The Consistency of Question-order Bias in a Changing Political Context

Six Large-scale Surveys on Trust and Perceptions of Pandemic Governance Effectiveness

Authors and institutions

Wouter Van Dooren[1], Morten Hjortskov[2], Steven F. De Vadder[1], and Koen Verhoest[1]

[1]Research Group Politics & Public Governance, University of Antwerp, Belgium
[2]VIVE - The Danish Center for Social Science Research, Denmark

Corresponding author:

Wouter Van Dooren, Faculty of Social Sciences, Sint-Jacobsstraat 2, 2000 Antwerpen, Belgium

Wouter.vandooren@uantwerpen.be

T +32 3 265 55 94

Bio's

Wouter Van Dooren is a professor of public administration in the Research Group Politics & Public Governance of the University of Antwerp. His research interests include public governance, performance information, accountability and learning, and productive conflict in public participation.

Morten Hjortskov is senior researcher at VIVE - The Danish Center for Social Science Research. His research interests include quality, organization and implementation in the area of children and young people. He also studies the interaction between the citizen and the public administration as well as survey methodology and experiments.

Steven De Vadder is PhD researcher in the Research Group Politics & Public Governance of the University of Antwerp. His research considers the performance-satisfaction gap and trust in governments. He studies social media communication by public sector organizations through automated text analysis.

Koen Verhoest is research professor in Comparative Public Administration and Globalisation at the research Group on Politics & Public Governance, University of Antwerp (Belgium). He is also promotor

of the GOVTRUST Centre of Excellence on Trust in Multi-level Governance, one of the 13 Centres of Excellence at the University of Antwerp.

Acknowledgement

We would like to thank the Research Foundation Flanders (FWO - G085719N) for supporting the COVID research (see the 'Grote Corona-studie', https://www.uantwerpen.be/nl/projecten/coronastudie/). This paper has also benefited from discussions within the UAntwerp GOVTRUST Centre of Excellence and the EGPA Study Group on Behavioural Public Administration.

1 Introduction

Media, politicians, and decision-makers in the public sector use surveys to gauge attitudes and behaviour of the population. During the COVID-19 pandemic in particular, several large-scale projects using survey methods informed decision-makers on, among other things, citizens' willingness to keep a distance to others, accept vaccines, and trust governments. Insights gained from such surveys are widely used, and form part of the basis of political strategies and political decision-making. Survey methods and the design of surveys are, therefore, essential for the quality of the knowledge obtained and the decisions made (Van de Walle and Van Ryzin 2011).

However, there is also strong evidence that survey measures are context dependent (Schuman and Presser 1981; Tourangeau et al. 1989; Strack 1992; Sudman, Bradburn, and Schwarz 1996; Celhay, Meyer, and Mittag 2022). A classic case of a context effect in surveys arises from the order of the questions presented to the respondent. Prior questions can affect how subsequent questions are answered, resulting in question-order bias (Van de Walle and Van Ryzin 2011; Thau et al. 2021). Question-order bias has the potential to change the results of items themselves (levels) as well as their correlation with other items, even when the items do not seem connected in terms of topic (Tourangeau et al. 1989; Bard and Weinstein 2017; Hjortskov 2017). The effects are potentially large, perhaps even outshining the effects of actual public scandals (Thau et al. 2021).

Current knowledge about question-order bias rests to a large extent on small-scale studies that are rarely replicated (Schuman and Presser 1981, Appendix A; Stark et al. 2020; Thau et al. 2021). This may explain why some of the basic questions regarding question-order bias remain only partly answered and

why the possible explanations of when and why question-order effects arise, and which direction they have, are numerous (Schuman and Presser 1981; Tourangeau and Rasinski 1988; Strack 1992; Sudman, Bradburn, and Schwarz 1996; Tourangeau, Rips, and Rasinski 2000; Moore 2002).

In this study, we present the results of six question-order experiments. The experiments were embedded in six large-scale COVID-19 surveys in Flanders, Belgium, carried out during the COVID-19 pandemic from May to December 2020. The overall aim of the surveys was to measure how the population experienced the crisis, with items on crisis perception, financial stress, mental health, and support for COVID measures. Part of the survey addressed government effectiveness and trust in the government. In each of the six surveys, one question-order experiment was carried out.

The question-order experiments randomly varied according to whether respondents first received a series of four evaluative items on the effectiveness of the COVID policy or a series of five items on the respondent's trust in the federal, regional, and local governments, as well as their trust in the European Union (EU) and in experts. Governmental effectiveness and trust in government are often conceptualized as being tightly interwoven through the performance-trust link (Bouckaert and Walle 2003; Bundi and Pattyn 2022), and both have been especially important during the COVID-19 pandemic (Agley 2020; Kreps and Kriner 2020). In the first three surveys, the order of the trust questions was as listed above, but in the last three surveys, this order was reversed (experts, EU, local, regional, and federal). In other words, the six survey experiments were divided into two sets consisting of two different survey experiments carried out three times.

Results show that the question-order bias between effectiveness and trust measures is quite probable and fairly consistent across replications and changes of saliency in the Belgian pandemic context. The direction of the bias changes when the internal order of the trust questions is reversed. When the 'trust in experts' question is asked first (rather than the 'trust in federal government' question), trust levels are affected negatively by the treatment (i.e., presenting the effectiveness questions first). These results highlight an under-appreciated point: prior questions in the question-order sense should also be thought of

as the first questions within an outcome battery, because these can also affect the direction of question-order bias.

We believe we contribute in three ways. First, we contribute methodologically to the progressing replication agenda in Public Administration (see for instance Guo, 2023; Hansen 2023 for recent examples). by providing data from large-scale surveys and by doing two replications of each of the two initial experiments (six experiments in total). Most studies cited in the literature have lower power and there are only a few attempts at replication. A remarkable amount of these replications lead to different conclusions than the original research (Schuman and Presser 1981, Appendix A; Bishop, Oldendick, and Tuchfarber 1985; Klein et al. 2014; Stark et al. 2020; Thau et al. 2021).

Second, we contribute by replicating the question-order hypothesis in the changing political circumstances of Belgium in 2020 and with the important case of questions on government effectiveness and trust in government. Not only did Belgium, like many other countries, experience immense political, democratic, and structural pressure from the COVID-19 pandemic, the country also had a governmental crisis with a long coalition formation in the study period. These mega-events create a high-salience context in which we test the question-order hypothesis with survey questions on citizens' perceptions of government effectiveness and their trust in government – indicators that could potentially change because of these events (Agley 2020; Rieger & Wang 2022). If question-order bias remains strong and stable under these circumstances, this underlines its general power.

Third, we contribute by showing that not only does question-order bias exist between different blocks of questions, it also seems to exist within them. The last three of our experiments flip the order of the trust questions, resulting in all effects of the treatment becoming negative, indicating that questions within an outcome battery can affect question-order bias by themselves.

2 Theory

It has long been known that attitude measurement is context dependent, and that the order of questions may determine the answers (Payne 1951). However, our knowledge remains scarce concerning

the conditions under which these question-order effects appear and which direction they may take (Krosnick and Alwyn 1987; Schwarz and Strack 1991; Stark et al. 2018; Thau et al. 2021).

2.1 Answering Questions - Context Effects in Surveys

When we ask survey questions, the respondent is faced with several tasks: interpreting the question, retrieving and using information stored in memory, generating (or recalling) an opinion, translating the response into some numeric scale, and perhaps editing the answer because of social desirability or other situational circumstances. The order of the questions in a survey may influence all of these processes. The tasks that are perhaps most open to contextual influences are the interpretation of the question and the retrieval and use of information (Strack and Martin 1987).

Asking good survey questions is an art. All questions will be attempts at asking the respondent to simplify otherwise complex attitudes and beliefs. However, some questions are more abstract and ambiguous than others, and the *interpretation* of the question, therefore, is essential. When ambiguity is high, respondents will most likely look for help for interpretation in the context. A respondent's context includes the surrounding context of the world they live in and the survey context of survey layout, question format, length and, of course, question order. Many respondents most likely assume that researchers (or whoever is asking questions) would like to receive relevant answers to their questions. So they look for clues in the survey context, which after all is constructed by the same people who want the answers (Schwarz and Strack 1991, 34). For example, if preceding questions in the survey had a common theme or topic, this may help the respondent to interpret the next question's meaning, or if the initial question is clearly a specific part of a more extensive set of questions, perhaps the present question is too? However, the survey context that respondents might use to interpret the questions may not always result in meaningful information (Sudman, Bradburn, and Schwarz 1996).

The *retrieval of information* used for answering a question is the next possible source of question-order bias. People are unlikely to retrieve all relevant information when answering questions, but most likely they will rely on a subset of the information. If the interpretation of the question has already determined

which information is relevant, selection and retrieval of a subset of that information will be the next task. Theories of context effects posit that information/belief accessibility is critical in this process. Respondents only retrieve information until they assess they have enough information to inform their judgment or evaluation with a satisfactory level of certainty. As a consequence, the most accessible information will have a higher likelihood of retrieval and therefore have a significant influence on the final answer (Krosnick and Alwyn 1987; Tourangeau et al. 1989).

Finally, the use of retrieved information is part of what makes up context effects in surveys and is thought to be influential in the direction of the possible bias (see below). Furthermore, in some instances, the retrieved information may be disregarded by the respondent if it is deemed redundant or unusable when trying to answer a question (Sudman, Bradburn, and Schwarz 1996). For example, as part of a conversational norm of providing quality information to the researcher, the respondent may reject information that is clearly irrelevant or is not new in the survey context, because a related question has been asked already (Schwarz, Strack, and Mai 1991; Strack, Martin, and Schwarz 1988).

2.2 Question-order Bias

Interpretation, retrieval, and use are essential parts of answering questions, and for context effects in general. However, question order in a survey can play a vital role in these processes because it determines what was seen and processed as the last thing before answering the current question. Therefore, the information retrieved to answer those (previous) questions is most likely highly accessible. In effect, people only recover and process a subset of the relevant information – possibly the information that comes to mind most easily. Recently viewed and answered questions may play a significant role in this process. As the order of questions generally should not influence how a question is responded to, the influence of question order is often called a bias (Tourangeau & Rasinski 1988; Schwarz et al. 1991; Sudman, Bradburn, and Schwarz 1996).

Question-order bias has been found in several studies across disciplines (see, e.g., Strack 1992; Auh 2003; Bowling and Windsor 2008; Van de Walle and Van Ryzin 2011; Bard and Wardstein 2017; Stark et

al. 2020; Thau et al. 2021). The extent to which the bias occurs and its size are described in the literature as dependent on attitude and belief accessibility as well as five other factors: recency of activation, frequency of activation, relation between the questions, ambiguity of the question and background of the respondent (Tourangeau et al. 2000; Thau et al. 2021).

Recency and frequency relate to how recently and how many times the question with possible influence on the current question has been encountered. The more recent and higher the frequency, the higher the probability of a question-order bias (Thau et al. 2021). The closer the relationship between the questions is perceived by the respondent, the higher the probability for a bias. If the ambiguity of the questions asked is high, there will generally be a higher probability of different biases slipping in, and this most likely also applies for question-order bias (Van de Walle and Van Ryzin, 2011, p. 1440; Hjortskov 2017). Furthermore, a general presumption in the question-order literature is that the background of respondents (especially high education and expertise) also plays a role, such that certain types of background are thought to protect against the question-order bias (Thau et al. 2021). Results on the influence of respondent background on the question-order bias are mixed, however (Stark et al. 2020).

Although there is no guarantee that question-order bias will result if a question-order scores highly within the five factors identified above (recency of activation, frequency of activation, relation between the questions, ambiguity of the question and background of the respondent), the probability of question-order bias should be higher.

2.2.1 The Direction of Question-order Bias: The Inclusion/Exclusion Model

Although a lot of evidence points to the existence of question-order bias, there is quite some debate about its direction. Much theory centres on two possible ways that the retrieved information can be used: assimilation and contrast (Sudman, Bradburn, and Schwarz 1996; Schwarz and Bless 1992a). In the following, we describe the Inclusion/Exclusion Model (IEM), which is one way of conceptualizing and predicting context effects and their direction based on assimilation and contrast theory (Schwarz and Bless 1992b).

In the IEM, the use of the retrieved information will result in a representation of the target (the question), but it may also result in a second representation in the form of a standard. Both of these representations are thought to be partly context dependent and hence likely influenced by preceding questions in the survey. Unsurprisingly, they may also be affected by information that is always accessible – and therefore context independent – which could, for example, reflect respondent characteristics such as education or expertise, as mentioned above (Sudman, Bradburn, and Schwarz 1996, 101-102; Stark et al. 2020).

If the evoked information from the previous questions is included in the temporary representation that respondents form of the target of judgment, the process is called *assimilation*. This basically means that any thoughts that might have been evoked by preceding questions, whether they be positive or negative, will be included in the final judgment and result in a positive correlation between previous and present questions. It would be assimilation if effectiveness questions about the government, on average, make respondents retrieve positive information about the government that they include in their judgment when subsequently asked about their trust in government. Inclusion of the retrieved information results in assimilation in the IEM (Bless & Schwarz 2010).

Instead of including the retrieved information, the respondent might exclude it. Importantly, exclusion is not ignoring the information (this would not result in question-order effects). Instead, exclusion is using the retrieved information from previous questions in another way that results in a contrast effect instead of an assimilation effect. *Contrast effects* can arise from two different exclusion processes according to the IEM. The first process is when the retrieved information is excluded from the evaluation or judgment required for the present question, creating a negative correlation between the previous and the current questions (Sudman, Bradburn, and Schwarz 1996). In our setting, answering the effectiveness questions might remind the respondents that effectiveness is not necessarily a relevant part of their evaluation of trust in government. If so, the (in this example, positive) retrieved information about effectiveness will be excluded from the judgment about trust. This is sometimes called a subtraction-based contrast effect, because the effectiveness evaluation is mentally subtracted from the trust evaluation (Schuman and Presser

1981; Sudman, Bradburn, and Schwarz 1996; Bless and Burger 2016). The result is that less positive information is included in the trust evaluation than would otherwise be the case, and therefore lower trust is the result – the opposite direction compared to assimilation.

The other process by which contrast effects can arise according to the IEM is when the retrieved information from the preceding questions is not only excluded from the representation of the target, but is also used to construct a comparison standard. The standard is then used in comparison-based contrast, in which trust is compared to the standard based on effectiveness. An example could be if government effectiveness is considered high, creating a high comparison standard against which the retrieved information on trust is compared. In comparison, the retrieved trust information may seem low, and therefore the reported trust will be lower than it would have been if the effectiveness questions were not asked first. This may vary according to the targets of the effectiveness and trust evaluations: are the respondents asked to evaluate independent experts or the EU?

In the IEM, assimilation is considered the default (Schwarz and Bless 1992a; Sudman, Bradburn, and Schwarz 1996, 112). Some studies have hypothesized that contrast effects require additional cognitive work compared to assimilation. Within the category of contrast effects, standards are thought to exert their effects over several of the subsequent questions as long as they are perceived relevant. In contrast, subtraction-based contrast effects are typically limited to a single question, where only the information that is left can be used (Sudman, Bradburn, and Schwarz 1996, 106-107).

2.2.2 The Mixed Results on Question-order Bias

In the literature, a number of questions are only superficially answered, and the empirical evidence is mixed (McFarland 1981; Schuman and Presser 1981; Stark et al. 2020). As the above section shows, the direction of question-order effects is hard to predict, and empirical studies also sometimes show different directions of similar treatments delivered in different contexts. For example, Thau et al. (2021) seek to replicate the question-order effects in Van de Walle and Van Ryzin (2011) but end up with a sizable effect in the opposite direction. In this case, the explanation for the opposite findings might be the specific contexts

and the relation between the specific questions and the overall satisfaction question that was the focus of these two different question-order experiments.

Context-wise, the sample in Van de Walle and Van Ryzin (2011) was a random sample of citizens who answered satisfaction questions on a range of governmental services and a question on their overall satisfaction with governmental services. In Thau et al. (2021), on the other hand, the sample consisted of well-educated professionals, who answered several specific questions about their satisfaction with a government agency that they had (successfully) applied for funding from, and an overall satisfaction question about the whole process. In terms of the relation between questions , the specific questions (i.e. not the overall question) in Van de Walle and Van Ryzin (2011) had different governmental services as their topic, while in Thau et al. (2021) all the specific questions concerned different parts of the funding process. The result was that placing the specific questions before the overall question in this setting resulted in lower overall satisfaction in Van de Walle and Van Ryzin (2011), but the opposite in Thau et al. (2021).

According to the logic of IEM theory, an explanation for the differing results in the two studies above could be that the more substantial topical relations within the specific questions as well as between those questions and the overall question in Thau et al. (2021) resulted in assimilation. In contrast, the more broad set of specific questions and their perhaps weaker relation to the overall question could have resulted in subtraction-based contrast in Van de Walle and Van Ryzin (2011) (Thau et al. 2021, 200). However, other factors could play a role as well, and one key point here is that we need more well-powered, exact replications to assess the direction of question-order effects.

2.2.3 Context and Replication in Question-order Effects

It would seem that the context of the specific question-order study may make a difference to the results. The context of the current study is the changing intensity of the COVID-19 pandemic and the handling of it in Belgium (see below for more details). We expect that the change in intensity also changes the saliency of the crisis and therefore the respondents' perceived importance of government effectiveness and trust in government (see e.g., Agley 2020; Kreps and Kriner 2020). In attitude research, attitude

importance is expected to cause higher attitude accessibility, consistency and stability (Boninger et al. 1995; Howe and Krosnick 2017). This could cause the possible question-order effects to diminish across the more intensive stages of the crisis.

However, Stark et al. (2020) find only some differences in classic question-order experiments carried out across different countries, and Fabrigar & Krosnick (1995) find no moderation by attitude importance on the false consensus effect in six different experiments. When attempting to replicate the famous Hyman and Sheatsley (1950) study, Klein et al. (2014) only replicated the question-order effect in 36 percent of the cases across different contexts. However, this might be explained by low power and the study still showed an overall significant question-order effect when the studies were combined (Stark et al. 2020, p. 592-593).

The mixed empirical results in the question-order literature in terms of both direction and context, and the pre-eminence of low-powered experiments also point to the need for more replication studies in the question-order literature (Klein et al. 2014). Below, we present our case, design, and data, which we believe add significant value to the question-order literature through two different order experiments that are both carried out 3 times. We use Pedersen and Stritch's (2018) Relevance, Number, Internal validity, Contextual realism, and External validity (RNICE) model to assess the value of our replications. First, the relevance of our replication study is high because surveys are used extensively by both researchers and practitioners, and we still know relatively little about the determinants of question-order effects. Second, the number of other replications is low, and even fewer are exact replications. We replicate our two own question-order studies twice in the same country (Belgium) during the changing political circumstances of the COVID-19 pandemic, so the third and fourth criteria, internal validity, and contextual realism, are equally high in the first experiments and the subsequent replications. In terms of the fifth replication criterion in the RNICE model, external validity, all our experiments are carried out in the same population and country with large citizen samples, but with political circumstances that change across the replications. This tests the generalizability of the causal relationship in our question-order experiments in spite of variation in the settings (Pedersen and Stritch, p. 609). It should be noted, though, that our samples are opt-in samples.

We hypothesize that we indeed will see question-order effects, but that they will partly depend on the status of the ongoing pandemic and the handling of it in Belgium. The mixed findings of the existing research make it difficult to clearly predict a direction of the effects. This approach makes our study more exploratory, with the main aim being estimation and interpretation of the directions of the effects and effect sizes.

2.3 Public Trust in Government and Experts

The context of our study is citizen trust in government (local, regional, federal and EU) and in experts in Belgium during the COVID-19 pandemic. We therefore not only rely on epistemic trust, i.e. trust in experts, but also trust in governmental actors (Bundi and Pattyn 2022). Studies show that trust is highly correlated with acceptance and compliance with restrictions introduced in various forms across countries during the pandemic (Bundi and Pattyn 2022; Agley 2020; Kreps and Kriner 2020). We use the definition of trust given by Mayer, Davis, and Schoorman (1995, 712):

...the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party (Mayer, Davis, and Schoorman 1995, 712)

One of the most enduring explanations of trust in government is government effectiveness and performance. The government performance-trust link predicts that perceived performance in government services and macroeconomic indicators such as inflation and unemployment will influence citizens' trust in government positively (Bouckaert et al. 2002; Van der Meer and Hakhverdian 2017; Hansen 2021). This also seems to have been the case during the COVID-19 crisis (Rieger and Wang 2022).

Therefore, in terms of possible question-order bias, the effectiveness and trust questions possibly have an initial topical relation which, as mentioned, is a prerequisite for question-order bias (Tourangeau, Rips, and Rasinski 2000; Hjortskov 2017; Thau et al. 2021). As the literature on the performance-trust link

often uses survey measures of both concepts, it is relevant to test the question-order bias hypothesis in this context.

3 Case, Design, and Data

The Belgian COVID-19 case is the context of the six surveys that contain our question-order experiments. The most important characteristic of this context for our replication study is that the crisis had different intensity levels during the years 2020 and 2021. Therefore, the effectiveness and trust questions we ask, and the order of them, are tested in quite varying contexts. The following section also presents the surveys and the experimental designs.

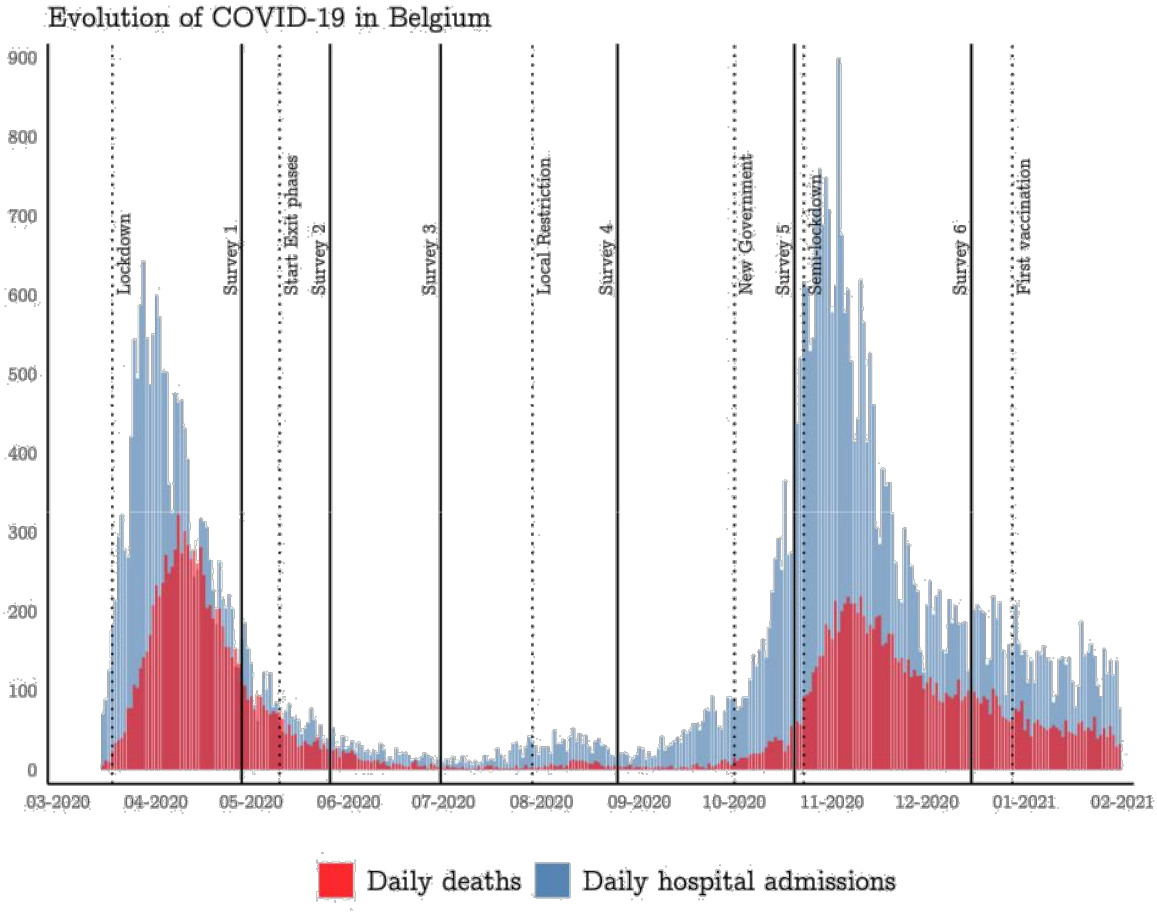
3.1 The Belgian COVID Pandemic

Belgium was strongly affected by the COVID-19 pandemic (Zaki et al. 2022, Six et al. 2021). The relative number of confirmed COVID-related deaths was substantially higher than the EU average for COVID-related deaths in relation to total population. The study surveys were conducted at various stages of the COVID-19 pandemic in Belgium (see Figure 1). The first survey was conducted on 28 April 2020. At that time, the first wave of infections was retreating, but lockdown measures were still in place. Belgium was still in a political crisis with a long period of coalition government formation after the general elections of 26 May 2019. The caretaker government was criticized for not providing personal protective equipment, but overall, support for the measures and trust in government was high. The pandemic drowned out media attention on all other issues, including attention on the protracted formation of a new government (Walgrave and Kuypers 2020). The relatively high levels of trust have been attributed to a rally-around-the-flag effect (Popelier et al. 2021; Sinardet and Pieters 2021; Baekgaard et al. 2020).

The second survey was conducted on 26 May 2020. Lockdown measures that had been in effect from 17 March 2020 were gradually eased as the pandemic lost force. Trust in government declined, and more critique of the handling of the crisis emerged. The dire situation in care homes was a core point of attention. The third survey was conducted on 30 June 2020 when the number of contaminations was low. Notably, trust on 30 June 2020 was at the lowest point of all six surveys. The fourth survey was conducted on 28 August 2020. At this time, cases had started to rise again, and several local governments implemented

new restrictions. The province of Antwerp had implemented a controversial curfew in early July 2020. However, trust in government was rising again. The fifth survey was conducted on 20 October 2020, shortly before the peak of the second wave. With the hospital system on the brink of collapse, the stress of the pandemic was again very high. Trust in government, however, was rising further. In October 2020, a new government was formed. The prime minister and minister of health took the leadership more firmly into their hands. The last survey was conducted in the wake of the second wave on 15 December 2020. At this time, trust in government reached its highest point across all six surveys.

Figure 1: Evolution of the COVID-19 Pandemic in Belgium



3.2 Survey and Experimental Design

The experiments were embedded in several editions of a weekly (and later biweekly) large-scale COVID survey. Respondents were recruited through social media, audiovisual media, and printed media. The respondents volunteered to join the survey. The response was very high in the first two surveys (108,415 in survey 1 and 46,020 in survey 2). Surveys 3 to 6 were completed by 20,000-25,000 respondents. In order to balance the dataset, we sampled 20,000 observations from each survey to make sure that each survey had an equal weighting in the pooled data set.

Participation in one or several surveys was not random, and our comparison across the surveys should therefore be interpreted with some caution. The self-selection of respondents into the surveys led to some deviations in the population characteristics of the study's dataset when compared with Belgium's general population characteristics. Higher educated people and women are overrepresented (see appendix 1). Older age cohorts are underrepresented in the first surveys, but not so much in the later surveys. Respondents from the province of Antwerp are strongly overrepresented, arguably because the survey emanated from the University of Antwerp. In the discussion section, we reflect on the implications of these deviations for our findings. Arguably, many respondents participated in several surveys, but we do not have unique identifiers across surveys to make an empirical assessment[1].

The experiment was inserted into the section of the survey that asked trust questions. First, all respondents answered a question on general trust in others. Next, respondents were randomly allocated to the treatment or control conditions (see Table 1). In the treatment condition, the respondents first had to fill out four items that asked about their perception of government effectiveness in handling of the crisis (see Table A4 in the appendix for question wording). The first item asked about general effectiveness, the second asked about effectiveness in reducing infections, the third asked about effectiveness in reducing the impact on the economy, and the fourth asked about effectiveness in dealing with mental welfare.

Next, the respondents were asked how much they trusted the federal government, the regional government, the local government, the European Union, and experts. In the fourth, fifth, and sixth

[1] One way repeated measurement might have an impact on the findings is if respondents recollect a different question order from previous waves, which would make them aware of the experiment. The question-order experiment was just one part of a larger survey. It seems unlikely that many respondents would remember the order of the questions and infer that a survey experiment was embedded, let alone answer the questions strategically afterwards.

experiment, we reversed the order of the trust items, i.e. first trust in experts, then the European Union, followed by local government, regional government, and finally federal government. The control group first received the trust items before answering the effectiveness items. In experiments 4-6, the question order in the battery of trust items (i.e. the outcome variable) was reversed. The change of question order within the battery of trust questions was not random since it happened between surveys: we changed the order in the last three surveys and did so for all respondents.

Table 1: Design of the Experiments: Survey Flow in Columns

Experiments 1-3		**Experiments 4-6**	
Treatment	**Control**	**Treatment**	**Control**
Effectiveness:		*Effectiveness:*	
General		General	
Infections		Infections	
Economy		Economy	
Mental Welfare		Mental Welfare	
Trust:	*Trust:*	*Trust:*	*Trust:*
Federal	Federal	Experts	Experts
Regional	Regional	EU	EU
Local	Local	Local	Local
EU	EU	Regional	Regional
Experts	Experts	Federal	Federal
	Effectiveness:		*Effectiveness:*
	General		General
	Infections		Infections
	Economy		Economy
	Mental Welfare		Mental Welfare

4 Results

Table 2 presents the descriptive statistics of the data. The effectiveness index is the mean of four items (general effectiveness, reducing infections, financial consequences, and mental welfare). By a narrow margin, effectiveness is judged to be adequate (4.2 on a 7-point scale). The trust index is the mean of the trust items in the federal government, regional government, local government, European Union, and in experts. The trust index is slightly higher than the effectiveness index because of the very high trust in experts (5.77). Trust in governments is lower than the effectiveness assessment[2].

Table 2: Descriptive Statistics

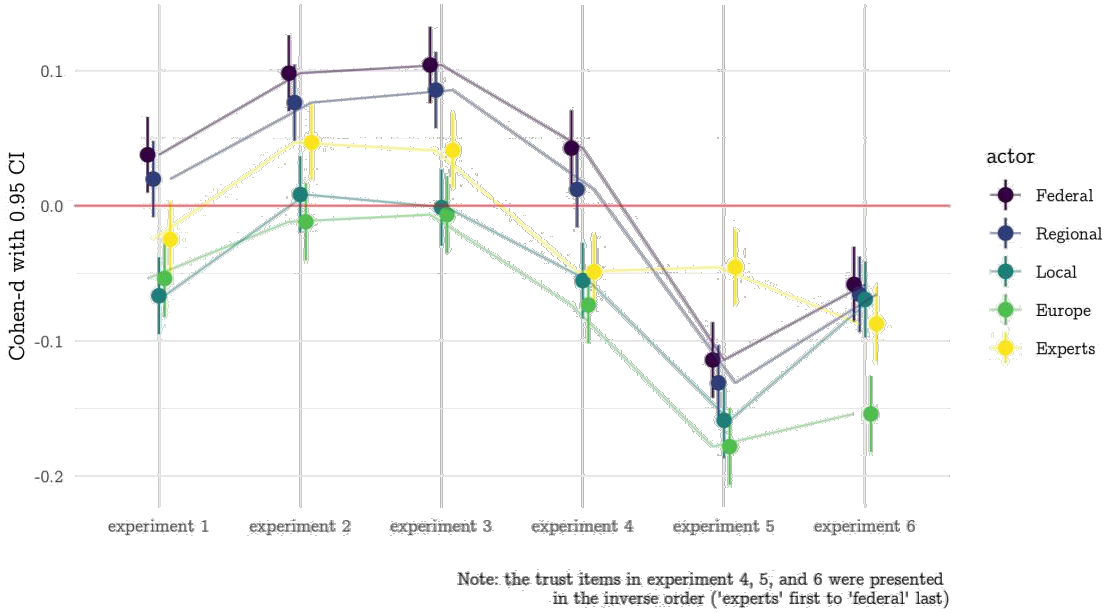
	N	Missing (%)	Mean	SD	Median	Min	Max
Effectiveness Index	111,636	7	4.20	1.30	4.25	1.00	7.00
Trust Index	110,259	8	4.35	1.28	4.40	1.00	7.00
Trust in Federal	117,479	2	4.11	1.70	4.00	1.00	7.00
Trust in Regional	116,304	3	4.00	1.59	4.00	1.00	7.00
Trust in Local	115,603	4	4.36	1.60	5.00	1.00	7.00
Trust in EU	113,360	6	3.59	1.69	4.00	1.00	7.00
Trust in Experts	118,449	1	5.77	1.39	6.00	1.00	7.00

Note: The descriptives in Table 2 are based on a random sampling of 20,000 respondents from each survey

Figure 2 shows the results for the different actors. The horizontal axis represents the six experiments that were embedded in the six surveys shown in Figure 1. The vertical access represents the effect size (Cohen's d) for the trust-in-government measures, including a 0.95 confidence interval. A positive number implies that seeing the government effectiveness items first increases trust in the government. In contrast, a negative number suggests that trust in government decreases for those respondents that first answered the effectiveness items. The exact estimates are in the appendix.

[2] Cronbach's alfa is 0.859 for the effectiveness index and 0.858 for the trust index

Figure 2: Effect of Seeing Government Effectiveness Items First (Results per Experiment)



The data suggest that question-order effects do exist. Twenty-three out of thirty Cohen's d estimates are significantly different than zero. The direction of the effects, however, is not straightforward. In experiments 1-3, answering the effectiveness items first increases trust in the federal and regional government. Trust in the local government is negatively affected in experiment 1, but the result is insignificant for experiments 2 and 3. Trust in the European Union decreases after answering effectiveness items in experiment 1, but the result is insignificant in experiments 2 and 3. Trust in experts is insignificant in experiment 1, but is positive and significant in experiments 2 and 3.

In experiments 4, 5, and 6, the question order of the dependent variables (i.e. the trust items) was reversed. In these experiments, we asked first for trust in experts, and next for trust in local government, the European Union, the regional government, and the federal government. As a result, the series of assessments of different governmental tiers that in experiments 1-3 appeared before the assessment of trust in experts was preceded by the assessment of trust in experts in experiments 4-6. In other words, in experiments 1-3,

the four items on effectiveness of government were immediately followed by items on trust in government. In experiments 4-6, the items on effectiveness of government were followed by an item asking for trust in experts, after which the trust in government items were asked.

The result of the above change in question order is that there seems to be a reversal in the direction of the question-order effects in experiments 4-6 compared to experiments 1-3. However, the drop in the effects in Figure 2 is most evident in experiments 5-6, while experiment 4 more resembles experiments 1-3. The rank order for the different government tiers is mostly preserved across all experiments, but the question-order effect on trust in experts seems to jump out of this order in experiments 5 and 6.

To dive a little more into the differences between the experiments and surveys, we have pooled the data from the six experiments and formed trust and effectiveness indices. We present the results from the pooled dataset in Table 3. The table presents the overall effect of the treatment on trust and the influence of the respondents' effectiveness perceptions using the pooled data of all six experiments and the effectiveness and trust indices. This enables us to address the general direction of the question-order effects conditional on the respondents' perceptions of the Belgian government's effectiveness. In terms of the replication aspect of this study, it also supplies us with more general estimates across the replications (see also Klein et al. (2014)).

Model 1 suggests that showing the effectiveness items first has a negative effect on trust. The effect of -0.038 is significant, as can be expected with a large sample. Asking the effectiveness items first reduces the trust index by 0.038 standard deviations. This effect may not be very strong, but notable. The second model includes a control variable for the effectiveness index to evaluate whether the level of the prior question influences the question-order effect. The very strong association of 0.7 between effectiveness and trust is not surprising given the performance-trust link (Bouckaert and Van de Walle 2003). The treatment effect becomes stronger when controlling for effectiveness, indicating that the level of the prior question is important for the question-order effect. Model 3 introduces an interaction term between effectiveness and the treatment. There is a small, negative interaction effect. People who have a more positive assessment of government effectiveness are more negatively affected by the question order. This result points to a contrast

effect of the question order: if people are reminded about their (high) evaluation of government effectiveness just before answering the trust questions, they have lower trust in government.

Table 3: Effect of Seeing Effectiveness Questions First on the Trust Index - Pooled Data[3]

	Model 1	Model 2	Model 3
Treatment (ref: control)	-0.038***	-0.101***	-0.102***
	[−0.050, −0.026]	[−0.110, −0.093]	[−0.110, −0.093]
	s.e. = 0.006	s.e. = 0.004	s.e. = 0.004
Effectiveness		0.721***	0.732***
		[0.716, 0.725]	[0.726, 0.738]
		s.e. = 0.002	s.e. = 0.003
Treatment (ref:control) x Effectiveness			-0.023***
			[0.032, 0.015]
			s.e. = 0.004
No. Obs.	110 259	104 796	104 796
R2	0.000	0.522	0.522
R2 Adj.	0.000	0.522	0.522
AIC	312866.3	220324.3	220296.6
BIC	312895.1	220362.5	220344.4

Standardized coefficients, † p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. 95% confidence intervals in brackets, s.e. = standard error

The flipped order of the outcome items seems to have a negative impact on the effect of the treatment. In the treatment group, trust is now lower for all items, except for trust in the federal government (positive) and trust in the regional government (insignificant) in experiment 4. The reversal of the order with the experts as the first question does seem to have a considerable effect. In Table A5 in the appendix,

[3] Note that the effectiveness variable is not randomized and is measured directly (not manipulated except for the order) in the survey. The effectiveness variable is used as an interaction variable with the randomized order variable estimating the conditional average treatment effect (CATE), which is often done in experimental research (Gerber & Green 2012, p. 296-299). We added two control variables from the orginal data: gender and education. As can be expected, gender (ref: male) and education are positively associated with trust. Yet, the interaction effect and the main effect remain stable. Adding additional control variables such as gender and education does not seem to alter the interaction effect of interest. The dataset pools 20,000 observations, sampled from six surveys. The main results are about the same if we estimate the models independently in each full sample. Therefore, we are fairly confident that the pooling in itself does not affect the main results. The interaction effect holds in four of the six samples.

the difference between experiments 1-3 and 4-6 is estimated as an interaction for each trust outcome question. All the interactions are significantly positive, meaning that, overall, the presentation of the federal question first in the trust outcome battery results in higher trust on all outcome questions compared to presenting the experts first.

5 Discussion

In this paper, we have presented a set of six question-order experiments embedded in six large-scale surveys conducted in Belgium during the COVID-19 pandemic of 2020. The experiments all randomly vary the order of a set of government effectiveness questions and a set of government trust questions. Three of the experiments presented respondents with the trust questions in one order, while the other three experiments presented respondents with the trust questions in the reverse order of the first three experiments.

The context of the pandemic might have affected our results. The results of experiments 3 and 4, both of which were carried out at a time with relatively low infection numbers but with a different order of trust questions, are not too different. In experiments 5 and 6, which were carried out when there were high infection numbers, we see large differences (compared with the other four experiments) in results when there are changes in the order of trust questions. However, we should note that experiment 1 was carried out at a time when infection numbers were similar to those for experiment 6, but with a different order of the trust questions, and the results are quite different. It is therefore not implausible that both the infection rate and the order of the trust questions could affect our results.

Despite the changing context of the experiments, the results show that question-order bias is present and fairly consistent across the replications and changing political conditions. With pooled data, we see that the treatment (seeing the effectiveness questions first) has an overall negative effect on the trust index (Table 3). The negative effect is larger when controlled for the respondents' effectiveness assessments, and interacting the treatment variable with the effectiveness assessment yields a significant, negative interaction effect. People who have a positive assessment of government effectiveness are more affected by the

treatment, i.e. the overall negative effect of the treatment is more negative for those who perceive government effectiveness as high.

The negative effect of the treatment and the negative interaction effect can be explained using the IEM. They seem to be subtraction-based contrast effects, since seeing the effectiveness questions first affects trust negatively, and this effect is emphasized for those with high government effectiveness perceptions. We posit that respondents react to the similar topic and feel that they have already answered part of the trust question by answering the effectiveness questions. They then perhaps conclude that the researchers probably are not interested in hearing the same perceptions twice and exclude this part of their trust assessment. Or perhaps they are reminded that effectiveness is something different than trust when they are presented with the effectiveness questions first. Therefore, we posit, they would subtract those perceptions from their trust perceptions. This results in lower trust and a negative relationship between effectiveness and trust, especially if respondents' effectiveness perceptions are high.

While the effects described above are the overall effects, Figure 2 shows that below the overall average, there are differences in different experiments. Responses to some of the trust questions — federal and regional government — are affected positively by the treatment in experiments 1-3. The context of these surveys was the first COVID wave. A 'rally-around-the-flag' effect may have led the effectiveness prompt to elicit more trust. In the first COVID wave, the federal and regional governments were most present in the media (Walgrave and Kuipers 2021). In later surveys, the effects become negative. The rank order between the levels of government is more or less preserved, with the exception of trust in experts in experiment 5.

Differences also emerge when we flip the order of the trust questions in experiments 4-6. While many of the effects of changing the order of the effectiveness and trust questions are positive in the federal-experts order in experiments 1-3, most of them are negative in experts-federal order in experiments 4-6. This points to the possibility of question-order effects within the outcome measurements. A test using the pooled data of the interaction between the treatment variable and a variable indicating whether the data are from either experiments 1-3 (federal government first) or experiments 4-6 (experts first) shows a positive

significant effect, indicating that the overall treatment effect is significantly higher in the first three experiments compared to the last three (see Table A5 in the appendix). It should be noted, though, that experiment 4 perhaps more resembles experiments 1-3 than 5-6, despite some of the treatment effects becoming significantly different from zero and a negative tendency. Experiments 5 and 6 are clearly different, with all treatment effects being negative and lower than the rest.

The study has its limitations. The fact that respondents opt into the survey has led to a dataset that deviates from the general population in three ways (see Table A1 in the appendix). Firstly, higher educated persons are overrepresented. Previous studies suggest that higher levels of education might protect against question-order effects (Thau et al., 2021). If this is the case, the results in our sample might underestimate the question-order effects in the population. Secondly, older people are underrepresented in the first two surveys, but not in the last four. The last four datasets are more comparable in terms of demographics. Yet, the most notable changes in effects are found between the last four surveys. Therefore, the reversal of the order of the outcome, in combination with changing contexts, seems to be the explanation for the changes in effects. Thirdly, women and people from the province of Antwerp are overrepresented. However, we do not have theoretical reasons to expect that order effects would play out differently based on gender or place of residence.

With fairly consistent question-order effects, this study fits well with much of the question-order literature. Our contribution offers a set of high-powered studies with replications in the same population but with changing circumstances. In terms of the RNICE model of replications (Pedersen and Stritch 2018), we contribute by having a high internal relevance of replication, and we deliver two-times-three exact replications in a field where replications are rare, with high internal validity and high contextual realism.

This study has some implications for research in the field of public administration in general. Our study shows that question-order effects are present in large-scale surveys and that they are replicable. Therefore, researchers should note that answers to questions might be influenced by the order in which they are presented – a feature that is rarely desirable. From the question-order literature, some general preconditions for question-order bias emerge: recency, frequency of activation, topic relation, and ambiguity

(Tourangeau et al. 2000; Thau et al. 2021). It is important as a researcher to be aware of these general preconditions when designing surveys, but it is also hard to avoid all five preconditions at once. Placing numerous, ambiguous questions on the same topic close to each other in the survey, however, may lead to ambiguous results.

Another implication for research is that question-order effects also seem to materialize within outcome batteries – in our case when the order of questions on trust in different actors was reversed. These resemble the response-order effects found in response-order studies (i.e., not question order but the order of response alternatives, e.g., Bishop 1987; Krosnick and Alwin 1987; Galesic et al. 2008), where participants focused more on the top response options than the lower options. The implication is that the first questions in outcome-batteries colour the answers to the rest of the questions. One easy-to-implement solution to both of these challenges to survey design is to randomize the order of questions, as long as the survey flow still makes sense to the respondent. This will even out the possible question-order bias and enable the researcher to assess the influence of possible question-order bias in large samples (Thau et al. 2021).

Our study also shows that the direction of question-order effects is still something that should be tested further in future research. We see both positive and negative effects along the way. Overall, the IEM seems good at describing possible question-order effects and their direction, but they are still very hard to predict. Further research is certainly warranted here, for example, one could perform systematic experiments that investigate the direction of question-order effects by randomly varying question order both between and within blocks of questions. Given the widespread use of public-opinion surveys among both researchers and practitioners, gaining more knowledge about how and when question order can cause differences in survey results is of great importance to the field.

References

Agley, Jon. 2020. “Assessing changes in US public trust in science amid the COVID-19 pandemic.” Publisher: Elsevier, Public health 183:122–125.

Auh, S., Salisbury, L.C., and Johnson, M.D. (2003) ‘Order Effects in Customer Satisfaction Modelling’, Journal of Marketing Management, 19(3–4), 379–400.

Baekgaard, Martin, Julian Christensen, Jonas Krogh Madsen, Kim Sass Mikkelsen, et al. 2020. "Rallying around the flag in times of COVID-19: Societal lockdown and trust in democratic institutions." Journal of Behavioral Public Administration 3 (2).

Bard, Gabriele, and Yana Weinstein. 2017. "The Effect of Question Order on Evaluations of Test Performance: Can the Bias Dissolve?" Publisher: SAGE Publications, Quarterly Journal of Experimental Psychology 70 (10): 2130–2140.

Biernat, Monica, and Scott Eidelman. 2007. "Standards." In Social Psychology: Handbook of Basic Principles, edited by Arie W. Kruglanski and E. Tory Higgins, 308–333. New York, NY: The Guilford Press.

Bishop, George F., Robert W. Oldendick, and Alfred J. Tuchfarber. 1985. "The Importance of Replicating a Failure to Replicate: Order Effects on Abortion Items." Public Opinion Quarterly 49 (1): 105.

Bishop, G.F. (1987) 'Context Effects on Self-Perceptions of Interest in Government and Public Affairs', in Hippler, H.-J., Schwarz, N. and Sudman, S., eds., Social Information Processing and Survey Methodology, Recent Research in Psychology, New York, NY: Springer, 179–199.

Bless, Herbert, and Axel M. Burger. 2016. "Assimilation and contrast in social priming." Current Opinion in Psychology 12:26–31.

Boninger, D.S., Krosnick, J.A., Berent, M.K., and Fabrigar, L.R. (1995) 'The Causes and Consequences of Attitude Importance', in Petty, R.E. and Krosnick, J.A., eds., Attitude Strenght: Antecedents and Consequences, London: Psychology Press.Bouckaert, Geert, Steven Van de Walle, Bart Maddens, and Jarl K. Kampen. 2002. Identity vs. Performance: An Overview of Theories Explaining Trust in Government. 2. Leuven: Public Management Institute.

Bouckaert, Geert, and Steven van de Walle. 2003. "Comparing Measures of Citizen Trust and User Satisfaction as Indicators of 'Good Governance': Difficulties in Linking Trust and Satisfaction Indicators." Publisher: SAGE Publications Ltd, International Review of Administrative Sciences 69 (3): 329–343.

Bundi, Pirmin, and Valérie Pattyn. 2022. "Trust, but verify? Understanding citizen attitudes toward evidence-informed policy making." _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/p Public Administration Early View:1–20.

Bowling, A. and Windsor, J. (2008) 'The effects of question order and response-choice on self-rated health status in the English Longitudinal Study of Ageing (ELSA)', Journal of Epidemiology & Community Health, 62(1), 81–85.

Celhay, Pablo A., Bruce D. Meyer, and Nikolas Mittag. 2022. What Leads to Measurement Errors? Evidence from Reports of Program Participation in Three Surveys. National Bureau of Economic Research.

Galesic, M., Tourangeau, R., Couper, M.P., and Conrad, F.G. (2008) 'Eye-Tracking Data: New Insights on Response Order Effects and Other Cognitive Shortcuts in Survey Responding', Public Opinion Quarterly, 72(5), 892–913.

Guo, Shihong. (2023). Policy advocacy of nongovernmental organizations in China: A quasi-replication of Zhan and Tang (2013, 2016). Public Administration, 1– 23. https://doi.org/10.1111/padm.12914

Hansen, Frederik Godt. 2021. "Trust in Governmental Institutions: How Government Performance Matters." PhD diss., Politicas ph.d.-serie.

Hansen, Paw. (2023). "A replication of "the effects of making public service employees aware of their prosocial and societal impact". Public Administration, 1– 14. https://doi.org/10.1111/padm.12913

Hjortskov, Morten. 2017. "Priming and Context Effects in Citizen Satisfaction Surveys." Public Administration 95 (4): 912–926.

Howe, L.C. and Krosnick, J.A. (2017) 'Attitude Strength', Annual Review of Psychology, 68(1), 327–351, available: https://doi.org/10.1146/annurev-psych-122414-033600.

Hyman, H.H. and Sheatsley, P.B. (1950) 'The Current Status of American Public Opinion', in Payne, J.C., ed., The Teaching of Contemporary Affairs, Twenty-First Yearbook of the National Council of Social Studies, 11–34.

Klein, R.A., Ratliff, K.A., Vianello, M., Adams, R.B., Bahník, Š., Bernstein, M.J., Bocian, K., Brandt, M.J., Brooks, B., Brumbaugh, C.C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W.E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E.M., Hasselman, F., Hicks, J.A., Hovermale, J.F., Hunt, S.J., Huntsinger, J.R., IJzerman, H., John, M.-S., Joy-Gaba, J.A., Barry Kappes, H., Krueger, L.E., Kurtz, J., Levitan, C.A., Mallett, R.K., Morris, W.L., Nelson, A.J., Nier, J.A., Packard, G., Pilati, R., Rutchick, A.M., Schmidt, K., Skorinko, J.L., Smith, R., Steiner, T.G., Storbeck, J., Van Swol, L.M., Thompson, D., van 't Veer, A.E., Ann Vaughn, L., Vranka, M., Wichman, A.L., Woodzicka, J.A., and Nosek, B.A. (2014) 'Investigating Variation in Replicability', Social Psychology, 45(3), 142–152.

Kreps, S. E., and D. L. Kriner. 2020. "Model uncertainty, political contestation, and public trust in science: Evidence from the COVID-19 pandemic." Publisher: American Association for the Advancement of Science, Science Advances 6 (43): eabd4563.

Krosnick, J.A. and Alwin, D.F. (1987) 'An Evaluation of a Cognitive Theory of Response-order Effects in Survey Measurement', Public Opinion Quarterly, 51(2), 201–219.

Mayer, Roger C., James H. Davis, and F. David Schoorman. 1995. "An Integrative Model Of Organizational Trust." Publisher: Academy of Management, Academy of Management Review 20 (3): 709–734.

McFarland, Sam G. 1981. "Effects of Question Order on Survey Responses." Public Opinion Quarterly 45 (2): 208–215.

Moore, David W. 2002. "Measuring New Types of Question-Order Effects Additive and Sub-tractive." Public Opinion Quarterly 66 (1): 80–91.

Payne, Stanley Le Baron. 1951. The Art of Asking Questions. Studies in Public Opinion 3. Princeton, NJ: Princeton University Press.

Pedersen, Mogens Jin, and Justin M. Stritch. 2018. "RNICE Model: Evaluating the Con-tribution of Replication Studies in Public Administration and Management Research." Public Administration Review 78 (4): 606–612.

Popelier, Patricia, Bjorn Kleizen, Carolyn Declerck, Monika Glavina, and Wouter Van Dooren. 2021. "Health crisis measures and standards for fair decision-making: A normative and empirical-based account of the interplay between science, politics and courts." European Journal of Risk Regulation 12 (3): 618–643.

Rieger, Marc Oliver, and Mei Wang. 2022. "Trust in Government Actions During the COVID-19 Crisis." Social Indicators Research 159 (3): 967–989.

Schuman, Howard, and Stanley Presser. 1981. Questions and answers in attitude surveys: Experiments on question form, wording, and context. New York: Academic Press.

Schwarz, Norbert, and Herbert Bless. 1992a. "Assimilation and Contrast Effects in Attitude measurement: an Inclusion/Exclusion Model." ACR North American Advances NA-19.

Schwarz, Norbert, and Herbert Bless. 1992b. Constructing reality and its alternatives: An inclusion/exclusion model of assimilation and contrast effects in social judgment. Lawrence Erlbaum Associates, Inc.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., and Simons, A. (1991) 'Ease of retrieval as information: Another look at the availability heuristic', Journal of Personality and Social psychology, 61(2), 195.

Schwarz, Norbert, and Fritz Strack. 1991. "Context effects in attitude surveys: Applying cognitive theory to social research." European review of social psychology 2 (1): 31–50.

Schwarz, Norbert, Fritz Strack, and Hans-Peter Mai. 1991. "Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis." Public opinion quarterly 55 (1): 3–23.

Sinardet, Dave, and Jade Pieters. 2021. “Nationalism and COVID in Belgium: A Surprisingly United Response in a Divided Federal Country.” Nationalism and Ethnic Politics 27 (3): 311–330.

Six, Frederique., de Vadder, Steven., Glavina, Monica., Verhoest, Koen and Koen Pepermans. 2021. “What drives compliance with COVID-19 measures over time? Explaining changing impacts with Goal Framing Theory.” Regulation & Governance. https://doi.org/10.1111/rego.12440

Stark, T.H., Silber, H., Krosnick, J.A., Blom, A.G., Aoyagi, M., Belchior, A., Bosnjak, M., Clement, S.L., John, M., Jónsdóttir, G.A., Lawson, K., Lynn, P., Martinsson, J., Shamshiri-Petersen, D., Tvinnereim, E., and Yu, R. (2020) ‘Generalization of Classic Question Order Effects Across Cultures’, Sociological Methods & Research, 49(3), 567–602.

Strack, Fritz. 1992. “Order Effects in Survey Research: Activative and Informative Functions of Preceding Questions.” In Context Effects in Social and Psychological Research, edited by Norbert Schwarz and Seymour Sudman. New York.

Strack, Fritz, Leonard Martin, and Norbert Schwarz. 1988. “Priming and Communication

Social Determinants of Information Use in Judgments of Life Satisfaction.” WOS:A1988R753900004, European Journal of Social Psychology 18 (5): 429–442.

Strack, Fritz, and Leonard L. Martin. 1987. “Thinking, Judging, and Communicating: A Process Account of Context Effects in Attitude Surveys.” In Social Information Process-ing and Survey Methodology, edited by Hans-J. Hippler, Norbert Schwarz, and Seymour Sudman, 123–148. Recent Research in Psychology. Springer New York.

Sudman, Seymour, Norman M. Bradburn, and Norbert Schwarz. 1996. Thinking about An-swers: The Application of Cognitive Processes to Survey Methodology. San Francisco, CA: Jossey-Bass.

Thau, Mads, Maria Falk Mikkelsen, Morten Hjortskov, and Mogens Jin Pedersen. 2021. “Question order bias revisited: A split-ballot experiment on satisfaction with public services among experienced and professional users.” Public Administration 99 (1): 189– 204.

Tourangeau, Roger, and Kenneth A. Rasinski. 1988. “Cognitive processes underlying context effects in attitude measurement.” Psychological Bulletin 103 (3): 299–314.

Tourangeau, Roger, Kenneth A. Rasinski, Norman Bradburn, and Roy D’Andrade. 1989. “Belief accessibility and context effects in attitude measurement.” Journal of Experimental Social Psychology 25 (5): 401–421.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. The Psychology of Survey Response. Cambridge, U.K. ; New York: Cambridge University Press.

Van de Walle, Steven, and Gregg G. Van Ryzin. 2011. “The order of questions in a survey on citizen satisfaction with public services: Lessons from a split-ballot experiment.” Public Administration 89 (4): 1436–1450.

Van der Meer, Tom, and Armen Hakhverdian. 2017. “Political Trust as the Evaluation of Process and Performance: A Cross-National Study of 42 European Countries.” Publisher:

SAGE Publications Ltd, Political Studies 65 (1): 81–102.

Walgrave, Stefaan, and Ine Kuypers. 2021. “Nieuwsmonitor 30: TV-nieuws in 2020: de coronastorm, grote veranderingen in het nieuws en de medische invalshoek”. Elektronisch Nieuwsarchief: Antwerpen.

Zaki, Bishoy, Pattyn, Valerie., and Ellen Wayenberg. 2022. “Policy learning type shifts during creeping crises: A storyboard of COVID-19 driven learning in Belgium.” European Policy Analysis, 1– 25.

Appendix

Table A1: survey waves (opt in) versus population characteristics

Variables	Population	Survey Wave 1	Survey Wave 2	Survey Wave 3	Survey Wave 4	Survey Wave 5	Survey Wave 6
Sample Size		*108 415*	*46 020*	*25 241*	*23 525*	*20 342*	*20 297*
Gender							
Male	49.1	28.6	27.5	29.4	29.5	30.2	33.2
Female	50.9	71.4	72.5	70.6	70.5	69.8	66.8
Age							
15-24	13.0	8.6	6.7	4.1	3.6	4.0	3.2
25-34	14.7	19.5	14.5	10.9	10.2	9.8	9.1
35-44	15.2	22.8	19.2	15.9	17.7	16.3	15.7
45-54	16.7	19.5	19.1	18.9	20.5	19.4	18.9
55-64	16.3	18.8	23.6	26.8	26.1	26.3	27.3
65+	24.1	10.8	16.9	23.4	21.9	24.2	25.8
Province							
Antwerp	28.0	47.9	50.5	49.4	49.9	48.8	47.3
Flemish Brabant	17.2	14.6	14.3	15.1	14.7	15.1	15.8
West Flanders	18.4	9.7	8.8	9.5	9.5	9.9	10.4
East Flanders	23.0	18.4	17.4	17.2	17.0	17.3	18.2
Limburg	13.4	9.4	9.0	8.8	8.8	8.8	8.4
Education							
Primary education	11.3	1.1	1.2	1.3	1.3	1.3	1.2
Secondary education	55.8	26.0	25.6	28.1	26.6	27.5	28.4
Bachelor-level education	16.2	37.6	37.6	37.3	37.9	37.2	37.2
Masters/PhD-level education	16.7	35.3	35.6	33.3	34.2	34.0	33.1

Table A2: Descriptives: means and standard deviation (sample size = 20,000)

Experiment 1	Treatment	Mean	4.290	4.192	4.323	3.436	5.958
		SD	1.638	1.575	1.621	1.693	1.265
	Control	Mean	4.228	4.160	4.431	3.529	5.990
		SD	1.686	1.622	1.619	1.735	1.292
Experiment 2	Treatment	Mean	4.179	3.981	4.192	3.443	5.740
		SD	1.568	1.530	1.576	1.617	1.323
	Control	Mean	4.022	3.863	4.179	3.462	5.676
		SD	1.634	1.566	1.609	1.641	1.399
Experiment 3	Treatment	Mean	3.934	3.842	4.099	3.327	5.825
		SD	1.608	1.553	1.582	1.603	1.324
	Control	Mean	3.764	3.709	4.102	3.338	5.770
		SD	1.640	1.568	1.624	1.647	1.358
Experiment 4	Treatment	Mean	3.663	3.861	4.439	3.507	5.509
		SD	1.642	1.581	1.599	1.681	1.520
	Control	Mean	3.592	3.841	4.529	3.631	5.583
		SD	1.687	1.642	1.645	1.699	1.524
Experiment 5	Treatment	Mean	4.072	3.907	4.316	3.540	5.651
		SD	1.727	1.561	1.569	1.649	1.419
	Control	Mean	4.273	4.115	4.567	3.836	5.715
		SD	1.783	1.614	1.583	1.675	1.408
Experiment 6	Treatment	Mean	4.566	4.196	4.487	3.850	5.865
		SD	1.702	1.563	1.556	1.709	1.390
	Control	Mean	4.666	4.299	4.594	4.112	5.982
		SD	1.712	1.566	1.541	1.693	1.304

Table A3: data for Figure 2

Experiment	Actor	Estimate	ConfInt Lower	Confint Upper
Experiment 1	Federal	- 0.038	-0.066	-0.010
	Regional	-0.020	-0.048	0.008
	Local	0.067	0.038	0.095
	Europe	0.054	0.025	0.083
	Experts	0.025	-0.003	0.053
experiment 2	Federal	-0.098	-0.126	-0.070
	Regional	-0.076	-0.105	-0.048
	Local	-0.008	-0.037	0.020
	Europe	0.012	-0.017	0.041
	Experts	-0.047	-0.075	-0.019
experiment 3	Federal	-0.104	-0.132	-0.076
	Regional	-0.086	-0.114	-0.057
	Local	0.001	-0.027	0.030
	Europe	0.007	-0.022	0.035
	Experts	-0.041	-0.069	-0.013
experiment 4	Federal	-0.043	-0.071	-0.015
	Regional	-0.012	-0.040	0.016
	Local	0.055	0.027	0.084
	Europe	0.074	0.045	0.102
	Experts	0.049	0.021	0.077
experiment 5	Federal	0.114	0.086	0.142
	Regional	0.131	0.103	0.159
	Local	0.159	0.131	0.187
	Europe	0.178	0.150	0.207
	Experts	0.046	0.018	0.073
experiment 6	Federal	0.058	0.030	0.086
	Regional	0.066	0.038	0.094
	Local	0.069	0.041	0.098
	Europe	0.154	0.126	0.182
	Experts	0.087	0.060	0.115

Table A4: Trust and Effectiveness Question Wording

Trust Questions	
Federal	To what extent do you trust the federal government to address the COVID-19 crisis adequately?
Regional	To what extent do you trust your regional government to address the COVID-19 crisis adequately
Local	To what extent do you trust your municipal or city government to address the COVID-19 crisis adequately?
EU	To what extent do you trust the European Union to address the COVID-19 crisis adequately?
Experts	To what extent do you trust the scientific experts to address the COVID-19 crisis adequately?
Effectiveness Questions	
General	To what extent do you think that public measures until now have been effective in addressing the COVID-19 crisis in general?
Infections	To what extent do you think that public measures until now have been effective in addressing the COVID-19 crisis with regards to reducing infections with the Coronavirus?
Economy	To what extent do you think that public measures until now have been effective in addressing the COVID-19 crisis with regards to reducing the negative economic consequences?
Mental Welfare	To what extent do you think that public measures until now have been effective in addressing the COVID-19 crisis with regards to reducing the negative social consequences?

Questions have been translated from Dutch

Table A5: Interaction effect of Treatment and Question Order

	Federal	Regional	Local	EU	Experts
General Effectiveness	0.742***	0.588***	0.507***	0.521***	0.445***
	[0.737, 0.746]	[0.583, 0.593]	[0.502, 0.512]	[0.515, 0.527]	[0.440, 0.449]
	s.e. = 0.002	s.e. = 0.003	s.e. = 0.003	s.e. = 0.003	s.e. = 0.002
Treatment (ref. control)	-0.108***	-0.124***	-0.174***	-0.250***	-0.103***
	[-0.129, -0.087]	[-0.146, -0.103]	[-0.196, -0.151]	[-0.275, -0.226]	[-0.123, -0.084]
	s.e. = 0.011	s.e. = 0.011	s.e. = 0.012	s.e. = 0.012	s.e. = 0.010
Trust Question Order (Federal to Experts)	-0.512***	-0.442***	-0.554***	-0.652***	-0.152***
	[-0.533, -0.491]	[-0.463, -0.420]	[-0.577, -0.531]	[-0.677, -0.628]	[-0.172, -0.132]
	s.e. = 0.011	s.e. = 0.011	s.e. = 0.012	s.e. = 0.013	s.e. = 0.010
Treatment*Trust Question Order	0.085***	0.096***	0.038*	0.105***	0.042**
	[0.056, 0.114]	[0.065, 0.126]	[0.006, 0.071]	[0.071, 0.140]	[0.014, 0.070]
	s.e. = 0.015	s.e. = 0.015	s.e. = 0.017	s.e. = 0.018	s.e. = 0.014
Num.Obs.	116,703	115,563	114,798	112,639	117,411
R2	0.435	0.315	0.237	0.231	0.235
R2 Adj.	0.435	0.315	0.237	0.231	0.235
AIC	388,219.0	391,121.6	402,893.2	407,609.3	378,568.8
BIC	388,277.0	391,179.6	402,951.1	407,667.1	378,626.8

† $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. 95% confidence intervals in brackets.