cohort, respectively. In the final wave of our experimental study, attrition rates ranged from 49.6% for the first cohort to 41.0% for the second and 31.6% for the third cohort (Schulz et al. 2021; Stadtmüller et al., 2022).

---

<sup>3</sup> The outcome rates provided by the GESIS Panel are specified for each recruited panel cohort, separately.

### Measures

*Attitudes toward abortion.* To investigate change in respondents' (reported) attitudes, we used 4-items measuring attitudes toward abortion which have been used within the US General Social Survey (GSS) (Adamczyk & Valdimarsdóttir, 2018; Carter et al., 2009). Respondents were presented with the four statements on different circumstances under which a woman should be able to obtain a legal abortion, and then asked to indicate to which extent they agree or disagree with each statement on a scale ranging from 1 "strongly disagree" to 7 "strongly agree." As a further response alternative, each item included a "don't know" category. The stated reasons for an abortion were (a) a woman simply not wanting more children, (b) a family which cannot afford any more children, (c) a high probability of serious birth defects, and (d) the endangerment of the woman's health due to the pregnancy (for question wording and implementation, see Online Appendix A). The items were presented in an item-by-item format with each of the four items shown on a separate screen with a vertical response scale.<sup>4</sup>

*Attitude certainty.* To gain further insights on the relevance of reflection processes for attitude change, we used follow-up measures which were administered both in the first and last wave of the study directly after respondents answered the 4-items assessing abortion attitudes. The first measure assessed the *certainty* of respondents' abortion attitudes. After responding to the four items on abortion, respondents were asked to indicate whether they have a clear opinion on the issue of abortion on a 7-point Likert-scale ranging from 1 "I do not have a clear opinion on that." to 7 "I have a clear opinion on that.". Response options additionally included a "don't know" – category.

*Knowledgeability.* The second measure captured respondents' level of *knowledgeability* on the issue. Respondents were asked to rate their familiarity with the topic

---

<sup>4</sup> In the offline mode of the GESIS Panel, the abortion items were not displayed on different pages of the paper questionnaire.

of legal abortion on a 7-point Likert-scale ranging from 1 “Not familiar at all” to 7 “Very familiar” with the option to choose “don’t know” as an explicit response category.

*Conditioning frequency.* To investigate whether topic-specific reflection processes are responsible for changes in respondents’ (reported) attitudes, we use the study’s experimental design and compare respondents with three different levels of exposure to identical attitude questions over the course of the six survey waves. We compare respondents who received identical attitude questions in every wave (i.e., fully conditioned) with respondents who received identical questions only three times from the second half of the study onwards (i.e., medium conditioned) and those who received the target attitude questions only once in the final wave of the study (i.e., unconditioned).

*Survey experience.* To explore the possible impact of experience with the general response task on answers to attitude questions, we use two different operationalization of survey experience. *First*, we compare the respondents across the probability and non-probability panels. We argue that respondents of non-probability panels, on average, have much more experience in participating in many different surveys compared to respondents of probability-based panel studies, who only participate, for instance, every two or three months. Administrative data obtained by the panel provider, indeed, shows that respondents of the non-probability panel had a median of 74 completed surveys in the last 12 months prior to our study, whereas respondents of the probability-based panel had participated in a maximum of 6 survey waves in the same period. Some nonprobability online panelists might even be a member of one or multiple other online panels (Callegaro et al., 2014). As a *second* operationalization of survey experience, we focus on respondents’ experience within a panel and differentiate our experimental groups within the probability-based panel by panel cohort. We compare respondents of the initial panel cohort (recruited in 2013 and highly experienced) with respondent of the second (medium experienced) and third cohort (lowest experienced),



which have been incorporated into the panel 3 (in 2016) and 5 years (in 2018) later, respectively.

### **Analytic strategy**

The overall goal of our analyses is to test the mechanisms responsible for observed changes in attitude answers in panel studies. For this, we first investigate differences in the proportion of “don’t know” answers to the target abortion measure between the experimental groups by conducting  $\chi^2$ - and Fisher’s exact tests (if the number of observations was not sufficiently high) to answer *Hypothesis 1*.

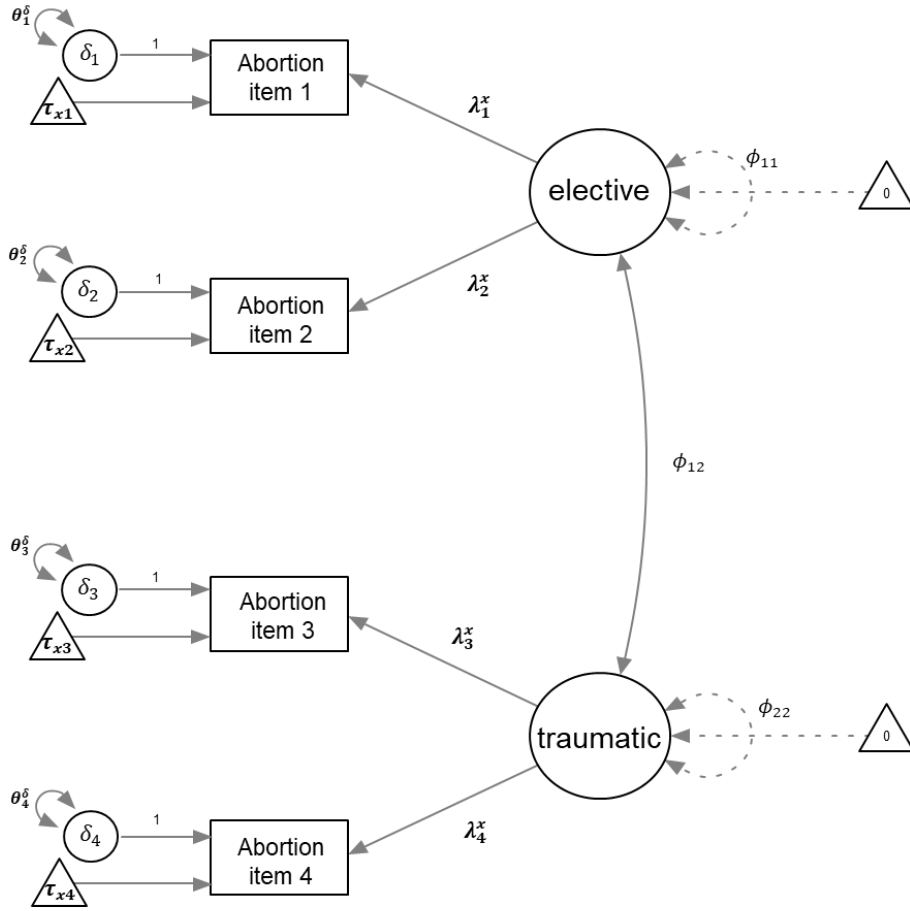
Subsequently, we conduct a between-group comparison across the six experimental groups regarding their observed abortion attitudes in the last wave of the study to answer *Hypothesis 2*. To compare the groups, we use multiple-group confirmatory factor analysis (MG-CFA) testing measurement invariance of the target abortion measures across the differently conditioned respondent groups (i.e., different exposure to identical abortion questions and general experience in survey participation) in Wave 6. For this and subsequent analyses based on structural equation modeling, we use the R-package lavaan (Rosseel, 2012). By conducting a MG-CFA, we can disentangle the effect of repeated interviewing on the measurement component of responses to the abortion questions (i.e., how respondents report their attitudes on abortion) from the effect of repeated interviewing on the actual underlying latent abortion attitudes. Testing measurement invariance consists of a series of pre-defined model comparisons that successively impose more stringent equality constraints on the parameters of a measurement model (Byrne, 1989).

The baseline measurement model that we compare across the groups in Wave 6 is a two-factor model distinguishing between abortion sought for elective reasons and abortion

sought for trauma- or health-related reasons<sup>5</sup>. In particular, the first two abortion items (i.e., the woman does not want more children and the family cannot afford any more children) load on a latent factor representing elective reasons, whereas the last two items (i.e., high probability of birth defects and serious endangerment of the woman's health) load on a second latent factor capturing the traumatic reasons for seeking out an abortion (for an illustration of the two-factor model, see Figure 2). The two-factor model has been validated in several previous studies (Hoffmann & Johnson, 2005; Osborne et al., 2022) and in various country contexts (Huang et al., 2016; Karpov & Kääriäinen, 2005; Osborne & Davies, 2012) against a model with only one latent factor.

---

<sup>5</sup> Using this terminology (i.e., 'elective' and 'traumatic'), we follow previous literature that investigates a two-factor measurement model of abortion attitudes in various country contexts (Osborne et al., 2022; Osborne & Davies, 2012). We distinguish between elective and traumatic abortion to capture the differences in the circumstances that lead to abortion. However, we are not suggesting that having an elective abortion is free from trauma or that traumatic abortion does not offer any choice to a woman.

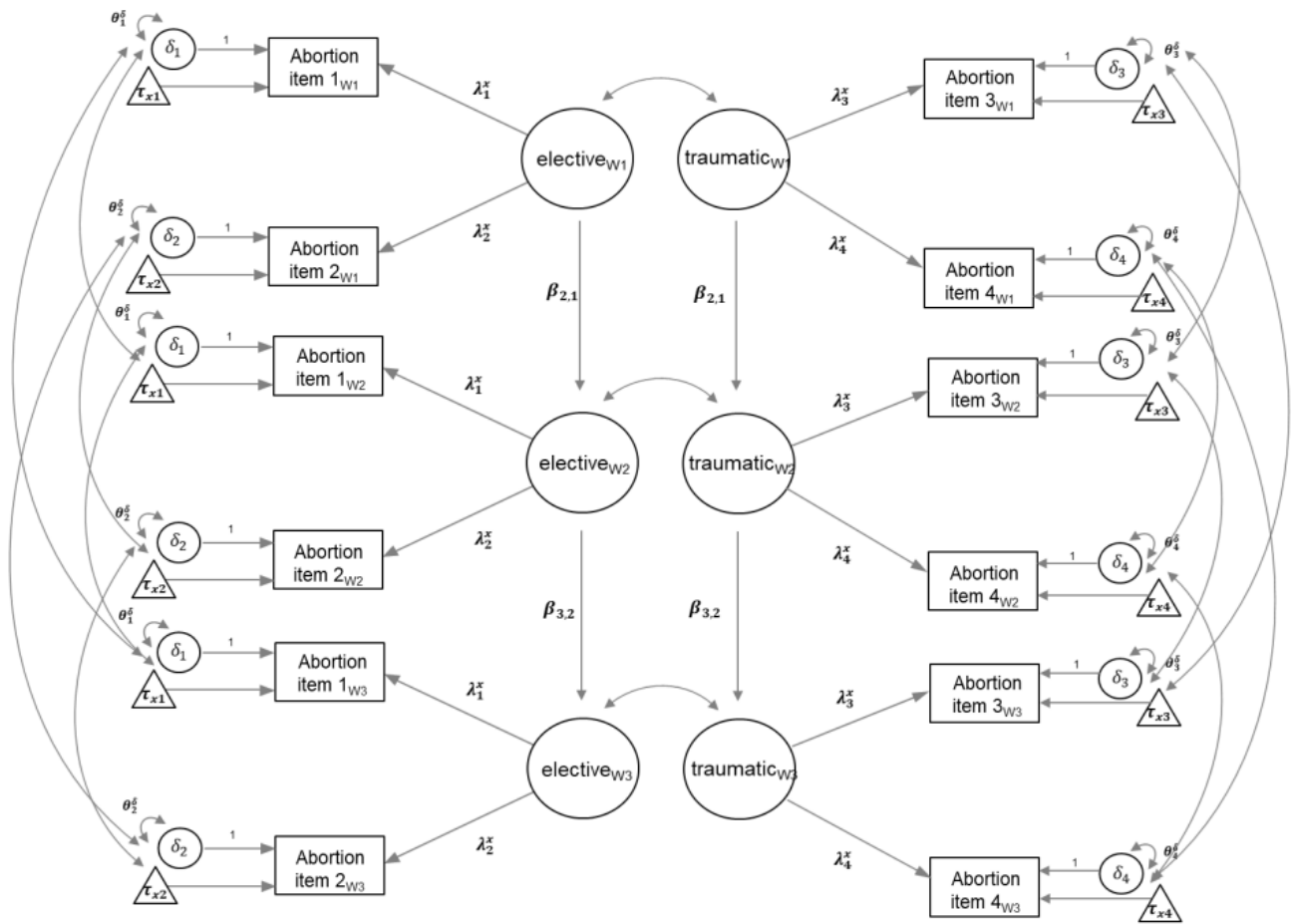


**Figure 2.** Two-factor measurement model on attitudes toward abortion.

Following the pre-defined model comparisons of a MG-CFA, we first test the overall factor structure and pattern of factor loadings of our measurement model (i.e., testing of configural invariance) across the six respondent groups (i.e., the three experimental groups in each panel study). Configural invariance exists if the two-factor structure as well as the loading patterns of the four abortion items are the same in every of the six experimental groups. If configural invariance can be established, we will proceed to test whether the magnitude of factor loadings is equal across the groups (i.e., metric invariance), and whether the intercepts of the items are equal across groups (i.e., scalar invariance). Finally, we will test whether residual variances of the indicators are equal across the six respondent groups (i.e., strict invariance).

## ATTITUDES OVER TIME

As a sensitivity analysis, we additionally investigate answers to the target abortion measure across the six survey waves to test *Hypothesis 2* further. We conduct this within-group comparison for fully conditioned respondents who received the abortion measure in each survey wave and estimate a longitudinal factor analysis that tests the measurement invariance of abortion attitudes over time differentiating between changes in reported attitudes and underlying latent attitudes. Again, we follow the pre-defined series of model comparisons testing differences in the measurement components (i.e., factor loadings, item intercepts, residual variances). Testing our baseline model over the six survey waves, we allow the residual variance of an item to correlate across the six waves. If scalar invariance as a minimum requirement can be established, we will then further employ the longitudinal design of our study and use the within-group comparison to investigate the stability of the latent abortion attitudes over time by estimating an autoregressive first-order model (*simplex model*; see Figure 3).



**Figure 3.** Autoregressive first-order model (simplex model); for reasons of conciseness, only the first three out of six waves are shown.

While the outlined analyses so far address the question of whether panel conditioning takes place, such a finding would be consistent with both theoretical models (i.e., cognitive stimulus and “better respondents” model).

In order to test *Hypothesis 3*, which specifically relates to the assumptions of the cognitive stimulus model, we incorporate (self-reported) attitude certainty and knowledgeability on the issue of abortion into the between-group comparison of the baseline measurement model in the last wave of the study. By doing so, we estimate a structural equation model including these external covariates, which affect both latent factors of the

abortion items. We test whether the means of respondents' attitude certainty and knowledgeability are significantly different across groups and how their impact on the latent abortion factors differs across groups.

Finally, to test *Hypothesis 4*, we compare the differently conditioned respondents of the probability and non-probability panels. Additionally, we differentiate the experimental groups of the probability-based panel by cohort and conduct multiple-group confirmatory factor analyses across the three different cohorts of the panel for each conditioning level (i.e., fully conditioned, medium conditioned, and unconditioned) separately.

## Results

### **Hypothesis 1: Differences in the proportion of “don’t know” answers**

Before investigating change in respondents' abortion attitudes and its underlying mechanisms, we first compared the dropout rates in the last wave of the study across the experimental groups to exclude the possibility of bias due to differential non-response possibly affecting differences in observed abortion attitudes across groups. In both the probabilistic and non-probabilistic panel study, we only find marginal differences in dropout rates (with a maximum difference of about 6 percentage points), indicating that nonresponse is not significantly different across the experimental groups (for the detailed dropout rates per experimental group, see Table B.6 and B.7 in the Online Appendix). To test *Hypothesis 1*, Table 1 displays the proportion of “don’t know” responses across the six experimental groups. The results show that the three respondent groups of the probabilistic panel significantly differ in their amount of given “don’t know” answers regarding two out of the four abortion items. The respondent groups of the non-probabilistic panel only differ significantly in their proportion of “don’t know” answers for one of the four abortion items. Nevertheless, fully conditioned respondents (i.e., respondents who were surveyed on abortion in each wave) in

both the probabilistic and non-probabilistic panel consistently show a smaller proportion in “don’t know” responses compared to the unconditioned respondents across all items, which is in line with the hypothesis. The magnitude of differences is greater among the groups of the non-probabilistic panel with differences amounting up to 4 percentage points (p.p.) on average, whereas differences across the respondent groups of the probabilistic panel are only 2 p.p. on average. For the last two abortion items (abortion due to birth defect and due to pregnancy endangering the woman’s health), the amount of “don’t know” responses did not differ across the respondent groups of the probabilistic panel, indicating that repeatedly being surveyed on these issues does not influence whether those respondents form a substantive opinion or report that they do not know.

Table 1. Percentage of “don’t know” answers for the abortion items per experimental group

<b>Experimental group</b>	<b>Abortion item 1</b>	<b>Abortion item 2</b>	<b>Abortion item 3</b>	<b>Abortion item 4</b>
<i>Probability panel</i>				
<b>Fully conditioned</b>	6.41	6.29	5.08	4.36
<b>Medium conditioned</b>	9.13	8.78	6.30	3.72
<b>Unconditioned</b>	9.20	9.03	5.77	4.51
$\chi^2$	11.10	10.23	2.10	.97
<b>p-value</b>	.004	.006	.349	.617
<i>Nonprobability panel</i>				
<b>Fully conditioned</b>	3.63	2.82	0.81	1.21
<b>Medium conditioned</b>	7.34	6.62	5.45	3.10
<b>Unconditioned</b>	7.76	6.93	4.33	4.76
$\chi^2$	4.35	4.99	-	-
<b>p-value</b>	.114	.083	.007	.066

*Note.* The estimated p-values for significance of the differences across groups are based on Pearson’s  $\chi^2$ -test and Fisher’s exact test for count data if observations of “don’t know” were  $n = <5$  in single cells of the contingency table.

When excluding “don’t know” and “skips” as a non-substantial answer for the further analyses below, we have complete data on all four abortion items for 4,370 respondents and partly missing data for 430 respondents, who are included in the between-group comparison analyses. For the within-group comparisons (on fully conditioned respondents of the

probabilistic panel), we can include complete data on all abortion items and all six measurement points for 1,289 respondents and partly missing data for 1,031 respondents. To account for the missing patterns in the data, we use Full Information Maximum Likelihood (FIML) to estimate the model coefficients (Enders & Bandalos, 2001).

### **Testing measurement invariance across the six experimental groups**

We first test *configural invariance* of the measurement model. For this, we fit the two-factor structure of the measurement model on abortion attitudes to the six different respondent groups while freely estimating the model's parameters (i.e., factor loadings, item intercepts, residual variances). The results provided in Table 2 show that the two-factor model fits the data in each group ( $\chi^2(6) = 5.98, p = .43$ , root mean square error of approximation (RMSEA) = .000, Comparative Fit Index (CFI) >.999, Akaike Information Criterion (AIC) = 60,539), according to the rules of thumb for model fit (Kline, 2023).

Next, we test for *metric invariance* by constraining all factor loadings to be equal across the six respondent groups but allowing for variation in item intercepts and residual variances. The results show that the metric invariance model does not significantly fit worse than the configural invariance model ( $\Delta\chi^2(10) = 11.80, p = .30$ ) and additionally has a smaller value for the AIC (AIC = 60531), indicating that the factor loadings do not differ between the six respondent groups.

Consequently, we test for *scalar invariance* by constraining all factor loadings and additionally, the item intercepts to be equal across the experimental groups but freely estimating residual variances of the items across groups. Again, results show that the more restrictive model fits the data better than the less restrictive metric invariance model ( $\Delta\chi^2(10) = 11.85, p = .30$ ; AIC = 60523), and accordingly that both factor loadings and item intercepts are equal across the six experimental group.



Lastly, we test for *strict invariance* and constrain all parameters (i.e., factor loadings, item intercepts, and residual variances) of the measurement model to be equal across the six experimental groups. The results indicate that equal residual variances across groups are not supported in the data by showing that the strict invariance model fits the data significantly worse than the less restrictive scalar invariance model ( $\Delta\chi^2(20) = 89.73, p < .001$ ; AIC = 60572). Thus, while there is no difference in the factor loadings and the item intercepts of the abortion measurement model, we find that residual variances of the abortion items differ significantly across our six experimental groups.

Table 2. Model fit statistics for measurement invariance across the experimental groups

Model/Test	$\chi^2$	df	CFI	RMSEA	AIC	$\Delta\chi^2$	$\Delta df$
<b>Configural invariance</b>	5.98	6	>.999	.000	60539	-	-
<b>Metric invariance</b>	17.77	16	>.999	.011	60531	11.80	10
<b>Scalar invariance</b>	29.62	26	>.999	.013	60523	11.85	10
<b>Strict invariance</b>	119.35***	46	.992	.045	60572	89.73***	20

Note. AIC = Akaike's Information Criterion; RMSEA = root mean square error of approximation; CFI = Comparative-Fit Index; \*\*\*  $p < 0.001$  \*\*  $p < 0.01$  \*  $p < 0.05$ .

## Hypothesis 2: Attitude crystallization

### *Between-group analyses*

*Reliability.* To test *Hypothesis 2* on the crystallization of attitudes due to repeated interviewing, we first investigate changes in attitude reliability. For this, we inspect the residual variances of the four abortion items – which have been shown to be unequal across the six respondent groups – in more detail. Similar to Jagodzinski et al. (1987), we infer a higher degree of attitude reliability from smaller residual item variances.

Table 3 shows the standardized estimates for the residual variances of each of the four abortion items (Columns  $\theta_1^\delta$  -  $\theta_4^\delta$ ; for unstandardized estimates, see Table C.1 in Appendix).

Although residual variances do not consistently decrease across the extent of exposure to the identical abortion questions (i.e., conditioning frequency) for all items, the fully conditioned respondent groups show the smallest residual variances for three out of the four abortion items. We calculated an average residual variance (see Table 3, Column  $\overline{\theta^\delta}$ ) over the four abortion items for reasons of conciseness, which shows similar results compared to the patterns we observe for the four individual items. Specifically, fully conditioned respondents show the smallest residual variances, which is in line with *Hypothesis 2*.

Table 3. Standardized parameter estimates of the scalar invariance model across the experimental groups

Experimental group/ parameter estimates	$\lambda_1^x$	$\lambda_2^x$	$\lambda_3^x$	$\lambda_4^x$	$\tau_{x1}$	$\tau_{x2}$	$\tau_{x3}$	$\tau_{x4}$	$\theta_1^\delta$	$\theta_2^\delta$	$\theta_3^\delta$	$\theta_4^\delta$	$\overline{\theta^\delta}$
<i>Probability panel</i>													
<b>Fully conditioned</b>	0.88	0.92	0.94	0.76	2.29	2.30	4.70	6.39	0.22	0.16	0.13	0.43	<b>0.24</b>
<b>Medium conditioned</b>	0.88	0.90	0.89	0.77	2.43	2.41	4.58	6.72	0.23	0.18	0.22	0.41	<b>0.26</b>
<b>Unconditioned</b>	0.86	0.88	0.93	0.77	2.36	2.35	4.31	5.99	0.27	0.22	0.13	0.41	<b>0.26</b>
<i>Non-probability panel</i>													
<b>Fully conditioned</b>	0.93	0.89	0.97	0.79	2.42	2.25	4.59	6.31	0.14	0.20	0.07	0.38	<b>0.20</b>

---

<b>Medium conditioned</b>	0.88	0.84	0.93	0.73	2.47	2.29	4.35	5.80	0.23	0.29	0.14	0.46	<b>0.28</b>
-------------------------------	------	------	------	------	------	------	------	------	------	------	------	------	-------------

---

<b>Unconditioned</b>	0.83	0.85	0.87	0.72	2.25	2.22	4.59	6.40	0.31	0.28	0.24	0.48	<b>0.33</b>
----------------------	------	------	------	------	------	------	------	------	------	------	------	------	-------------

---

We additionally estimate and compare reliability coefficients (i.e., McDonald's third coefficient omega (McDonald, 2013)) across the six groups, which provide further support for *Hypothesis 2*, showing the highest reliability coefficients (for three out of four items .90 or higher) among fully conditioned respondents (see Online Appendix, Table C2)<sup>6</sup>.

### ***Within-group analyses***

As a sensitivity analysis, we additionally investigate within-group changes in respondents' attitudes toward abortion over time. We focus on the analysis of fully conditioned respondents (who received the identical abortion measure in each of the six waves) from both the probability-based and the non-probability panel and first examine measurement invariance of the abortion items across the six measurement points. Our analysis of measurement invariance over time for respondents of the probability-based panel shows that only partial scalar invariance can be established (i.e., equal factor loadings and equal item intercepts except from the fourth abortion item in Wave 1), and that similar to the between-group analyses residual variances are not equal across the six waves. For fully conditioned respondents of the non-probability panel, measurement invariance analyses show similarly that only scalar invariance is supported in the data and that residual variances of the four

---

<sup>6</sup> Besides investigating changes in the measurement component of respondents' abortion attitudes, we also examined whether repeated interviewing on the identical issues leads to changes in respondent's latent attitudes and found no differences in the latent factor means between the experimental groups. This indicates, that repeatedly answering identical questions on abortion does not change respondents' attitudes toward the subject (for detailed results, see Table D.1 and D.2 in the Online Appendix).

## ATTITUDES OVER TIME

abortion items are significantly different across the six measurement points (for detailed results, see Table E.1 for probability-based panel respondents and Table E.4 for non-probability panel respondents in the Online Appendix).

*Reliability.* The initial screening of the items' residual variances over time (Table 4, Columns  $\theta_1^\delta$  -  $\theta_4^\delta$ , for unstandardized estimates, see Table E.2 in the Online Appendix) for respondents of the probability-based panel, shows a pattern of decreasing error variances across the six survey waves. We calculated an average residual variance over the four individual items (see Table 4, Column  $\overline{\theta^\delta}$ ). The average residual variance across the six measurement points supports the general pattern we observe across the four individual items, showing a somewhat consistent decrease across waves (until Wave 6) and smaller error variances in the subsequent waves compared to Wave 1.

Table 4. Standardized parameter estimates for the longitudinal partial scalar invariance model

Survey wave/ parameter estimates	$\lambda_1^x$	$\lambda_2^x$	$\lambda_3^x$	$\lambda_4^x$	$\tau_{x1}$	$\tau_{x2}$	$\tau_{x3}$	$\tau_{x4}$	$\theta_1^\delta$	$\theta_2^\delta$	$\theta_3^\delta$	$\theta_4^\delta$	$\overline{\theta^\delta}$
<b>Wave 1</b>	0.88	0.88	0.88	0.77	2.29	2.28	3.91	5.48	0.24	0.22	0.23	0.41	<b>0.28</b>
<b>Wave 2</b>	0.90	0.89	0.91	0.79	2.34	2.27	4.14	5.73	0.20	0.22	0.17	0.38	<b>0.24</b>
<b>Wave 3</b>	0.88	0.88	0.93	0.79	2.30	2.28	4.33	5.82	0.23	0.22	0.13	0.38	<b>0.24</b>
<b>Wave 4</b>	0.90	0.90	0.93	0.79	2.37	2.34	4.34	5.84	0.20	0.20	0.13	0.39	<b>0.23</b>
<b>Wave 5</b>	0.90	0.91	0.94	0.78	2.32	2.30	4.38	5.81	0.18	0.18	0.12	0.39	<b>0.22</b>

<b>Wave 6</b>	0.90	0.91	0.92	0.78	2.28	2.27	4.50	6.03	0.20	0.18	0.15	0.40	<b>0.23</b>
---------------	------	------	------	------	------	------	------	------	------	------	------	------	-------------

---

Turning to the fully conditioned respondents of the non-probability panel, we similarly observe an overall pattern of decrease in residual variances across waves and smaller residual variances in subsequent waves compared to Wave 1 of the study (for detailed results, see Table E.5 and E.6 in the Online Appendix). The longitudinal patterns of the items' residual variances in both panels provide additional tentative support for *Hypothesis 2*, which expects an increase in attitude reliability due to repeated interviewing on the identical issue.

Additional analyses comparing McDonald's third coefficient omega across the six measurement points for both fully conditioned respondents of the probability-based and the non-probability panel provide further tentative evidence on an increase in reliability supporting *Hypothesis 2* (for detailed results, see Table E.3 and Table E.7 in the Online Appendix).

*Stability.* We further examine the crystallization of attitudes due to repeated interviewing (*Hypothesis 2*) by investigating changes in the stability of respondents' abortion attitudes over time. For this, we use within-group comparison across the six survey waves and focus on fully conditioned respondents of the probabilistic panel<sup>7</sup>. As the data does support partial scalar invariance of our measurement model over time, we can investigate change patterns in respondents' latent abortion attitudes over the course of the six waves. We estimate an autoregressive first-order model (*simplex model*) modelling respondents' latent abortion attitudes at measurement point  $t$  as being influenced solely by their abortion attitudes at the previous measurement point ( $t - 1$ ). The estimated autoregression coefficients across waves

---

<sup>7</sup> Parameter estimates of the longitudinal measurement model for fully conditioned respondents of the non-probability panel have shown negative residual covariances for the third abortion item across selected measurement points. Due to the possibility of model misspecification and overfitting, we did not test relationships in the underlying abortion attitudes for respondents of the non-probability panel (for detailed results, see Table E.4 in the Online Appendix).

thus indicate the stability of respondents' latent abortion attitudes over time. We compared a first-order model with unequal autoregression coefficients to a first-order model in which we restricted the autoregression coefficients to be equal over time (i.e., stability of respondents' latent abortion attitudes is equal over time). Model fit statistics show that the model with equal autoregression coefficients fits the data significantly worse than the model with unequal autoregression coefficients over time ( $\Delta\chi^2(8) = 61.89, p < .001$ ; AIC = 142699; for detailed model fit statistics, see Table E.8 in the Online Appendix), suggesting that the stability of respondents' abortion attitudes changes over time.

To investigate the pattern of the autoregression coefficients over time, we inspect the coefficients' estimates provided in Table 5 (for the unstandardized estimates, see Table E.9 in the Online Appendix). The results show that autoregression coefficients between Wave 1 and Wave 2 are lowest compared to subsequent waves, indicating that the stability of respondents' latent abortion attitudes does increase after the second interview. Consequently, observed changes over time are not restricted to changes in the measurement component of the reported abortion attitudes but are also reflected in respondents' latent abortion attitudes, which become more stable over time. It should be noted that the results on changes in autoregression coefficients over time should be interpreted cautiously as the simplex model fit the data worse than the previously established partial scalar invariance model, possibly due to model complexity or complex patterns in our data that cannot be fully captured by the simplex model alone. Nevertheless, the increase in stability over time provides further tentative evidence in line with our *Hypothesis 2*, which assumes that repeatedly answering identical attitude questions leads to the crystallization of attitudes.

Table 5. Standardized estimates of first-order model with unequal autoregression coefficients

	$\beta_{w2, w1}$	$\beta_{w3, w2}$	$\beta_{w4, w3}$	$\beta_{w5, w4}$	$\beta_{w6, w5}$
<b>Elective factor</b>	0.83	0.91	0.93	0.93	0.91
<b>Traumatic factor</b>	0.83	0.89	0.88	0.89	0.86

### **Hypothesis 3: Increase in attitude certainty and knowledgeability as a result of reflection processes**

To test *Hypothesis 3*, we investigate the role of respondents' attitude certainty and knowledgeability. We first fit a structural model with attitude certainty and knowledgeability as external covariates that influence both latent factors across the six experimental groups. The model fits the data sufficiently well ( $\chi^2(30) = 33.50$ ,  $p = .30$ , RMSEA = .012, CFI = .999, AIC = 95,522). We then compare this model to a structural equation model that includes equality constraints on the means of attitude certainty and knowledgeability across groups. We find that the model with equality constraints on the covariates' means does not fit the data significantly worse than the less restrictive model ( $\Delta\chi^2(10) = 15.08$ ,  $p = .13$ ; AIC = 95,517; for detailed model fit statistics, see Table F.1 in the Online Appendix). This indicates that respondents do not hold their abortion attitudes more strongly or show higher levels of knowledgeability when they are repeatedly asked on the issue of abortion compared to those respondents who are not repeatedly surveyed on abortion.

After testing the covariates' means across the respondent groups, we then estimate a model with equality constraints regarding the covariates' impact on the latent abortion attitudes (i.e., equal regression slopes between the covariates and the latent factors). We find

that the model with equal regression slopes fits the data better than the model with unequal regression slopes ( $\Delta\chi^2(20) = 30.89, p = .06$ ; AIC = 95508, see Table F.1. in the Online Appendix for detailed results), suggesting that the impact of attitude certainty and knowledgeability on latent abortion attitudes also does not differ across the experimental groups. Altogether, these results do not provide support for *Hypothesis 3*, as we do neither observe an increase in certainty or knowledgeability nor an increase in their impact on underlying abortion attitudes.

#### **Hypothesis 4: Survey experience and increased attitude reliability (“better respondent” model)**

##### ***Probability-based vs. non-probability panel respondents***

To test our fourth hypothesis regarding the impact of survey experience (i.e., experience with the response task in general) on the reliability of reported abortion attitudes, we first compare residual variances as an indicator of attitude reliability between fully conditioned respondents of the probability-based against respondents of the non-probability panel. To do so, we first constrain the residual variances of the abortion items to be equal between respondents of the probabilistic and the non-probabilistic panel. The results show that the model with equal residual variances among respondents of the probabilistic and non-probabilistic panel fits the data significantly worse than the scalar invariance model with unequal residual variances ( $\Delta\chi^2(12) = 45.27, p < .001$ ; AIC = 60544; for detailed results, see Table G.1 in the Online Appendix), indicating that respondents from the probability-based and the non-probability panel significantly differ from each other.

When comparing the standardized residual variances between the two panels (see parameter estimates of the scalar invariance model in Table 3), the residual variances of the fully conditioned respondents are larger among probability-based compared to non-probabilistic panel members, which is line with Hypothesis 4. With respect to the reliability



coefficients (see Table C.2 in Online Appendix), we observe a similar pattern with fully conditioned respondents of the non-probabilistic panel showing (slightly) higher reliability coefficients compared to fully conditioned members of the probability-based panel. Because the observed differences are of small magnitude, our analyses provide tentative evidence for Hypothesis 4, assuming that higher levels of survey experience lead to increased attitude reliability.

### *Comparison of panel cohorts within the probability-based panel*

We additionally differentiated the respondent groups of the probability-based panel by panel cohort to test Hypothesis 4 further. For every conditioning level, we tested differences in our measurement model between respondents of the three cohorts who have different tenure within the panel. Results of the MG-CFA across the panel cohorts show no significant differences in model parameters (for detailed results, see Online Appendix H). Consequently, factor loadings, item intercepts, and the residual variances of the abortion items are equal across cohort, indicating that experience within a panel does not change respondents' reported abortion attitudes. Overall, results imply that respondents' abortion attitudes are not affected by the amount of experience, so that *Hypothesis 4* is not supported for this comparison.

## Discussion

Previous research suggests that repeated interviewing on the same issues stimulates reflection processes within respondents which can lead to the crystallization of attitudes, that is, more reliable and stable attitudes (*cognitive stimulus hypothesis*; see Sturgis et al., 2009). In this paper, we tested the *cognitive stimulus hypothesis* against the *better respondents model* (Yan & Eckman, 2012) which explains changes in attitude answers over time with respondents' increased experience with the general response task and more accurate self-

reports. As the empirical base, we investigated changes in respondents' abortion attitudes and the underlying mechanisms of those changes over the course of six consecutive panel waves. Our study used data from a longitudinal survey experiment conducted within both a German probabilistic and non-probabilistic panel. In this experiment, we manipulated the frequency of receiving identical attitude questions (*1 time vs. 3 times vs. 6 times*). We estimated structural equation models to further differentiate between changes in the measurement of respondents' reported abortion attitudes and changes in their latent attitudes toward abortion.

Overall, our study provides evidence for panel conditioning effects that result from repeatedly asking respondents questions on the same issues. Respondents who received the identical abortion questions in each wave of the study provided significantly fewer "don't know" answers to most items of the abortion battery compared to respondents who received the attitudinal target questions less often (*Hypothesis 1*). Following previous literature, this decrease in "don't know" responses could either be attributed to improved response behavior (Waterton & Lievesley, 1989) or an increased opinionation on the topic (Sturgis et al., 2009). Moreover, the results show that attitude reliability is significantly different across the experimental groups and is highest for those respondents who received the identical abortion questions in each survey wave (i.e., fully conditioned respondents). Increases in attitude reliability, could be due to both better response behavior leading to more accurate self-reports as well as triggered reflection processes on surveyed topics. However, the observed differences in attitude reliability between the groups who differ in their survey intensity on identical topics, point to the attitude crystallization assumption of the *cognitive stimulus hypothesis*. Within-group comparisons of fully conditioned respondents provide further evidence in line with the findings of the between-group comparisons: Respondents' attitude reports on abortion become more reliable over time compared to the first wave of the study which can in this case be due to, both the increasing exposure to the identical issue of abortion or respondents simply becoming better at answering multi-item scales.

The results of the first-order autoregressive model, investigating change in the latent traits further show that respondents' latent abortion attitudes become more stable in subsequent waves of the study following the second interview. Although the results of the autoregressive model should be interpreted with caution given its model fit statistics, this implies that the attitude reliability we observe is at least partially due to a change in respondents' latent abortion attitudes rather than only the result from a change in how they report their attitudes. Therefore, the findings of the longitudinal factor analysis further support the assumption of crystallized abortion attitudes due to repeated interviewing on an identical issue (*Hypothesis 2*). Altogether, our findings on increased attitude reliability and stability are consistent with previous studies which document more reliable and stable attitudes among more experienced panelists (Bergmann & Barth, 2018; Kroh et al., 2016; Sturgis et al., 2009; Sun et al., 2019).

Whereas both the findings of the between- and within-group comparisons point toward the *cognitive stimulus hypothesis* as an explanation for observed changes, we did not find evidence on increased attitude certainty and knowledgeability on abortion after respondents answered the identical abortion question multiple times (*Hypothesis 3*). This finding contradicts the assumptions of the *cognitive stimulus hypothesis* (Sturgis et al., 2009) which argues that repeated survey participation triggers reflection processes and further information search, possibly leading to an increased attitude strength and increased knowledge levels. Noteworthy, however, we used self-reported measures of attitude strength and knowledgeability which could have impacted this result.

We further tested the assumptions of triggered reflection processes against the theory of better response behavior by comparing respondents from the probability-based panel to the respondents of the non-probabilistic panel as well as comparing different panel cohorts of the probability-based panel. Whereas we do not find any evidence for differences in reported abortion attitudes across the three cohorts of the probabilistic panel, we find that fully

conditioned respondents of the non-probabilistic panel show smaller residual variances and somewhat higher reliability coefficients compared to fully conditioned respondents of the probability-based panel. This finding indicates that general experience with responding to surveys might further enhance the reliability of reported attitudes (*Hypothesis 4*). In addition, as we did not only find changes in reliability of the measured attitudes but also an increased stability of the underlying abortion attitudes, we can assume that reflection processes on the survey topic might be triggered after repeatedly answering the same attitude question and that these processes are responsible for changes in the “true” underlying attitudes and not only for changes in their measurement component. This further supports the *cognitive stimulus hypothesis* (Sturgis et al., 2009).

Our study is not without limitations and opportunities for future research. First, for the operationalization of general survey experience – to test its effect on changes in respondents’ reported attitudes –, we compare different panel cohorts of the probabilistic panel. However, even the most recent cohort included in our analysis and in comparison, the least experienced cohort of the probabilistic panel has been part of the panel for about two years and respondents have already participated in up to 16 survey waves at the beginning of our study, which might have impacted our results.

Second, we used subjective measures on attitude certainty and knowledgeability on the issue of abortion to test the *cognitive stimulus hypothesis*, possibly affecting conclusions on the relevance of reflection processes for attitude change in panel studies. Future research could use objective measures (e.g., response latency) and test alternative measures of peer discussions and media attention on surveyed topics to deepen the understanding of the underlying mechanism of attitude change.

In addition, future research could experimentally vary the topic and task difficulty while still manipulating the frequency of receiving identical question content to shed further

light on which mechanism (triggered reflection processes vs. improved response behavior) is most likely responsible for observed changes in respondents' (attitude) answers over time.

Overall, our study provides valuable insights into the validity of attitudinal responses in longitudinal studies. By using an experimental study design and distinguishing between changes in the measurement of respondents' attitudes and their latent traits, we showed that repeated interviewing increases the reliability of reported attitudes and increases how stable latent attitudes are being held by respondents, both supporting the cognitive stimulus hypothesis. In addition, we found that experience with the overall response task further enhances attitude reliability when answering identical attitude questions repeatedly. Consequently, our study suggests that repeated measurement can improve the accuracy of attitude answers in panel studies, which allows researchers a valid analysis of change and stability over time.

## References

- Adamczyk, A., & Valdimarsdóttir, M. (2018). Understanding Americans' abortion attitudes: The role of the local religious context. *Social Science Research*, 71, 129–144.
- Bailar, B. A. (1989). Information needs, surveys, and measurement errors. *Panel surveys*, 1–24.
- Basso, M. R., Lowery, N., Ghormley, C., & Bornstein, R. A. (2001). Practice effects on the Wisconsin Card Sorting Test–64 card version across 12 months. *The Clinical Neuropsychologist*, 15(4), 471–478.
- Battaglia, M. P., Zell, E. R., & Ching, L. Y. H. (1996). Can Participating in a Panel Sample introduce Bias into Trend Estimates? *Proceedings of Survey Research Methods Section of the American Statistical Association*, 1010–1013.
- Bergmann, M., & Barth, A. (2018). What was I thinking? A theoretical framework for analysing panel conditioning in attitudes and (response) behaviour. *International Journal of Social Research Methodology*, 21(3), 333–345.  
<https://doi.org/10.1080/13645579.2017.1399622>
- Binswanger, J., Schunk, D., & Toepoel, V. (2013). Panel Conditioning in Difficult Attitudinal Questions. *Public Opinion Quarterly*, 77(3), 783–797.  
<https://doi.org/10.1093/poq/nft030>
- Byrne, B. M. (1989). Multigroup Comparisons and the Assumption of Equivalent Construct Validity Across Groups: Methodological and Substantive Issues. *Multivariate Behavioral Research*, 24(4), 503–523. [https://doi.org/10.1207/s15327906mbr2404\\_7](https://doi.org/10.1207/s15327906mbr2404_7)
- Callegaro, M., Baker, R., Bethlehem, J., Göritz, A. S., Krosnick, J. A., & Lavrakas, P. J. (2014). Online panel research. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Hrsg.), *Online Panel Research* (1. Aufl., S. 1–22). Wiley. <https://doi.org/10.1002/9781118763520.ch1>

- Callegaro, M., & DiSogra, C. (2008). Computing response metrics for online panels. *Public opinion quarterly*, 72(5), 1008–1032.
- Carter, J. S., Carter, S., & Dodge, J. (2009). Trends in abortion attitudes by race and gender: A reassessment over a four-decade period. *Journal of Sociological Research*, 1(1), 1–17.
- Clausen, A. R. (1968). Response Validity: Vote Report. *The Public Opinion Quarterly*, 32(4), 588–606.
- Clinton, J. D. (2001). *Panel Bias from Attrition and Conditioning: A Case Study of the Knowledge Networks Panel*. 1–34.
- Cornesse, C., Blom, A. G., Sohnius, M.-L., Gonzalez Ocanto, M., Rettig, T., & Ungefucht, M. (2023). Experimental evidence on panel conditioning effects when increasing the surveying frequency in a probability-based online panel. *Survey Research Methods: SRM*, 17(3), 323–339. [https://madoc.bib.uni-mannheim.de/66055/1/7990\\_artcl.pdf](https://madoc.bib.uni-mannheim.de/66055/1/7990_artcl.pdf)
- Eckman, S., & Bach, R. (2021). Panel Conditioning in the US Consumer Expenditure Survey. *Journal of Official Statistics (JOS)*, 37(1).
- Enders, C., & Bandalos, D. (2001). The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(3), 430–457.  
[https://doi.org/10.1207/S15328007SEM0803\\_5](https://doi.org/10.1207/S15328007SEM0803_5)
- Frick, J. R., Goebel, J., Schechtman, E., Wagner, G. G., & Yitzhaki, S. (2006). Using Analysis of Gini (ANOGI) for Detecting Whether Two Subsamples Represent the Same Universe: The German Socio-Economic Panel Study (SOEP) Experience. *Sociological Methods & Research*, 34(4), 427–468.  
<https://doi.org/10.1177/0049124105283109>
- GESIS (2022). GESIS Panel Standard Edition. GESIS Data Archive, Cologne. ZA5665 Datafile Version 46.0.0, doi: 10.4232/1.14007

- Halpern-Manners, A., Warren, J. R., & Torche, F. (2017). Panel conditioning in the general social survey. *Sociological Methods & Research*, 46(1), 103–124.
- Hoffmann, J. P., & Johnson, S. M. (2005). Attitudes toward abortion among religious traditions in the United States: Change or continuity? *Sociology of Religion*, 66(2), 161–182.
- Huang, Y., Davies, P. G., Sibley, C. G., & Osborne, D. (2016). Benevolent sexism, attitudes toward motherhood, and reproductive rights: A multi-study longitudinal examination of abortion attitudes. *Personality and Social Psychology Bulletin*, 42(7), 970–984.
- Jagodzinski, W., Kühnel, S. M., & Schmidt, P. (1987). Is there a „Socratic Effect“ in Nonexperimental Panel Studies?: Consistency of an Attitude toward Guestworkers. *Sociological Methods Research*, 15, 259–302.  
<https://doi.org/10.1177/0049124187015003004>
- Kalton, G., Kasprzyk, D., & McMillen, D. (1989). *Panel surveys*. New York: Wiley.
- Karpov, V., & Kääriäinen, K. (2005). “Abortion culture” in Russia: Its origins, scope, and challenge to social development. *Journal of Applied Sociology*, 2, 13–33.
- Kartsounidou, E., Kluge, R., Silber, H., & Gummer, T. (2023). Survey experience and its positive impact on response behavior in longitudinal surveys: Evidence from the probability-based GESIS Panel. *International Journal of Social Research Methodology*, 1–14.
- Kline, R. B. (2023). *Principles and practice of structural equation modeling*. Guilford publications.
- Kraemer, F., Silber, H., Struminskaya, B., Sand, M., Bosnjak, M., Koßmann, J., & Weiß, B. (2023). Panel Conditioning in A Probability-based Longitudinal study: A Comparison of Respondents with Different Levels of Survey Experience. *Journal of Survey Statistics and Methodology*, smad004.



- Kraemer, F., Silber, H., Struminskaya, B., Weiß, B., Bosnjak, M., Koßmann, J., & Sand, M. (2023). Satisficing Response Behavior Across Time: Assessing Negative Panel Conditioning Using an Experimental Design with Six Repetitions. *Survey Research Methods*, 17(3), 269–300. <https://ojs.ub.uni-konstanz.de/srm/article/view/7986>
- Kroh, M., Winter, F., & Schupp, J. (2016). Using Person-Fit Measures to Assess the Impact of Panel Conditioning on Reliability. *Public Opinion Quarterly*, 80(4), 914–942. <https://doi.org/10.1093/poq/nfw025>
- Kruse, Y., Callegaro, M., Dennis, J. M., Subias, S., Lawrence, M., DiSogra, C., & Thompson, T. (2009). *New Methodological Research on Web Panel Conditioning and Attrition*. Accuracy's Impact on Research. A Knowledge Networks Newsletter.
- Lazarsfeld, P. F. (1940). „Panel“ studies. *The Public Opinion Quarterly*, 4(1), 122–128.
- McDonald, R. P. (2013). *Test theory: A unified treatment*. psychology press.
- Nancarrow, C., & Cartwright, T. (2007). Online access panels and tracking research. The conditioning issue. *International Journal of Market Research*, 49(5), 573–594.
- Osborne, D., & Davies, P. G. (2012). When Benevolence Backfires: Benevolent Sexists' Opposition to Elective and Traumatic Abortion 1. *Journal of Applied Social Psychology*, 42(2), 291–307.
- Osborne, D., Huang, Y., Overall, N. C., Sutton, R. M., Petterson, A., Douglas, K. M., Davies, P. G., & Sibley, C. G. (2022). Abortion attitudes: An overview of demographic and ideological differences. *Political Psychology*, 43, 29–76.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of statistical software*, 48, 1–36. <https://www.jstatsoft.org/article/view/v048i02>
- Scherpenzeel, A. C., & Zandvliet, R. (2011). *Slapers en inactieven binnen online panels*. Tilburg University, School of Economics and Management. <https://ideas.repec.org/p/tiu/tiutis/0eb1dc85-8b8c-4802-ace1-496499154ee4.html>
- Schulz, M., Minderop, I., Weyandt, K. (2021). GESIS Panel Wave Report: Wave he (GESIS

- Panel Wave Report). GESIS - Leibniz Institute for the Social Sciences.
- Silber, H., Schröder, J., Struminskaya, B., Stocké, V., & Bosnjak, M. (2019). Does panel conditioning affect data quality in ego-centered social network questions? *Social Networks*, 56, 45–54. <https://doi.org/10.1016/j.socnet.2018.08.003>
- Stadtmüller, S., Schwerdtfeger, M., Weyandt, K. (2022). GESIS Panel Wave Report: Wave id (GESIS Panel Wave Report). GESIS - Leibniz Institute for the Social Sciences.
- Struminskaya, B. (2016). Respondent Conditioning in Online Panel Surveys: Results of Two Field Experiments. *Social Science Computer Review*, 34(1), 95–115. <https://doi.org/10.1177/0894439315574022>
- Sturgis, P., Allum, N., & Brunton-Smith, I. (2009). Attitudes over time: The psychology of panel conditioning. *Methodology of longitudinal surveys*, 113, 126.
- Sun, H., Tourangeau, R., & Presser, S. (2019). Panel Effects: Do the Reports of Panel Respondents Get Better or Worse over Time? *Journal of Survey Statistics and Methodology*, 7(4), 572–588. <https://doi.org/10.1093/jssam/smy021>
- Warren, J. R., & Halpern-Manners, A. (2012). Panel Conditioning in Longitudinal Social Science Surveys. *Sociological Methods & Research*, 41(4), 491–534. <https://doi.org/10.1177/0049124112460374>
- Waterton, J., & Livesley, D. (1989). Evidence of conditioning effects in the British Social Attitudes Panel. In D. Kasprzyk, G. Duncan, G. Kalton, & M. P. Singh (Hrsg.), *Panel Surveys* (S. 319–339). Wiley.
- Yalch, R. F. (1976). Pre-election Interview Effects on Voter Turnout. *Public Opinion Quarterly*, 40, 331–336.
- Yan, T., & Eckman, S. (2012). Panel Conditioning: Change in True Value versus Change in Self-Report. *Survey Research Methods*, 4726–4736.

## Online Appendix

### Appendix A: Question wordings and implementation

#### Attitudes toward abortion

Below you will find statements about potential reasons for a legal abortion. In Germany, an abortion is considered legal until the 12th week of pregnancy. We are interested in your personal opinion on this topic and to what extent you agree or disagree with the following statements.

- It should be possible for a pregnant woman to have a legal abortion if she is married and no longer wants children.
- It should be possible for a pregnant woman to have a legal abortion if the family has a very low income and cannot afford to have any more children.
- It should be possible for a pregnant woman to have a legal abortion if the probability of a serious birth defect of the baby is high.
- It should be possible for a pregnant woman to have a legal abortion if the woman's health is seriously endangered by the pregnancy.

1 Strongly disagree – 7 Strongly agree; 99 Don't know

#### Attitude certainty

The previous questions were about potential reasons that might lead a pregnant woman into having a legal abortion.

Do you have a clear opinion on legal abortions or not?

1 I do not have a clear opinion on that – 7 I have a clear opinion on that; 99 Don't know

#### Knowledgeability

In the following, we would like to ask you one more question about the reasons that may lead a pregnant woman into having a legal abortion.

How familiar are you with this topic?

1 Not familiar at all – 7 Very familiar; 99 Don't know

**Appendix B: Outcome rates****Table B.1***Completion rates (COMR) per wave in the non-probability panel (in percent)*

Wave 1	93.90
Wave 2	80.70
Wave 3	90.42
Wave 4	92.57
Wave 5	93.16
Wave 6	92.87

**Table B.2***Attrition rates (ATTR) per wave and respondent group in the non-probability panel (in percent)*

Wave 1	-
Wave 2	19.30
Wave 3	27.16
Wave 4	32.80
Wave 5	37.43
Wave 6	42.00

*Note.* All attrition rates refer to those respondents who completed the first wave.**Table B.3***Completion rates (COMR) per wave and panel cohort in the probability-based panel (in percent)*

	First cohort	Second cohort	Third cohort
Wave 1	93.17	93.73	89.72
Wave 2	94.82	94.00	91.11
Wave 3	94.00	93.42	91.82
Wave 4	93.03	92.44	91.26
Wave 5	93.06	93.27	91.48
Wave 6	92.24	92.16	90.26

*Note.* Completion rates are provided for each panel cohort, separately.

**Table B.4**

*Cumulative response rates (CUMR) per wave and panel cohort in the probability-based panel (in percent)*

	First cohort	Second cohort	Third cohort
Wave 1	11.45	9.68	9.62
Wave 2	11.51	9.54	9.58
Wave 3	11.24	9.32	9.41
Wave 4	11.01	9.08	9.12
Wave 5	10.93	9.10	8.99
Wave 6	10.69	8.84	8.75

*Note.* The cumulative response rates provided by the GESIS Panel are dependent on recruitment and profile rates which are specific for each recruited panel cohort. CUMR is therefore, provided for each panel cohort, separately.

**Table B.5**

*Attrition rates per wave of our experimental study and per panel cohort in the probability-based panel (in percent)*

	First cohort	Second cohort	Third cohort
Wave 1	46.60	36.55	24.33
Wave 2	47.25	37.66	25.82
Wave 3	48.04	38.71	27.69
Wave 4	48.60	39.65	29.50
Wave 5	48.97	40.06	30.62
Wave 6	49.64	41.05	31.61

*Note.* As the GESIS Panel is an already established panel infrastructure comprising more than 50 waves, all attrition rates refer to those respondents who completed the first survey wave after recruitment. Consequently, attrition rates are differentiated for each recruited panel cohort.

**Table B.6**

*Attrition rates in Wave 6 of the experimental study per respondent group in the non-probability panel (in percent)*

Fully conditioned group	Medium conditioned group	Unconditioned group
44.12	41.32	47.15

*Note.* The attrition rates refer to those respondents who completed the first wave.

**Table B.7**

*Attrition rates in Wave 6 of the experimental study per respondent group in the probability-based panel (in percent)*

<b>Fully conditioned group</b>	<b>Medium conditioned group</b>	<b>Unconditioned group</b>
1.41	1.29	1.55

*Note.* As the GESIS Panel is an already established panel infrastructure comprising more than 50 waves, we calculated specific attrition rates for Wave 6 of our experimental study, referring to the respondents who completed the first wave of our study. We additionally differentiated for the different respondent groups. The rates are calculated based on the procedure for the calculation of attrition rates per panel cohort, which are provided by the GESIS Panel (GESIS, 2023).

## Appendix C: Multiple-group confirmatory factor analysis

**Table C.1**

Unstandardized parameter estimates in the experimental groups for the scalar invariance model

Experimental group/ parameter estimates	$\lambda_1^x$	$\lambda_2^x$	$\lambda_3^x$	$\lambda_4^x$	$\tau_{x1}$	$\tau_{x2}$	$\tau_{x3}$	$\tau_{x4}$	$\theta_1^\delta$	$\theta_2^\delta$	$\theta_3^\delta$	$\theta_4^\delta$
<i>Probability-based panel</i>												
<b>Fully conditioned (prob)</b>	1.96	1.99	1.25	0.77	5.10	5.00	6.26	6.54	1.09	0.74	0.22	0.45
<b>Medium conditioned (prob)</b>	1.96	1.99	1.25	0.77	5.10	5.00	6.26	6.54	1.01	0.79	0.41	0.39
<b>Unconditioned (prob)</b>	1.96	1.99	1.25	0.77	5.10	5.00	6.26	6.54	1.23	1.01	0.27	0.49
<i>Non-probability panel</i>												
<b>Fully conditioned (non-prob)</b>	1.96	1.99	1.25	0.77	5.10	5.00	6.26	6.54	0.62	1.01	0.13	0.41
<b>Medium conditioned (non-prob)</b>	1.96	1.99	1.25	0.77	5.10	5.00	6.26	6.54	0.95	1.39	0.29	0.59
<b>Unconditioned (non-prob)</b>	1.96	1.99	1.25	0.77	5.10	5.00	6.26	6.54	1.59	1.41	0.45	0.50

**Table C.2**

Reliability coefficients for both factors across the experimental groups

<b>Experimental group</b>	<b>Reliability <math>\omega_3</math></b>	
	Elective factor	Traumatic factor
<i>Probability-based panel</i>		
<b>Fully conditioned</b>	0.90	0.85
<b>Medium conditioned</b>	0.88	0.83
<b>Unconditioned</b>	0.86	0.87
<i>Non-probability panel</i>		
<b>Fully conditioned</b>	0.91	0.90
<b>Medium conditioned</b>	0.85	0.85
<b>Unconditioned</b>	0.82	0.80

The results for the elective factor show that reliability consistently rises with the level of conditioning for both groups of respondents: probabilistic and non-probabilistic panelists. Fully conditioned respondents show the highest reliability coefficients. For the traumatic factor, we only observe this pattern for respondents of the non-probability panel. For respondents of the probability-based panel, we observe a less consistent pattern with unconditioned respondents showing the highest reliability coefficient and only an increased reliability of the fully conditioned respondent group compared to medium conditioned respondents. Similar to the residual variances of the abortion items, fully conditioned respondents of the non-probability panel show even higher reliability than those of the probability-based study.



## Appendix D: Multiple-group confirmatory factor analysis – Test of equal factor means

**Table D.1**

Model fit statistics of measurement model with equal factor means across the experimental groups

Model/Test	$\chi^2$	df	CFI	RMSEA	AIC	$\Delta\chi^2$	$\Delta df$
<b>Scalar invariance model</b>	29.62	26	>.999	.013	60523	11.85	10
<b>Model with equal factor means</b>	41.70	36	.999	.014	60515	12.08	10

*Note.* AIC = Akaike's Information Criterion; RMSEA = root mean square error of approximation; CFI = Comparative-Fit Index; \*\*\* p<0.001 \*\* p<0.01 \* p<0.05.

**Table D.2**

Standardized differences in latent factor means across the six experimental groups

	<b>Elective factor</b>	<b>Traumatic factor</b>
	Mean difference	Mean difference
<b>Fully conditioned (prob) vs. medium conditioned (prob)</b>	-.233 (.041)	-.021 (.043)
<b>Fully conditioned (prob) vs. unconditioned (prob)</b>	-.208 (.042)	-.627 (.046)
<b>Fully conditioned (prob) vs. fully conditioned (non-prob)</b>	-.056 (.073)	.360 (.077)
<b>Fully conditioned (prob) vs. medium conditioned (non-prob)</b>	-.440 (.068)	.252 (.078)
<b>Fully conditioned (prob) vs. unconditioned (non-prob)</b>	-.179 (.077)	-.539 (.076)
<b>Medium conditioned (prob) vs. Unconditioned (prob)</b>	.025 (.047)	.648 (.053)
<b>Medium conditioned (prob) vs. fully conditioned (non-prob)</b>	.178 (.076)	.381 (.081)
<b>Medium conditioned (prob) vs. medium conditioned (non-prob)</b>	-.206 (.071)	.273 (.082)

# ATTITUDES OVER TIME

<b>Medium conditioned (prob) vs. unconditioned (non-prob)</b>	.054 (.080)	-.518 (.080)
<b>Unconditioned (prob) vs. fully conditioned (non-prob)</b>	.152 (.076)	-.267 (.082)
<b>Unconditioned (prob) vs. medium conditioned (non-prob)</b>	-.232 (.072)	-.375 (.084)
<b>Unconditioned (prob) vs. unconditioned (non-prob)</b>	.029 (.081)	-1.166 (.082)
<b>Fully conditioned (non-prob) vs. medium conditioned (non-prob)</b>	-.384 (.093)	-.108 (.104)
<b>Fully conditioned (non-prob) vs. unconditioned (non-prob)</b>	-.123 (.100)	-.899 (.102)
<b>Medium conditioned (non-prob) vs. unconditioned (non-prob)</b>	.261 (.097)	-.792 (.103)

*Note.* Standard errors in parentheses.

## Appendix E: Longitudinal confirmatory factor analysis

**Table E.1**

Model fit statistics for tested models across the six survey waves for fully conditioned respondents from the probability-based panel

Model/Test	$\chi^2$	df	CFI	RMSEA	AIC	$\Delta\chi^2$	$\Delta df$
<b>Configural invariance model</b>	236.91***	126	.998	.019	140994	-	-
<b>Metric invariance model</b>	246.89***	136	.998	.019	140984	9.98	10
<b>Scalar invariance model</b>	269.40***	146	.997	.019	140987	22.51*	10
<b>Partial scalar invariance model<sup>a</sup></b>	260.18***	145	.998	.019	140980	13.29	9
<b>Strict invariance model</b>	388.50***	164	.995	.024	141070	128.33***	19

Note. AIC = Akaike's Information Criterion; RMSEA = root mean square error of approximation; CFI = Comparative-Fit Index; \*\*\*  $p < 0.001$  \*\*  $p < 0.01$  \*  $p < 0.05$ , <sup>a</sup> in comparison to the metric invariance model.

*Configural invariance over time* was tested by fitting the two-factor structure of the measurement model across the six waves, while freely estimating the model's parameters (i.e., factor loadings, item intercepts, residual variances) over time. For the longitudinal comparison of the measurement model, we allow residuals of the abortion items to correlate over the six measurement points (i.e., correlated measurement errors). The results show that our longitudinal two-factor measurement model of abortion attitudes fits the data well ( $\chi^2(126) = 236.91$ ,  $p < .001$ , RMSEA = .019, CFI = .998, AIC = 140,994), according to the rules of thumb for model fit (Kline, 2023).

After establishing the two-factor structure of the measurement model over time we then test for *metric invariance over time*, restricting all factor loadings to be equal across the six measurement points. We find that metric invariance holds and equal factor loading over

time are supported in the data ( $\Delta\chi^2(10) = 9.98, p = .44$ ; AIC = 140,984). As a next step, we test for *scalar invariance over time* by constraining all factor loadings and the item intercepts to be equal across the survey waves. Our results show that the longitudinal scalar invariance model does fit the data significantly worse than the longitudinal metric invariance model ( $\Delta\chi^2(10) = 22.51, p = .01$ ; AIC = 140,987), so that only *partial scalar invariance* can be established across the six measurement points ( $\Delta\chi^2(9) = 13.29, p = .15$ ; AIC = 140,980). In the partial scalar invariance model, we freely estimated the intercept of the fourth abortion item (abortion due to endangerment of the woman's health by the pregnancy) in Wave 1, while we restricted the remaining item intercepts to be equal across time (for detailed results, see Table E.2). After establishing equal factor loadings and equal item intercepts (except from the fourth item in Wave 1), we test for *strict invariance* over time. Similar to the MG-CFA, the results of the longitudinal CFA show that strict invariance is not supported in the data ( $\Delta\chi^2(19) = 128.33, p < .001$ ; AIC = 141,070) and that the residual variances of the abortion items are significantly different across the six measurement points.

**Table E.2**

Unstandardized parameter estimates for longitudinal partial scalar invariance model for fully conditioned respondents from the probability-based panel

Survey wave/ parameter estimates	$\lambda_1^x$	$\lambda_2^x$	$\lambda_3^x$	$\lambda_4^x$	$\tau_{x1}$	$\tau_{x2}$	$\tau_{x3}$	$\tau_{x4}$	$\theta_1^\delta$	$\theta_2^\delta$	$\theta_3^\delta$	$\theta_4^\delta$
<b>Wave 1</b>	1.93	1.92	1.36	.90	5.05	5.0	6.03	6.42	1.14	1.03	.54	.56
<b>Wave 2</b>	1.93	1.92	1.36	.90	5.05	5.0	6.03	6.37	.91	1.03	.37	.47

<b>Wave 3</b>	1.93	1.92	1.36	.90	5.05	5.0	6.03	6.37	1.08	1.05	.25	.46
<b>Wave 4</b>	1.93	1.92	1.36	.90	5.05	5.0	6.03	6.37	.90	.88	.26	.46
<b>Wave 5</b>	1.93	1.92	1.36	.90	5.05	5.0	6.03	6.37	.88	.82	.24	.47
<b>Wave 6</b>	1.93	1.92	1.36	.90	5.05	5.0	6.03	6.37	.97	.86	.27	.44

**Table E.3**

Reliability coefficients for both factors across the six survey waves for fully conditioned respondents from the probability-based panel

	<b>Reliability <math>\omega_3</math></b>	
	Elective factor	Traumatic factor
<b>Wave 1</b>	0.87	0.82
<b>Wave 2</b>	0.88	0.86
<b>Wave 3</b>	0.87	0.87
<b>Wave 4</b>	0.89	0.86
<b>Wave 5</b>	0.90	0.86
<b>Wave 6</b>	0.90	0.85

**Table E.4**

Model fit statistics for tested models across the six survey waves for fully conditioned respondents from the non-probability panel

<b>Model/Test</b>	<b><math>\chi^2</math></b>	<b>df</b>	<b>CFI</b>	<b>RMSEA</b>	<b>AIC</b>	<b><math>\Delta\chi^2</math></b>	<b><math>\Delta df</math></b>
<b>Configural invariance model</b>	214.23***	126	.987	.039	21570	-	-
<b>Metric invariance model</b>	240.37***	136	.985	.041	21576	26.14**	10
<b>Partial metric invariance model<sup>a</sup></b>	223.77***	135	.987	.038	21561	9.53	9

<b>Scalar invariance model</b>	236.57***	145	.987	.038	21554	12.80	10
<b>Strict invariance model</b>	314.67***	164	.979	.045	21594	78.10***	19

Note. AIC = Akaike's Information Criterion; RMSEA = root mean square error of approximation; CFI = Comparative-Fit Index; \*\*\*  $p < 0.001$  \*\*  $p < 0.01$  \*  $p < 0.05$ , <sup>a</sup> in comparison to the configural invariance model.

We also investigated measurement invariance of abortion attitudes over time for the fully conditioned respondents of the non-probability panel. Following the pre-defined series of model comparisons, we first fitted our two-factor measurement model to the longitudinal data of the non-probability panel to test *configural invariance over time*. Fit statistics of the configural invariance model are sufficiently good ( $\chi^2(126) = 214.23$ ,  $p < .001$ , (RMSEA) = .039, (CFI) = .987, (AIC) = 21,570). Further tests on metric invariance show that only *partial metric invariance over time* can be established in the non-probability panel with factor loadings being equal over time except for the factor loading of abortion item 3 at Wave 6 ( $\Delta\chi^2(9) = 9.53$ ,  $p = .39$ ; AIC = 21,561). We further find that item intercepts are significantly equal across the six survey waves, establishing *scalar invariance over time* ( $\Delta\chi^2(10) = 12.80$ ,  $p = .24$ ; AIC = 21,554). Similar to our analysis of the fully conditioned respondents in the probabilistic panel, model fit statistics further show that a more restricted model with equal residual variances across the six measurement points fits the data significantly worse ( $\Delta\chi^2(19) = 78.10$ ,  $p < .001$ ; AIC = 21,594). Accordingly, *strict invariance over time* is not supported in the data and the items' residual variances are significantly different across the six survey waves. The detailed parameter estimates of the scalar invariance model (see Table E.5) show in line with the analysis of respondents from the probability-based panel and our postulated *Hypothesis 2* an overall decrease in residual variances and smaller residual variances in subsequent waves compared to the first wave of the study (see Column  $\overline{\theta\delta}$ ). However, estimates of the model's parameters show negative residual covariances for the third abortion

item across selected measurement points (i.e.,  $\Gamma_{6,2}$ ,  $\Gamma_{6,3}$ ,  $\Gamma_{6,4}$ ). A possible reason for this might be the small sample size ( $n = 449$ ) combined with the complexity of our fitted model ( $n_{parameter} = 218$ ), affecting the estimation of parameters. We, therefore, did not proceed to test relationships in the latent factors over time for the respondents of the non-probability panel.

**Table E.5**

Standardized parameter estimates for longitudinal scalar invariance model for fully conditioned respondents from the non-probability panel

Survey wave/ parameter estimates	$\lambda_1^x$	$\lambda_2^x$	$\lambda_3^x$	$\lambda_4^x$	$\tau_{x1}$	$\tau_{x2}$	$\tau_{x3}$	$\tau_{x4}$	$\theta_1^\delta$	$\theta_2^\delta$	$\theta_3^\delta$	$\theta_4^\delta$	$\overline{\theta^\delta}$
<b>Wave 1</b>	.87	.87	.87	.79	2.29	2.31	4.15	5.37	.25	.25	.24	.38	<b>.28</b>
<b>Wave 2</b>	.88	.90	.88	.81	2.31	2.38	4.37	5.72	.23	.19	.23	.35	<b>.25</b>
<b>Wave 3</b>	.90	.88	.90	.84	2.28	2.25	4.34	5.77	.20	.23	.19	.30	<b>.23</b>
<b>Wave 4</b>	.93	.94	.91	.80	2.30	2.35	4.34	5.50	.14	.12	.18	.36	<b>.20</b>
<b>Wave 5</b>	.92	.90	.88	.80	2.28	2.25	4.02	5.18	.15	.19	.22	.37	<b>.23</b>
<b>Wave 6</b>	.94	.88	.98	.78	2.33	2.20	4.26	6.06	.12	.23	.04	.39	<b>.20</b>

**Table E.6**

Unstandardized parameter estimates for longitudinal scalar invariance model for fully conditioned respondents from the non-probability panel

Survey wave/ parameter estimates	$\lambda_1^x$	$\lambda_2^x$	$\lambda_3^x$	$\lambda_4^x$	$\tau_{x1}$	$\tau_{x2}$	$\tau_{x3}$	$\tau_{x4}$	$\theta_1^\delta$	$\theta_2^\delta$	$\theta_3^\delta$	$\theta_4^\delta$
<b>Wave 1</b>	1.90	1.86	1.28	.94	5.01	4.96	6.10	6.40	1.18	1.14	.52	.54
<b>Wave 2</b>	1.90	1.86	1.28	.94	5.01	4.96	6.10	6.40	1.06	.85	.45	.44
<b>Wave 3</b>	1.90	1.86	1.28	.94	5.01	4.96	6.10	6.40	.96	1.13	.37	.37
<b>Wave 4</b>	1.90	1.86	1.28	.94	5.01	4.96	6.10	6.40	.67	.55	.36	.48
<b>Wave 5</b>	1.90	1.86	1.28	.94	5.01	4.96	6.10	6.40	.72	.90	.51	.56
<b>Wave 6</b>	1.90	1.86	1.60	.94	5.01	4.96	6.10	6.40	.56	1.16	.09	.43

**Table E.7**

Reliability coefficients for both factors across the six survey waves for fully conditioned respondents from the non-probability panel

	Reliability $\omega_3$	
	Elective factor	Traumatic factor
<b>Wave 1</b>	0.85	0.83
<b>Wave 2</b>	0.89	0.84
<b>Wave 3</b>	0.89	0.86
<b>Wave 4</b>	0.93	0.85
<b>Wave 5</b>	0.91	0.83
<b>Wave 6</b>	0.91	0.90



**Table E.8**

Model fit statistics of first-order autoregressive models

Model/Test	$\chi^2$	df	CFI	RMSEA	AIC	$\Delta\chi^2$	$\Delta df$
<b>Simplex model w/ unequal autoregressive coefficients</b>	2033.48	195	.960	.064	142653	-	-
<b>Simplex model w/ equal autoregressive coefficients</b>	2095.37	203	.959	.063	142699	61.89***	8

Note. AIC = Akaike's Information Criterion; RMSEA = root mean square error of approximation; CFI = Comparative-Fit Index; \*\*\* p<0.001 \*\* p<0.01 \* p<0.05.

**Table E.9**

Unstandardized estimates of first-order model with unequal autoregression coefficients

	$\beta_{w2, w1}$	$\beta_{w3, w2}$	$\beta_{w4, w3}$	$\beta_{w5, w4}$	$\beta_{w6, w5}$
<b>Elective factor</b>	.82	.89	.92	.96	.92
<b>Traumatic factor</b>	.80	.86	.88	.91	.84

## Appendix F: Structural equation model including subjective attitude certainty and knowledgeability

**Table F.1**

Model fit statistics of the structural equation models including attitude certainty and knowledgeability

<b>Model/Test</b>	<b><math>\chi^2</math></b>	<b>df</b>	<b>CFI</b>	<b>RMSEA</b>	<b>AIC</b>	<b><math>\Delta\chi^2</math></b>	<b><math>\Delta df</math></b>
<b>Multiple group structural equation model w/ covariates</b>	33.50	30	>.999	.012	95522	-	-
<b>Structural equation model w/ equal covariate means</b>	48.59	40	.999	.016	95517	15.08	10
<b>Structural equation model w/ equal covariate means and equal regression slopes</b>	79.48*	60	.998	.024	95508	30.89	20

*Note.* AIC = Akaike's Information Criterion; RMSEA = root mean square error of approximation; CFI = Comparative-Fit Index; \*\*\* p<0.001 \*\* p<0.01 \* p<0.05.

## Appendix G: Comparison of respondents from the probability-based panel and the non-probability panel

**Table G.1**

Model fit statistics of models including equality constraints between respondents of the probability-based and the non-probability panel

<b>Model/Test</b>	<b><math>\chi^2</math></b>	<b>df</b>	<b>CFI</b>	<b>RMSEA</b>	<b>AIC</b>	<b><math>\Delta\chi^2</math></b>	<b><math>\Delta df</math></b>
<b>Scalar invariance model</b>	29.62	26	>.999	.013	60523	11.85	10
<b>Model with equal residual variances between prob and non-prob respondents</b>	70.61***	38	.996	.033	60539	40.99***	12

*Note.* AIC = Akaike's Information Criterion; RMSEA = root mean square error of approximation; CFI = Comparative-Fit Index; \*\*\* p<0.001 \*\* p<0.01 \* p<0.05.

## Appendix H: Comparison of panel cohorts within probability-based panel for each conditioning level

**Table H.1**

Model fit statistics of tested models across the cohorts of the probability-based panel (fully conditioned respondents)

<b>Model/Test</b>	<b><math>\chi^2</math></b>	<b>df</b>	<b>CFI</b>	<b>RMSEA</b>	<b>AIC</b>	<b><math>\Delta\chi^2</math></b>	<b><math>\Delta df</math></b>
<b>Configural invariance model</b>	3.40	3	>.999	.014	25649	-	-
<b>Metric invariance model</b>	5.93	7	>.999	.000	25644	2.53	4
<b>Scalar invariance model</b>	10.64	11	>.999	.000	25641	4.71	4
<b>Strict invariance model</b>	20.93	19	>.999	.012	25635	10.29	8

*Note.* AIC = Akaike's Information Criterion; RMSEA = root mean square error of approximation; CFI = Comparative-Fit Index; \*\*\*  $p < 0.001$  \*\*  $p < 0.01$  \*  $p < 0.05$ .

**Table H.2**

Model fit statistics of tested models across the cohorts of the probability-based panel (medium conditioned respondents)

<b>Model/Test</b>	<b><math>\chi^2</math></b>	<b>df</b>	<b>CFI</b>	<b>RMSEA</b>	<b>AIC</b>	<b><math>\Delta\chi^2</math></b>	<b><math>\Delta df</math></b>
<b>Configural invariance model</b>	7.58	3	.998	.067	12578	-	-
<b>Metric invariance model</b>	8.98	7	.999	.029	12571	1.40	4
<b>Scalar invariance model</b>	12.71	11	.999	.021	12567	3.73	4
<b>Strict invariance model</b>	21.41	19	.999	.019	12560	8.69	8

*Note.* AIC = Akaike's Information Criterion; RMSEA = root mean square error of approximation; CFI = Comparative-Fit Index; \*\*\*  $p < 0.001$  \*\*  $p < 0.01$  \*  $p < 0.05$ .

**Table H.3**

Model fit statistics of tested models across the cohorts of the probability-based panel  
(unconditioned respondents)

<b>Model/Test</b>	<b><math>\chi^2</math></b>	<b>df</b>	<b>CFI</b>	<b>RMSEA</b>	<b>AIC</b>	<b><math>\Delta\chi^2</math></b>	<b><math>\Delta df</math></b>
<b>Configural invariance model</b>	6.01	3	.998	.055	12887	-	-
<b>Metric invariance model</b>	9.41	7	.999	.032	12883	3.40	4
<b>Scalar invariance model</b>	13.19	11	.999	.024	12878	3.78	4
<b>Strict invariance model</b>	23.96	19	.997	.028	12873	10.77	8

*Note.* AIC = Akaike's Information Criterion; RMSEA = root mean square error of approximation; CFI = Comparative-Fit Index; \*\*\* p<0.001 \*\* p<0.01 \* p<0.05.