Journal of Behavioral and Experimental Economics 104 (2023) 102013

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

journal homepage: www.elsevier.com/locate/jbee

Priming and the value of a statistical life: A cross country comparison

Henrik Andersson [a,*], Benjamin Ouvrard [b,c]

[a] Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France
[b] Toulouse School of Economics, INRAE, University of Toulouse Capitole, Toulouse, France
[c] Université Grenoble Alpes, INRAE, UMR GAEL, Grenoble, France

ARTICLE INFO

Keywords:
Air pollution
Discrete choice experiments
Priming
Value of a statistical life
Willingness to pay

ABSTRACT

Using a discrete choice experiment this study examines whether different types of priming may influence the respondents' answers when choosing between different policies aimed at reducing the mortality risk due to ambient air pollution. We focus on two types of priming: (i) two versions of an oath where respondents commit to answer truthfully during the survey, and (ii) a priming scenario that combines information about the social cost of ambient air pollution and questions on the respondents' experiences related to the topic. To test the robustness of the findings the same survey is implemented in two different countries, the United States (US) and the United Kingdom (UK). Results show that respondents behave as expected in the choice situations and the two estimates of the value of statistical life (VSL) obtained are in line with values recommended for policy purposes in both countries. Regarding the priming treatments, we find that the oath treatments have different effects in the US and in the UK, and that the priming scenario has an effect on those who have already been suffering from air pollution (US), or on those who are willing to change and undertake actions to protect the environment (UK).

1. Introduction

Air pollution not only causes negative environmental effects but is also responsible for several negative health effects. It has been shown to cause, e.g., mental health issues (Lu, 2020; Zhang, Zhang, & Chen, 2017) and respiratory diseases (Beatty & Shimshack, 2014; Ruchiraset & Tantrakarnapa, 2020), but that it also can lead to premature death (Lelieveld, Evans, Fnais, Giannadaki, & Pozzer, 2015). According to some recent studies, the number of annual premature deaths due to fine particulate ambient air pollution are between 3.4 and 8.9 million (Burnett et al., 2018; Global Burden of Disease Study, 2017).

Improving air quality therefore has the potential to provide large social benefits from reducing health risks related to air pollution. Such policies come at a cost, though, and to ensure that society's resources are used efficiently the policies should be evaluated. One economic tool that can be used to provide information on whether the benefits of the measures exceed the costs is cost–benefit analysis (CBA). This tool favored by economists is founded in welfare theory and if the analysis shows that the benefits are larger than the costs, then it suggests that the policy is desirable for society. However, the use of CBA requires that all benefits and costs are measured in a common metric, which is usually money. Hence, it is necessary to monetize the different effects from the policy that do not have easily observable prices, like the reduced risk of dying from air pollution.

To empirically elicit preferences for health risk reductions and converting them into monetary values analysts rely on revealed- (RP) or stated-preference (SP) methods (see, e.g., Freeman, Herriges, & Kling, 2014). Whereas the former, i.e. RP methods, rely on individuals' actual decisions in markets, such as accepting to take a riskier job if financially compensated (Gentry & Viscusi, 2016), or paying more for a property in an area with better air quality (Chay & Greenstone, 2005), the latter, i.e. SP methods, rely on individuals' answers from hypothetical scenarios to measure their preferences. Economists have traditionally favored the RP approach since it relies on actual decisions. However, weaknesses with the approach include the necessity of markets (which can be an issue with public goods), access to good data (e.g. different environmental effects like noise and air pollution may be highly correlated making it difficult to identify the preferences for one of the effects), and the assumption that individuals make well-informed decisions in markets. Therefore, SP methods have gained ground due to their flexibility which allows them to construct the market scenario of interest, and to control the choice situations providing transparency for

* Corresponding author.
E-mail addresses: henrik.andersson@tse-fr.eu (H. Andersson), benjamin.ouvrard@inrae.fr (B. Ouvrard).

https://doi.org/10.1016/j.socec.2023.102013
Received 8 April 2022; Received in revised form 19 October 2022; Accepted 17 March 2023
Available online 3 April 2023

the analyst of the respondents' decision alternatives. The hypothetical nature of the SP approach is its main weakness, though, but how to address this has been a major research area and several manuals and guidelines on best practice for SP methods are available, including the recent one by Johnston et al. (2017).

Among SP methods, the two techniques that dominate are contingent valuation (CV) and discrete choice experiments (DCE) (see, e.g., Champ, Boyle, & Brown, 2017). Mahieu, Andersson, Beaumais, Crastes dit Sourd, Hess, and Wolff (2017) showed in their review that DCE have gained popularity among economists to elicit respondents' willingness to pay (WTP) for different non-market goods. The use of DCE has also gained ground when it comes to monetizing the value for health risk reductions (e.g., Adamowicz, Dupont, Krupnick, & Zhang, 2011; Andersson, Hole, & Svensson, 2016; Cameron, DeShazo, & Stiffler, 2010; Jin, Andersson, & Zhang, 2020). In many of the studies estimating the WTP for health risk reductions the focus has been on eliciting WTP to reduce mortality risks, which is commonly normalized and referred to as the value of a statistical life (VSL) (Andersson, Hole, & Svensson, 2019). In this study, in addition to eliciting the WTP for a policy that aims at reducing the mortality risk due to ambient air pollution, we use priming treatments (e.g., Carlsson et al. 2013, Jacquemet, Joule, Luchini, and Shogren 2013) to assess whether they may influence the respondents' answers when choosing between different policies in DCE. We focus on this type of mechanism as it is simple to implement and, if effective, can be of particular interest for practitioners to obtain reliable results. Our hypotheses in this paper are that priming, depending on the type of priming, can either address the issue of hypothetical bias in SP studies, and hence provide a more conservative (lower) WTP for the good of interest in the survey, or push respondents to care more for the good of interest, and hence, result in a higher WTP.

The priming treatments are described in more detail later in the paper but we focus on two types: (i) an oath presented at the beginning of the survey, and (ii) a priming scenario that combines information and questions on the respondents' experiences. Regarding the potential difference in how the priming may influence among different populations we implement the same survey in the United States (US) and the United Kingdom (UK). We focus on these two countries for two reasons: (i) they share many common features, including the language, which mitigate the risk of too many confounding factors influencing the results, and (ii) the US are recognized as having a "distinctive culture" (Sunstein & Reisch, 2019) that may suggest that findings from that country may not be generalized even to a country that may be perceived as "similar" as the UK. Indeed, except in Carlsson et al. (2013) where the authors compare the effect of an oath in a survey in China and Sweden, evidence is lacking regarding the effect of such priming mechanisms to induce reliable results in SP studies.

The contributions of our study are: (1) since priming may influence respondents' answers in SP studies, and hence estimated values to be used for policy purposes, it is of importance to examine whether priming does have an effect, and how robust these effects are between contexts (e.g. countries); and (2) in addition to what have already been tested in the literature, we extend the analyses by comparing what has already been used together with different forms of priming. Hence, the aim of the study is to be of both policy and research relevance by contributing to the literature on the behavioral aspects related to the design of SP studies that can have an effect on elicited preference estimates, such as the VSL, through the comparison of the effects of different forms of priming.

In the following section we briefly describe the concept of valuing mortality risk reductions and behavioral aspects of relevance when eliciting preferences in SP studies. After that, in Section 3, we describe the survey, i.e., how data were collected and the structure and content of it. Section 4 contains the model and empirical methods used followed by the results section. The paper ends with a discussion and some concluding comments.

2. Valuing safety and behavioral influences

In this section we first briefly describe the theory behind the VSL, the monetary value of interest to this study. We then discuss different behavioral aspects of relevance to SP studies that may influence respondents when they make their choices, and hence may affect the estimated monetary values.

2.1. Valuing reduced mortality risk

In both the RP and SP approaches described earlier, individuals' preferences are monetized by eliciting their WTP (or willingness to accept, WTA). Let p, w, a, and d denote the probability of survival, the wealth level, and the states of staying alive or being dead, and $u_l(w)$, $l \in \{a, d\}$, the state-dependent utility of wealth, then the WTP approach to monetizing health risk can be framed using the state-dependent expected utility model (Jones-Lee, 1974),

$$EU(w, p) = pu_a(w) + (1 - p)u_d(w). \tag{1}$$

Total differentiation of Eq. (1), while holding expected utility constant, results in the monetary amount an individual is willing to forgo to reduce the risk level,

$$\text{VSL} = \left.\frac{dw}{dp}\right|_{EU \text{ constant}} = \frac{u_a(w) - u_d(w)}{pu'_a(w) + (1 - p)u'_d(w)}, \tag{2}$$

where the prime denotes the first derivative. Hence, Eq. (2) is the marginal rate of substitution (MRS) between mortality risk and wealth, which is the VSL.

The standard assumptions are that the utility and marginal utility of wealth is higher when alive than when dead, marginal utility is non-negative, and individuals are weakly risk averse, i.e.,

$$u_a > u_d, u'_a > u'_d \geq 0, \text{ and } u''_l \leq 0, \tag{3}$$

which are sufficient for the VSL to be positive and increasing with wealth and baseline risk, and further it can be shown that WTP should be nearly proportional to small changes in risk (Corso, Hammitt, & Graham, 2001).[1]

2.2. Stated preferences studies and behavioral considerations

There are several aspects that may influence respondents' answers in SP surveys. Below we provide a brief review of behavioral aspects that have been shown to be influential and are relevant to the priming we propose. As emphasized in the introduction, we do not intend to propose formal tests to assess whether our priming addresses specific behavioral aspects, our objective is to assess whether and how priming can influence the choices made in SP studies, which in our case also means the effect on the estimated VSL.

Since respondents in SP studies face hypothetical choice situations they do not experience any consequences from their decisions. One concern is, therefore, that they exaggerate their WTP compared to how much they would pay if it would have been an actual decision. This is usually referred to as hypothetical bias (Carson & Groves, 2007; Murphy, Allen, Stevens, & Weatherhead, 2005) and, from a behavioral point of view, hypothetical bias may be related to salience since, compared to real situations, economic incentives are not at stake in SP surveys (Fifer, Rose, & Greaves, 2014). Whereas the use of cheap talk to address hypothetical bias has led to mixed results (Aadland & Caplan, 2006; Carlsson, Frykblom, & Lagerkvist, 2005), there is some

[1] The MRS between mortality risk and wealth is often used to define the VSL. A stricter definition is that the VSL is the population mean of the MRS when the individual MRS and the personal risk change are uncorrelated (Jones-Lee, 2003). This section has provided a very brief introduction to the topic. For a fuller description see, e.g., Andersson, Hole, and Svensson (2019).

evidence that preference-certainty questions can mitigate or eliminate this bias (e.g., Blumenschein, Blomquist, Johannesson, Horn, & Freeman, 2008). Related to the hypothetical nature is also the issue of perceived non-consequentionality, i.e., respondents do not believe that their answers have any consequences and are therefore not important. Compared to the hypothetical bias this could also lead to lower stated WTP compared to actual decisions and solutions to enhance consequentiality have been discussed in the literature (Carson & Groves, 2007; Vossler, Doyon, & Rondeau, 2012; Vossler & Evans, 2009).

Another, but related, concern in SP studies is whether respondents are attentive enough during the survey so that their answers are accurate and their preferences are truly revealed (Sandorf, 2019). Thus, there is a risk that low-quality data are collected if respondents do not pay enough attention. Malone and Lusk (2018) suggest that inattention in SP studies may result in: (i) the use of heuristics (e.g., always choosing the same option in a DCE), or (ii) a more random response behavior.

To tackle these concerns, one may rely on behavioral interventions, such as the use of oaths. This approach is grounded in the theory of commitment in psychology according to which agents want to maintain an internal consistency between their intentions and their actions if they commit to a given behavior (see, e.g., Joule & Beauvois, 2010; Joule, Bernard, & Halimi-Falkowicz, 2008). Concerning SP studies, committing to an oath makes agents answer the survey more truthfully, since if they deviate from the committed behavior (answering truthfully) they incur a moral cost and they also suffer from cognitive dissonance (Festinger, 1957). Hence, not stating the truth after having committed to it would impose a psychological cost on them, which means they have incentives to answer truthfully.

Oath mechanisms have already been tested in SP studies (e.g., Carlsson et al. 2013, Donfouet, Mahieu, and Macha 2013, Jacquemet et al. 2013, Kemper, Popp, and Nayga 2020). For instance, Carlsson et al. (2013), De-Magistris, Gracia, and Nayga (2013) and Kemper et al. (2020) find that the use of an oath significantly reduces the (marginal) WTP, thus potentially mitigating any hypothetical bias. However, in a more recent study, Mamkhezri, Thacher, Chermak, and Berrens (2020) do not find evidence of any effect of the solemn oath to mitigate the hypothetical bias for solar energy.[2] This could indicate that the solemn oath may not be enough. Therefore, in addition to asking respondents in one treatment to only ticking a box, we also ask respondents in another treatment to write the oath. Evidence suggests that when the respondents are asked to write or sign a commitment instead of simply ticking a box, their commitment is likely to be stronger since their self-image is perceived to be at higher stake (e.g., Cialdini 2009, Lokhorst, Werner, Staats, van Dijk, and Gale 2013).

The use of an oath has the potential to address both the issue of the hypothetical nature of SP studies, i.e. mitigating hypothetical bias, and the issue of inattention. Other possibilities exist to foster respondents' attention, though, such as asking individuals questions to recollect their past experiences, which are generally associated with some emotions. This type of behavioral intervention has been developed in social psychology to trigger a behavior in a given direction using individuals' mental representations (e.g., Bargh, Gollwitzer, Lee-Chai, Barndollar, & Trötschel, 2001; Molden, 2014). The intuition is that individuals are more likely to pay attention to a task if they link it with past experiences that involved some emotions. From a theoretical point of view, there is now evidence that emotions affect individual decisions (see Lerner, Li, Valdesolo, and Kassam 2015 for a review). In their experiment, Lerner, Gonzalez, Small, and Fischhoff (2003) emphasize that subjects who are asked to report an anxious/fearful event are more pessimistic regarding their judgments on different risks (related to future events) compared to those who are asked to report sad events. More recently, Callen, Isaqzadeh, Long, and Sprenger (2014) have shown that priming individuals to recall fear make them more often choose the certain option in lotteries.

3. Survey

The survey was programmed in LimeSurvey and implemented using MTurk in November and December of 2019 in the US and UK. In the past, there were some concern regarding the use of Mturk for economic experiments. However, several studies have replicated findings from lab experiments on MTurk (e.g., Arechar, Gächter, & Molleman, 2018; Coppock, Leeper, & Mullinix, 2018; Gandullia, Lezzi, & Parciasepe, 2020) and, nowadays, this platform is regularly used to conduct online experiments (e.g., Gandullia, 2019; Horton, Rand, & Zeckhauser, 2011; Jacquemet et al., 2021). One concern has been that respondents on MTurk might not pay sufficient attention to questions or instructions, but recent evidence has found that MTurkers pay more attention to these elements compared to students in the lab (Hauser & Schwarz, 2016).[3] Prior to the main survey a pilot with 100 respondents was run on MTurk that resulted in some minor changes.

3.1. Survey structure

Our survey consists of four parts. In the introduction of the survey, respondents are informed that their participation is important and that their answers can provide important knowledge for policy makers, this to address the importance of perceived consequentiality in SP studies (Mariel et al., 2021; Vossler et al., 2012). In the first part of the survey, respondents are also asked questions regarding environmental pollution and government spending on public services, to get them engaged in the questionnaire and to mitigate the risk of a framing bias towards air pollution, and it includes the priming for the treated respondents. In the second part, we include a training session on risk-money tradeoffs with the aim to prepare the respondents for the choice situations where they are going to choose between different air pollution policies (see Appendix A). The training section also allows us to identify inconsistent answers.[4] The third part consists of the DCE, and in the fourth and final part the respondents are asked some debriefing questions regarding the DCE as well as some on background information (age, gender, etc.). In the following sections, we will describe the priming treatments and the DCE part in more detail.

3.2. Priming treatments

To assess whether priming may influence respondents in a DCE our sample was split into four different subsamples, with one subsample not exposed to any priming, and hence is our control group. The four groups (control and the three treatments) are summarized in Table 1, and our respondents were randomly allocated to one treatment only (they could not be involved in other treatments).

[2] The solemn oath has also been tested in lab and online experiments (e.g., Jacquemet, James, Luchini, Murphy, & Shogren, 2021; Jacquemet, Luchini, Rosaz, & Shogren, 2019) with mixed results. While Jacquemet et al. (2019) find a reduction by 50% of non-payoff maximizing choices when a truth-telling oath is implemented compared to the control group, in Jacquemet et al. (2021) the only significant effect detected is on the time to complete the task (compared to the control group, those in the oath treatment spend more time on the task).

[3] To mitigate the risk of "low performers", i.e. respondents who do not pay attention during the survey, only MTurk workers with an approval ratio of 75 percent or greater were allowed to participate in the survey (Hunt & Scheetz, 2019).

[4] The training session was designed using a CV format to allow for an estimation of the respondents' WTP for a mortality risk reduction if the results suggested that answers provided valid preference estimates. Whereas the bid level had the predicted negative effect on the respondents' acceptance of paying for the policy, we could not reject that the size of the risk reduction had no effect. We, therefore, only use these answers for the training purpose and to identify inconsistent answers.

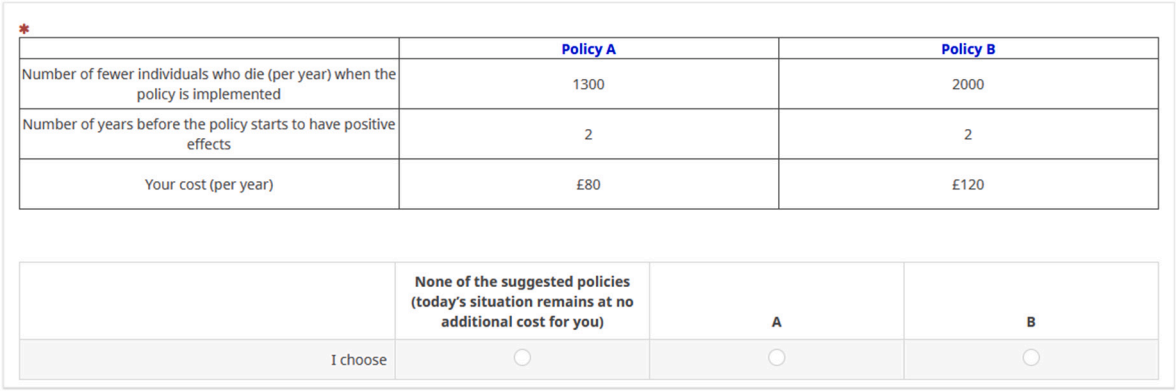


Fig. 1. Example of a choice set (UK version).

Table 1
Priming treatments.

Treatment	Description
Control	No manipulation
Priming scenario	Information provision and questions on respondents' experience
Oath	Respondents have to tick a box if they agree with the oath
Oath script	Respondents have to write the oath if they agree with it

We implemented two types of priming. The first type was split into two different versions and relies on respondents' commitment to tell the truth during the survey by agreeing to an oath (Carlsson et al., 2013; Donfouet et al., 2013; Jacquemet et al., 2013; Kemper et al., 2020). We implement two types of oaths: (i) one where the respondents only have to tick a box if they agree to tell the truth, therefore directly replicating Carlsson et al. (2013), De-Magistris and Pascucci (2014) and Jacquemet et al. (2013) (*Oath* treatment), and (ii) one in which the respondents after having agreed with the oath also have to confirm it by writing it (*Oath script* treatment), following the recommendations by Cialdini (2009) and Lokhorst et al. (2013). The different versions are shown in Appendix B.1.

The second type of priming involves providing information about the social cost from air pollution and asking about the respondents' opinion and personal experience related to air pollution and the environment (*Priming* scenario). The combination of the information and the questions may again give incentives to respondents to more carefully consider the information they receive and the answers they give (Bargh et al., 2001). More precisely, before the DCE respondents were presented with a cost estimate from the OECD related to air pollution and then they were asked a series of three questions related to: (i) their perception of their personal involvement to protect the environment, (ii) whether they, or someone close to them, have suffered from health issues due to air pollution, and (iii) their commitment to protect the environment. The presentation of the cost and the questions are shown in Appendix B.2. Regarding the cost estimate, one concern is that it could lead to anchoring (see, e.g., Vossler & Zawojska, 2020). We argue that this is a minor concern here; the cost estimate from the OECD describes healthcare costs, while in our DCE we ask about the respondents' WTP to reduce their mortality risk. Regarding the three questions, the first one aims at pushing the respondents to think about the actions they have been conducting so far to protect the environment and to assess whether or not they do enough (*Think* question). We, therefore, here play on respondents' self-image. The second question aims at capturing whether or not the respondents have been, or know someone who has been, suffering from air pollution (*Suffer* question). We here play on respondents' affect which we expect, following our discussion on attention and self-image in Section 2.2, make respondents caring more about the mortality attribute. With the last question, we assess whether respondents would be willing to change their behavior in the future to undertake actions to protect the environment (*Action* question).

3.3. Discrete choice experiment

The choice situations in the DCE contained three attributes and were inspired by previous studies on eliciting preferences to reduce mortality risk related to air pollution (e.g. Andersson et al., 2016; Jin et al., 2020).[5] Before being presented with the first choice set, and following the training part in which the general setup was presented, the respondents were shown the following information,

> Assume now that the government is considering different policies. In addition to the reduced mortality risk and the cost of the program as described before the policies also differ in when they will have an effect. That is, some policies will have an immediate effect on the reduced mortality risk, whereas for other policies the effect will be delayed by between 2 and 5 years, and then last for 5 years. As before, the payment starts today and lasts for 5 years for all policies.
>
> You will be asked to make 6 choices. In each choice situation you choose between two alternatives and you also have the option to choose neither of them. Treat each choice as a new situation, that is you should treat each situation as independent from the others.
>
> We also want to remind you that a higher cost means that you have less money left to consume other goods and services.

Hence, respondents were, in addition to being informed about the attributes and how to choose among the choice sets, reminded about the budget constraint (Johnston et al., 2017). As described, the attributes and their levels were inspired by previous studies on the same topic and to allow for a sufficiently large range of potential VSL estimates. Table 2 details the attributes and their levels. Regarding the mortality attribute, since individuals have difficulties understanding very small probabilities we decided to use frequencies instead of probabilities to describe the risk change (Andersson et al., 2016).

Fig. 1 provides an example of a choice set from the UK version. As shown, respondents choose between two alternatives (that differ

[5] The VSL concept in Section 2.1 is based on the individual tradeoff between wealth and change in risk of death, but following other studies eliciting preferences for public safety/health measures we framed air pollution as a public risk, and not an individual risk (e.g., Adamowicz et al., 2011; Jin et al., 2020). An individual risk scenario would avoid potential issues related to altruistic motives, which may influence the VSL, but since air pollution is a public bad we argue that a public safety scenario is more realistic and most likely to obtain valid preference estimates. For a discussion, see, e.g., Andersson, Levivier, and Lindberg (2019) and Robinson and Hammitt (2011).

Table 2
Description of the attributes and their level (US and UK).

Attribute	Description	Levels (US)	Levels (UK)
Mortality	Number of fewer annual deaths[a]	0, 6000, 9000	0, 1300, 2000
Delay	Number of yeas before policy has an effect	0, 2, 5	0, 2, 5
Cost	Annual cost	$0, $50, $100, $150, $250	£0 £40, £80, £120, £200

[a]Basee (annual number of deaths), US = 107,000 and UK = 33,000.

regarding the levels of the attributes) and one status quo (SQ). The choice sets were generated with the Ngene software using a D-efficient design based on priors from Jin et al. (2020). Twelve choice sets were created that were blocked into two different groups, meaning that each respondent answered six choice sets. Respondents were randomly assigned to one block and the order of the choice sets were randomized within each block.

4. Empirical methods and hypotheses

4.1. Baseline model

The respondents are assumed to derive utility from the attributes of the choice situations and their choices are analyzed using the Random Utility Model (RUM) (McFadden, 1974). According to the RUM, respondents maximize their utility, but the true one being unknown, a random component is considered. As previously explained, respondents were asked to choose their preferred option among $J = 3$ alternatives (two policy options and one status quo) in $T = 6$ choice sets. Individual i's utility for alternative j in choice set t therefore writes,

$$U_{ijt} = \beta_0 sq + \beta_1 mortality_{ijt} + \beta_2 cost_{ijt} + \beta_3 delay_{ijt} + \epsilon_{ijt}, \tag{4}$$

where sq is an alternative-specific constant (ASC) for the status quo, $\beta_0, \dots, \beta_3$ are coefficients to be estimated, and ϵ the random term, which is assumed to be I.I.D. extreme value type 1 distributed.

Using Eq. (4), we can obtain the marginal WTP for a change in mortality risk, i.e., VSL (dropping the subscripts of the utility function),

$$\text{VSL} = -\frac{\partial U/\partial mortality}{\partial U/\partial cost} = -\frac{\beta_1}{\beta_2}, \tag{5}$$

which, as explained in Section 2.1, is the measure of the MRS between mortality risk and wealth, i.e., the increase in cost necessary to keep the individual at the same utility level when the mortality risk is reduced due to the policy.

We extend Eq. (4) by interacting our attributes of interest for the estimation of the VSL, i.e., *mortality* and *cost*, with the treatment variables, which allows us to examine if the VSL varies between the treatments:

$$\begin{aligned} U_{ijt} &= \beta_0 sq + \beta_1 mortality_{ijt} + \beta_2 cost_{ijt} + \beta_3 delay_{ijt} \\ &\quad + \sum_{n=1}^{3} \left(\beta_{1,n} mortality_{ijt} + \beta_{2,n} cost_{ijt}\right) priming_n + \epsilon_{ijt}, \end{aligned} \tag{6}$$

where the subscript $n \in \{1, 2, 3\}$ refers to the different treatments, besides the control group, and *priming* is a dummy equal to 1 if the priming n (oath, oath script or priming scenario) is implemented, and 0 otherwise. The expression for the VSL is derived in the same way as in Eq. (5),

$$\text{VSL}_n = -\frac{\partial U/\partial mortality}{\partial U/\partial cost} = -\frac{\beta_1 + \beta_{1,n}}{\beta_2 + \beta_{2,n}}. \tag{7}$$

Hence, the coefficients $\beta_{1,n}$ and $\beta_{2,n}$ reflect the influence of the priming treatment on how the respondents react to the *mortality* and *cost* attributes, compared to the control group.

Based on the overall evidence in the literature from using an oath in SP studies (e.g., Carlsson et al., 2013; De-Magistris et al., 2013; Jacquemet et al., 2013), and the discussion on commitment and attention in Section 2.2, we expect that hypothetical bias is mitigated in the two oath treatments. Since respondents have committed to telling the truth and may be influenced to pay more attention to the cost attribute and the fact that they are expected to pay the amount shown to them, our first hypothesis is:

H1: Compared with the control, respondents treated with the oath treatments more carefully consider the cost attribute, i.e. $\beta_{2,oath}$ and $\beta_{2,oath\ script}$ are expected to be negative.

That is, hypothesis **H1** predicts, by mitigating the hypothetical bias, a more conservative (lower) VSL. However, the effect of the oath treatments on the VSL could be ambiguous if respondents also more carefully consider the mortality attribute when treated with one of the oaths, resulting in this attribute having a larger impact on respondents' choices. Such a result is in line with the recent findings in Mamkhezri et al. (2020). Our expectation, though, is that the oath will mainly address hypothetical bias, and hence, that hypothesis **H1** is met.

Moreover, following our discussion on self-image in Section 2.2, we expect a stronger effect in the *Oath script* treatment compared to the *Oath* treatment (only ticking a box), which is described in our second hypothesis:

H2: Compared to the oath, the effect of the oath script is stronger.

This effect, based on **H1**, would suggest that $\text{VSL}_{oath\ script} < \text{VSL}_{oath}$. However, if the *Oath script* also has an effect on how respondents react to the mortality attribute, the effect on the VSL would be ambiguous.

In the *Priming scenario* treatment we first hypothesize that respondents are influenced to be willing to pay more for reducing the mortality risk because of the sentiments that the questions may trigger. Second, since respondents are reminded of the existence of health issues and the social cost to society from air pollution, they may be more willing to accept a policy that would mitigate them, therefore paying more attention to the mortality attribute, but less attention to the cost attribute. Hence our two hypotheses related to priming are:

H3a: Compared with the control, respondents treated with the priming treatment more carefully consider the mortality attribute, i.e., $\beta_{1,priming}$ is positive.

H3b: Compared with the control, respondents treated with the priming treatment more carefully consider the mortality attribute but not the cost attribute, i.e. the expectation is that both $\beta_{1,priming}$ and $\beta_{2,priming}$ are positive.

Both hypotheses therefore predict a higher VSL, since we expect an increase in the effect from the mortality attribute, but no (**H3a**) or a smaller effect (**H3b**) from the cost attribute. This can be compared with the two oath treatments where hypotheses **H1** and **H2** predict a negative (or ambiguous) effect on the VSL.

Regarding the three questions in the *Priming scenario* treatment, we may observe different effects from this treatment depending on the answers to the questions themselves. Therefore, in addition to the estimation of treatment effects, we also compare the respondents' choices depending on their answers to the questions in the priming treatment. We first compare the effect of answering "yes" to the *Suffer* question compared with those having answered "no". Second, we construct the variable *Change*, a dummy equal to 1 if the respondent believes that

he/she does not do enough to protect the environment (*Think* question) but is willing to undertake actions now to protect future generations (*Action* question), and assess the impact of this variable on the choices made. We may expect that respondents who believe they do not do enough but are willing to take action care more about the mortality attribute, where the priming may influence respondents to perceive the mortality attribute as not only reflecting a health risk reduction but also an element of environmental protection. Our last hypotheses are therefore:

H4a: Respondents who suffer from air pollution, oneself or someone close, more carefully consider the mortality attribute, i.e., the coefficient for the interaction between *Mortality* and *Suffer* is positive.

H4b: Respondents who do not believe they do enough for the environment, but are willing to take action, more carefully consider the mortality attribute, i.e., the coefficient for the interaction between *Mortality* and *Change* is positive.

Hence, both **H4a** and **H4b** predict a higher VSL due to the expectations that the priming will mainly influence respondents to care more about the mortality attribute.

4.2. Preference heterogeneity

4.2.1. Generalized multinomial logit model

In the baseline model the estimated parameters of the model, i.e. the βs, are assumed identical between respondents and that any preference heterogeneity is examined by interacting either treatments or individual characteristics with the attributes. Hence, in Eq. (4) preferences are assumed homogeneous for the policy attributes. However, previous studies have emphasized that taking into account the panel structure of the data and also allowing for unobserved preference heterogeneity is more representative of the reality (McFadden & Train, 2000). We, therefore, extend our analysis to also take into account the panel structure and allow for unobserved preference heterogeneity by employing the generalized multinomial logit model (GMNL) (Fiebig, Keane, Louviere, & Wasi, 2010).

As above, N individuals choose among $J = 3$ alternatives in $T = 6$ choice sets. The probability that individual i chooses alternative j in choice set t is, following Fiebig et al. (2010),

$$Pr(choice_{it} = j|\beta_i) = \frac{exp(\beta_i' x_{ijt})}{\sum_{k=1}^{J} exp(\beta_i' x_{ikt})} \tag{8}$$

with x_{ikt} the vector of observed attributes of alternative j and β_i an individual specific vector of coefficients defined as following

$$\beta_i = \sigma_i \beta + [\gamma + \sigma_i(1-\gamma)]\theta_i. \tag{9}$$

In our application we set $\gamma = 0$, and hence estimate what has been referred to as the "scaled mixed logit model" (Greene & Hensher, 2010), i.e.,

$$\beta_i = \sigma_i(\beta + \theta_i), \tag{10}$$

where θ captures preference heterogeneity. As described by Hess and Train (2017), σ will not only capture scale heterogeneity but also correlations between utility parameters. We include σ to capture heterogeneity from pooling data from four different treatments. As explained, σ will reflect both scale heterogeneity and other behavioral correlations, but for simplicity we refer to is as scale from now on. To examine if scale differs between treatments we estimate the GMNL in Stata using the command by Gu, Hole, and Knox (2013) who, based on Fiebig et al. (2010), assume that σ_i is lognormally distributed with standard deviation τ and we include our treatments as explanatory variables for σ_i. We also run regressions solely on the *Priming scenario* treatment where we set $\sigma = 1$, and hence, Eq. (10) reduces to the mixed logit model, which is also estimated in Stata (Hole, 2007).

4.2.2. Latent class models

We finally analyze heterogeneity by estimating latent class models (Hole, 2008). Compared to the GMNL model above where we assume a distribution for θ, we here assume that preferences may vary between groups, but that within groups (classes) preferences are homogeneous, and hence represented by the same parameter estimate. The utility function is represented by,

$$U_{ijt} = \beta_{c0} sq_{ijt} + \beta_{c1} mortality_{ijt} + \beta_{c2} cost_{ijt} + \beta_{c3} delay_{ijt} + \epsilon_{ijt}, \tag{11}$$

with $c = 1, \ldots, C$ the individual respondent class membership. Then, conditional on membership in class c, the probability that respondent i chooses alternative j in choice set t is also given by Eq. (8). In our case, we use the treatments to determine class membership. The number of classes C being, *a priori*, unknown, we rely on goodness-of-fit measures such as the Akaike criteria to determine it. We detail this point in the next section.

5. Results

5.1. Data

In total, our sample consists of 806 US respondents and 694 UK respondents, corresponding to 1500 respondents.[6] As described in Appendix A, we were able to identify respondents who gave inconsistent answers during the training session which we decided to use to remove those respondents who did from our analysis. A total of 170 respondents were removed (94 in the US sample and 76 in the UK one).[7] We therefore consider an effective sample of 1330 respondents. Table 3 shows descriptive statistics, per treatment, for our two samples. Treatment sample sizes differ since the random allocation of respondents to different treatments was made through a random variable, and therefore, it was not possible to insure a perfect balance between treatments.

Conducting Kruskal Wallis tests for each sample to test for significant differences regarding the socio-economics characteristics (age, gender, education level and income), we find that the different treatments in the US significantly differ in terms of income (at the 5% level), with a lower mean income in the *Oath script* treatment. Regarding the UK sample, we do not detect any significant differences between the treatments. However, the two samples appear to be different, with younger respondents in the UK, more females and more respondents having a university diploma in the US.

Finally, and as emphasized in Section 2.2, one way to assess respondents' attention is to investigate the time spent to complete the DCE. Focusing only on the time spent to complete the six choice sets, we find that the time spent is similar between treatments and between the UK and US sample, and we do not detect any statistically significant difference.

5.2. Regression analysis

We now turn to the econometric analysis of the respondents' decisions to test our different hypotheses. Our preliminary analyses suggested both preference and scale heterogeneity from pooling the different treatments and our preferred model is therefore the GMNL as described in Eq. (10). In the following section we therefore report our results from the GMNL where we start by assessing the effects of the treatments, followed by examining the channels for both types of priming (oaths and priming scenario). We then report the results from the latent class models, which are our preferred models to estimate VSL for policy purposes.

[6] The objective was to also have 800 UK respondents, but due to the smaller number of potential MTurkers in the UK, we ended with a smaller sample for that country when the survey closed after 4 weeks.

[7] Overall, the number of inconsistent respondents is similar across treatments, with a slightly higher proportion in the *Oath* treatment in the US.

Table 3
Descriptive statistics.

	US				KW	UK				KW
	Control	Priming scenario	Oath	Oath script	test	Control	Priming scenario	Oath	Oath script	test
Age	40.18	38.66	39.79	37.46	0.310	29.63	27.86	30.46	29.25	0.191
	(12.33)	(12.26)	(13.01)	(10.56)		(9.370)	(8.076)	(10.07)	(9.669)	
Gender	0.374	0.426	0.506	0.426	0.105	0.307	0.313	0.322	0.336	0.954
	(0.485)	(0.496)	(0.501)	(0.496)		(0.463)	(0.465)	(0.469)	(0.474)	
University	0.719	0.713	0.676	0.645	0.390	0.593	0.551	0.636	0.591	0.499
	(0.451)	(0.454)	(0.469)	(0.480)		(0.493)	(0.499)	(0.483)	(0.493)	
Income[a]	63.10	60.22	59.64	52.94	0.017	34.84	34.56	32.42	33.95	0.693
	(31.09)	(30.69)	(31.23)	(31.41)		(18.57)	(17.45)	(18.29)	(17.96)	
Observations	171	188	170	183		150	176	143	149	

[a]Income in USD for the US, and in GBP for the UK.

5.2.1. Treatment effects

Table 4 reports the results from the GMNL for the US and UK separately.[8] For both samples, we estimate two models. First, in models (1) and (3), we only consider the attributes and include the treatments as explanatory variables for scale heterogeneity. Then, in models (2) and (4), we interact the cost and mortality attributes with the dummies for the treatments (to test **H1**, **H2**, **H3a** and **H3b**), which also means that we no longer include the treatments to explain scale heterogeneity. We do not interact the delay attribute since our main focus is on the preferences for reducing the mortality risk, and as our intuition is that the priming treatments may in particular influence the respondents' perception of the cost and/or of the mortality attribute.

Regarding the main effects, the results are in line with our expectations: respondents prefer programs that cost less, provide a larger mortality risk reduction, and provide it sooner. Moreover, the coefficient for the ASC for the SQ is negative and significant, which suggests that respondents have a preference for the policy to be implemented rather than the contrary. Turning to the interactions, some differences appear between the two countries. First, whereas both types of oaths have a significant effect when interacted with the cost attribute in the UK sample, only the *Oath* treatment is statistically significant when interacted with the cost attribute in the US sample. Moreover, the direction of the effect is different in the two countries. Whereas in the US the coefficient is positive, they are negative in the UK. Hence, whereas US respondents belonging to the *Oath* treatment are less sensitive to the cost attribute compared to the control group, the UK respondents are more sensitive. Second, the oath treatments also have an effect on respondents' reaction to the mortality attribute. While in the US those who received the oath treatments are less sensitive to higher mortality rates (negative and statistically significant coefficients), in the UK this results is only found for those who received the *Oath* treatment. Third, the *Priming scenario* treatment does not have any effect on respondents' sensitivity to the cost attribute in either country, as shown by the statistically insignificant interaction coefficients, but a negative and statistically significant effect from priming on mortality is found in the UK sample.

Regarding the effect from the treatments on the scale component, which is reported in the section *Het* of Table 4, we expected that the two oath treatments and the priming scenario treatment would have a lower variance compared with the control group, i.e., a positive scale parameter. While the scale parameter is positive and statistically significant for both oath treatments for the US, and thus a lower variance of respondents' choices compared to the control group in line with our expectations, we find no evidence that the treatments have any effect on the scale in the UK sample. As discussed in Section 4.2.1, σ also captures other behavioral correlations besides scale which may explain the results.

Table 4
GMNL: Main and treatments effects.

	US		UK	
	(1)	(2)	(3)	(4)
Mean				
ASC_SQ	−6.737***	−7.913***	−9.082***	−8.530***
	(0.665)	(0.799)	(0.857)	(0.776)
Cost	−0.00840***	−0.0202***	−0.0305***	−0.0231***
	(0.001)	(0.003)	(0.004)	(0.003)
Mortality	0.131***	0.293***	0.0549***	0.144***
	(0.023)	(0.036)	(0.019)	(0.021)
Delay	−0.470***	−0.951***	−0.537***	−0.583***
	(0.079)	(0.108)	(0.089)	(0.061)
Cost × Oath		0.00535**		−0.00881**
		(0.003)		(0.004)
Cost × Oath script		0.00295		−0.0144***
		(0.003)		(0.005)
Cost × Priming scenario		0.000565		−0.00189
		(0.003)		(0.004)
Mortality × Oath		−0.122***		−0.0715***
		(0.022)		(0.026)
Mortality × Oath script		−0.0596***		0.00506
		(0.022)		(0.024)
Mortality × Priming scenario		0.0354		−0.0601***
		(0.022)		(0.023)
Het				
Oath	0.479**		0.0341	
	(0.191)		(0.174)	
Oath script	0.687***		0.145	
	(0.175)		(0.222)	
Priming scenario	0.0572		−0.0944	
	(0.200)		(0.146)	
SD				
ASC_SQ	5.479***	−6.351***	5.224***	−4.793***
	(0.536)	(0.662)	(0.729)	(0.554)
Mortality	−0.183***	0.358***	−0.0655***	0.151***
	(0.029)	(0.039)	(0.025)	(0.021)
Delay	0.463***	0.963***	0.517***	0.505***
	(0.078)	(0.107)	(0.093)	(0.062)
τ				
Constant	−1.012***	1.450***	1.498***	1.286***
	(0.103)	(0.087)	(0.080)	(0.074)
N	12816	12816	11124	11124
Log-likelihood	−3187.5	−3183.6	−2587.1	−2581.1

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

To summarize, it appears that the two types of oaths have different effects in the US and in the UK. Whereas respondents become more sensitive to the cost attribute in the UK when being treated with one of the two oath treatments, which is line with hypothesis **H1**, we find the opposite for the US (only statistically significant for the *Oath* treatment, though). Regarding the mortality attribute, we find the same qualitative result in both the US and UK samples for the *Oath*

[8] Following the recommendation by Gu et al. (2013) multiple starting values were tested for the GMNL and the run with the largest likelihood was used for each regression. The number of draws used for the simulations, GMNL and Mixed logit, were 1000.

Table 5
GMNL: Assessment of the effect of writing the oath.

	US		UK	
	(1)	(2)	(3)	(4)
Mean				
ASC_SQ	−7.008***	−7.207***	−8.403***	−8.823***
	(0.995)	(1.009)	(1.038)	(1.222)
Cost	−0.00761***	−0.00828***	−0.0460***	−0.0373***
	(0.001)	(0.001)	(0.010)	(0.006)
Mortality	0.119***	0.109***	0.0578***	0.0179
	(0.020)	(0.024)	(0.018)	(0.020)
Delay	−0.478***	−0.548***	−0.592***	−0.551***
	(0.070)	(0.068)	(0.148)	(0.100)
Cost × Oath script yes		−0.000905		−0.00652
		(0.001)		(0.007)
Mortality × Oath script yes		0.0454		0.0542**
		(0.030)		(0.027)
Het				
Oath script	0.224		−0.112	
	(0.160)		(0.228)	
SD				
ASC_SQ	−6.512***	−6.884***	3.424***	4.003***
	(0.936)	(1.036)	(0.696)	(0.977)
Mortality	0.164***	0.185***	0.0719***	−0.0414***
	(0.023)	(0.025)	(0.021)	(0.015)
Delay	0.442***	0.507***	−0.602***	−0.548***
	(0.068)	(0.065)	(0.165)	(0.119)
τ				
Constant	−0.376*	−0.468***	−1.748***	1.711***
	(0.228)	(0.172)	(0.143)	(0.133)
Agree with oath		98.82%		99.30%
Agree with oath script		93.99%		93.29%
N	6192	6192	5166	5166
Log-likelihood	−1475.0	−1475.0	−1182.5	−1181.4

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

treatment, but only an effect from the *Oath script* treatment in the US and *Priming scenario* treatment in the UK. Overall we, therefore, reject hypotheses **H3a** and **H3b**, since they both predicted a positive effect on the mortality attribute in the priming treatments. The VSL estimates do vary depending on the treatment, but they are either not statistically significantly different, or based on interaction coefficients that were not statistically significantly different. The exception is the VSL for the UK *Oath* treatment which is statistically significantly lower than the control group. Therefore, **H1** is only verified with the *Oath* treatment in the UK.

5.2.2. Channels for the priming treatments

To better understand the effects of the priming treatments, we start by examining the effect from writing the oath script compared to "just" ticking a box (to test **H2**). That is, we are interested in understanding the additional effect of writing the text compared to the action of ticking a box. Therefore, the analysis is restricted to the two oath treatments and only respondents agreeing to the oath. Results are reported in Table 5 where we, thus, compare the *Oath script* treatment to the *Oath* treatment. As above, we first estimate models without any interaction terms (models 1 and 3), and then we interact the cost and mortality attributes with the variable *Oath script yes*, a dummy equal to 1 if the respondent agreed to the oath and wrote it (models 2 and 4). Moreover, as above we test whether the treatments have any effect on the scale in the regressions without any interactions, but as shown in Table 5 we find no effect from belonging to the *Oath script* treatment on the scale.

We again find the expected sign and significance of the main attributes, except for the coefficient of the mortality attribute in model (4), which is no longer statistically significant. We also detect a difference between the US and the UK, as *Oath script* interacted with the mortality attribute is statistically significant for the UK sample, in line with hypothesis **H2**. We, therefore, reject **H2** for the US sample and only find weak evidence that asking respondents to write the text, compared to only ticking a box, has an effect, i.e., only in the UK.

Finally, we investigate whether respondents' answers in the *Priming scenario* treatment depend on the respondents' type by interacting the cost and mortality attributes with *Suffer* and *Change* (to test **H4a** and **H4b**). Hence, within the priming scenario treatment we compare those who answered that they have been suffering from air pollution to those who have not, and those who are willing to change to those who do not want to (see Section 3.2 for the description of these variables). The results are reported in Table 6, and similarly to our past observations, the US and UK samples do not react the same way to the same priming. Indeed, contrary to the UK respondents, the US respondents who have been suffering from pollution (or know someone who has been suffering) are more sensitive to the mortality attribute, as suggested by hypothesis **H4a**. However, in the UK sample, respondents who are willing to change are more sensitive to both the cost and mortality attribute, where the latter is in line with hypothesis **H4b**, but the former not in line with our expectations. In the US sample, we do not detect any such an effect from those who are willing to change. Regarding the treatment effect on the VSL, as above we find that levels vary with treatments/interactions, but that they are either not statistically significantly different, or based on interaction coefficients not being statistically significant.

5.3. Latent class analyses and VSL for policy purposes

As described, contrary to the previous models, latent class analysis (LCA) allows to capture unobserved preference heterogeneity through the identification of subgroups in the population. One issue is to establish the number of classes, and another one to consider is that the standard errors increase with the number of classes (Hole, 2008). The standard approach is to decide the number of classes based on information criteria like Bayes or Akaike. In our case the total number of classes we could consider was limited by the number of parameters we could estimate. This restriction combined with our tests of best model fit resulted in 3 classes for our LCA.

Table 7 shows the results for three classes in each country. The results suggest that there are two classes in the US sample for which our validity requirements of the mortality and cost attributes having, respectively, a statistically significant positive and negative effect on respondents' choices are not met, i.e., classes 2 and 3. For the UK sample, our validity requirements are not met for class 2. The results also show that class membership does not depend on the treatments.

Since the LCA allows us to detect groups of respondents who do not meet our validity requirements, we prefer to use the LCA for the estimation of our preferred VSL. Our preferred estimates are therefore based on the results in Table 7 and for the US sample the VSL is based on class 1 and is USD 7.81 million, which is lower but close to the USD 9.5 million recommended by the US EPA. The VSL for the UK sample is a weighted average based on classes 1 and 3 and equals GBP 3.11 million, which is higher but also close to the GBP 1.93 million proposed by the UK Dept. for Transportation for a VSL for the UK, which is the reference value also used for air pollution policies (UK HM Treasury, 2020).[9] Hence, both the US and UK values are in line with what we would expect, which could suggest them being reliable estimates of preferences to reduce the mortality risk related to air pollution in both countries.

[9] The US value obtained from https://www.epa.gov/environmental-economics/mortality-risk-valuation#whatisvsl (2021-01-30), whereas the UK from DFT, TAG Data Book, July 2020 v1.13.1. Original values have been converted to 2020 price level using CPI from stats.oecd.org.

Table 6
Mixed logit: Assessment of the effect of priming scenario questions.

	US			UK		
	(1)	(2)	(3)	(4)	(5)	(6)
Mean						
ASC_SQ	−5.429***	−5.176***	−5.184***	−8.235***	−8.201***	−7.965***
	(0.897)	(0.856)	(0.831)	(1.442)	(1.382)	(1.377)
Cost	−0.00654***	−0.00708***	−0.00615***	−0.0121***	−0.0122***	−0.00903***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Mortality	0.0923***	0.0724***	0.0815***	0.0770***	0.0736***	0.0327
	(0.018)	(0.020)	(0.023)	(0.018)	(0.021)	(0.027)
Delay	−0.290***	−0.291***	−0.300***	−0.361***	−0.363***	−0.364***
	(0.041)	(0.042)	(0.044)	(0.045)	(0.045)	(0.045)
Cost × Suffer		0.00162			0.000312	
		(0.001)			(0.002)	
Mortality × Suffer		0.0676**			0.0157	
		(0.034)			(0.039)	
Cost × Change			−0.00107			−0.00469**
			(0.001)			(0.002)
Mortality × Change			0.0275			0.0722**
			(0.033)			(0.033)
SD						
ASC_SQ	4.751***	4.350***	4.445***	6.799***	6.659***	6.570***
	(0.938)	(0.898)	(0.897)	(1.586)	(1.278)	(1.333)
Mortality	0.135***	0.130***	0.143***	0.129***	0.132***	0.120***
	(0.022)	(0.022)	(0.022)	(0.026)	(0.025)	(0.024)
Delay	−0.292***	−0.287***	0.308***	0.269***	0.272***	0.272***
	(0.049)	(0.049)	(0.052)	(0.052)	(0.053)	(0.051)
Prop. of Suffer = 1		29.79%			26.14%	
Prop. of Change = 1		48.94%			64.77%	
N	3384	3384	3384	3168	3168	3168
Log-likelihood	−865.9	−862.0	−865.5	−737.3	−737.2	−733.1

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7
Latent class models.

	US		UK	
Class 1				
ASC SQ	−2.850***	(0.194)	−4.426***	(0.277)
Mortality	0.0845***	(0.010)	0.0588***	(0.011)
Delay	−0.133***	(0.018)	−0.209***	(0.018)
Cost	−0.0108***	(0.001)	−0.0164***	(0.001)
Class 2				
ASC SQ	−3.049***	(0.611)	−1.836***	(0.427)
Mortality	0.108***	(0.020)	0.155***	(0.037)
Delay	−0.517***	(0.051)	−0.653***	(0.091)
Cost	0.00487***	(0.001)	0.00482**	(0.002)
Class 3				
ASC SQ	−2.126***	(0.476)	−1.520***	(0.368)
Mortality	0.0210	(0.014)	0.0426***	(0.013)
Delay	−0.102*	(0.057)	−0.235***	(0.053)
Cost	−0.0507***	(0.006)	−0.0406***	(0.005)
Share 1[a]	0.511		0.650	
Priming scenario	0.00445	(0.344)	−0.109	(0.342)
Oath	−0.379	(0.345)	−0.325	(0.348)
Oath script	−0.414	(0.335)	−0.0231	(0.355)
Constant	1.402***	(0.266)	1.573***	(0.265)
Share 2[a]	0.335		0.198	
Priming scenario	−0.0918	(0.367)	−0.340	(0.400)
Oath	0.0163	(0.359)	−0.642	(0.416)
Oath script	−0.225	(0.352)	−0.580	(0.436)
Constant	0.855***	(0.288)	0.644**	(0.300)
Estimated VSL[b]	7.84 (5.98–9.69)		3.11 (2.19–4.02)	
Observations	12816		11124	
Log-likelihood	−3082.3		−2542.8	

[a]Share 3 = (1 − (Share 1 + Share 2))
[b]VSL in millions USD and GBP with 95% conf. intervals in parentheses.
Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
VSL for the US based on Class 1 and for the UK a weighted average based Class 1 and 3.

6. Discussion

In this study, we use a DCE to elicit US and UK respondents' preferences for reducing mortality risks related to air pollution. In addition, we test whether the respondents' preferences, when choosing between different policies, can be influenced using two types of priming — two versions of an oath presented at the beginning of the survey, and a priming scenario that combines information and questions on the respondents' experiences.

We used self-recruited MTurkers to conduct the study. A concern with such a sample could be the quality of the answers. However, as described, evidence suggests that surveys conducted on MTurk can obtain just as valid and reliable findings as more traditional surveys or experimental approaches in economics. Nothing in our analysis suggests otherwise. The coefficient estimates of our attributes in the choice situations all have the expected signs and are statistically significant: respondents prefer programs that cost less, that provide larger risk reductions, and where benefits appear sooner rather than later. Moreover, the level of our VSL for both the US and the UK are in line with values recommended by responsible agencies in each country. Hence, nothing in our analyses suggests that we should reject that our findings provide valid and reliable estimates of preferences to reduce mortality risk due to air pollution in the US and the UK.

We have tested several hypotheses, summarized in Table 8. We find some evidence that preferences can be influenced. Regarding the oath treatments, our hypothesis was that respondents would be more involved in the survey and consider more carefully the choice sets and, in particular, the cost attribute. Moreover, we expected the *Oath script* treatment to trigger a stronger effect because of the cost associated with writing the oath. That is, writing the full oath instead of just ticking a box, make the respondents pay more attention to their self-image (Baca-Motes, Brown, Gneezy, Keenan, & Nelson, 2013). We find that both oath treatments do have an effect on respondents' answers, but that there are differences both within and between countries, and that even if the level of the VSL is affected, it is often not statistically

Table 8
Summary of the behavioral hypotheses and of the results.

	Country	
Hypotheses	US	UK
H1: Effect of oath treatments on the cost attribute	×	✓
H2: Stronger effect of the oath script compared to the (simple) oath	×	(✓)
H3a: Effect of the priming treatment on the mortality attribute	×	×
H3b: Effect of the priming treatment on both the mortality and cost attributes	×	×
H4a: Effect of suffering from air pollution on the mortality attribute	✓	×
H4b: Effect of the willingness to take action on the mortality attribute	×	✓

Note: ✓ = Evidence in favor, (✓) = Weak evidence, × = No evidence.

significantly different between treatments. The exception is the UK where the oath (i.e., no script) resulted in a statistically significantly lower VSL compared with the control group, due to the combined effect of higher sensitivity to the cost attribute and less to the mortality attribute.

Our findings only partially confirm those in Carlsson et al. (2013) and Donfouet et al. (2013), who find that a simple oath implemented before asking WTP questions mitigates the hypothetical bias. In our case, this is the case only for the UK respondents. In the US, the global effect of the simple oath is ambiguous (as the coefficients of the cost and mortality attributes change in opposite directions). Regarding the oath script, even if it induces a lower VSL (through a lower coefficient of the mortality attribute in the US, and a higher coefficient of the cost attribute in the UK), the difference with the control is not statistically significant. Our results are therefore more in line with those in Mamkhezri et al. (2020), who do not observe any effect of the solemn oath on the mitigation of the hypothetical bias. Still, we observe in the US a positive effect of both the oath and oath script on the scale parameter. This seems to indicate that respondents treated with these honesty priming measures provide less random answers in the DCE, i.e. their decision may be more carefully thought through. Although additional evidence would be necessary to confirm this point, this would justify the use of oath mechanisms to enhance data quality.

We, therefore, highlight cultural differences between the US and the UK regarding the respondents' reaction to the same priming. Carlsson et al. (2013) also found cultural differences between China and Sweden when implementing an oath, with the oath inducing a stronger effect in China. They argued that it could be explained by a difference in individualism/collectivism (Hofstede, 2011), given that individualism has been found to increase the hypothetical bias in SP studies (Ehmke, Lusk, & List, 2008). Similarly to their study, our data do not allow us to test the reason for the cultural difference, but although both countries share many common features the US are recognized as having a "distinctive culture" (Sunstein & Reisch, 2019), which may explain our result.

Regarding the *Priming scenario* treatment, this priming could influence the marginal WTP in both directions. The sentiments emerging from the questions and of being reminded of the social cost related to air pollution could influence the respondents to pay more for reducing the mortality risks. However, by being reminded of their own experiences and actions taken, respondents could also be willing to pay less for reducing the mortality risks if they feel that they already have done enough to protect the environment, and have been influenced to treat the health attribute as also capturing environmental effects. Our hypothesis was that if one of these effects would be stronger, it would be the former one, i.e., a higher marginal WTP. However, we do not detect any effects of the *Priming scenario* treatment on how respondents react to the cost attribute, but we do find a negative one for the mortality attribute in the UK sample. Moreover, we do show that there is an effect that depends on the type of respondents. More precisely, after having characterized two types of respondents (those who have been, or know someone who has been, suffering from air pollution and those who are willing to change), we show two different types of reactions between the two countries. In the US, the *Priming* treatment is more influential on those who have been suffering (or know someone who has) from air pollution (compared to those who have not been suffering), while in the UK it is more influential on those who are willing to change (compared to those who are not willing to change). One contributing factor could be that in the UK individuals are enjoying a better air quality than in the US, thus explaining the significant coefficient of the *Suffer* variable in the US only.[10] In a sense, this last result is in line with the one obtained in Callen et al. (2014) and Lerner et al. (2003): using a prime based on the recollection of past experiences (having suffered from air pollution in our case) does affect respondents' decisions.

7. Conclusions

Our results highlight a heterogeneity in respondents' reaction to different priming. In a sense, we can make a link with the literature that focuses on the assessment of the effects of nudges and, in particular, on the effect of social norm comparisons (Allcott, 2011; Brent, Cook, & Olsen, 2015; Ferraro & Price, 2013).[11] Similarly in SP studies, Ouvrard, Abildtrup and Stenger (2020) show that the effect of their nudges to increase respondents' WTP for an environmental policy depends both on their environmental sensitivity and their perception of the fairness of the measure. Moreover, our results based on choices in an SP study confirm those obtained by Reisch, Sunstein, and Gwozdz (2017) and Sunstein and Reisch (2019), who find that we cannot generalize the results obtained following the implementation of a nudge in one country to another one, because of cultural differences. Therefore, our results only partially confirm those obtained in Carlsson et al. (2013) and Jacquemet et al. (2013).

The fact that respondents' preferences can be influenced is the essence of using priming when people make actual decisions, but here our results show that they can also be used to influence respondents when making hypothetical decisions in SP studies, in line with the results obtained by Ouvrard, Préget, Reynaud and Tuffery (2020). Moreover, as shown, results regarding how priming influence decisions in an SP study obtained in one context cannot necessarily be generalized to other contexts. Hence, our results show that care needs to be taken to make sure that the chosen priming has the intended impact on the targeted population, like mitigating hypothetical bias, making respondents more attentive, or with the purpose of obtaining valid estimates of the respondents' preferences.

The main purpose of this study regarding the effect of different types of priming was to examine if they had an effect and whether they were country specific. We did find effects, but our design did not allow us to examine whether they are due to mitigated hypothetical bias, increased attention, higher perception of consequentiality, or some other driving behavioral aspects. Whether or not such mechanisms

[10] See: https://aqicn.org/rankings/ (2021-01-30).

[11] In all these studies, the authors highlight a high heterogeneity among individuals in their reaction to the nudge.

could be generalized when implementing DCE is an open question. Still, because oath mechanisms can significantly reduce random behaviors (as seen with the scale parameter for the US respondents) without causing, *a priori*, any "harm" on respondents, we argue in this direction.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

We are grateful for valuable feedback from anonymous reviewers and participants during seminars and at conferences. Henrik Andersson acknowledges funding from ANR, France under grant ANR-17-EURE-0010 (Investissements d'Avenir program).

Appendix A. Training session

Most individuals do not have any experience of trading money for mortality risk-reductions related to air pollution. The aim of the training session was therefore to familiarize the respondents with such a situation. The training session was designed as a single-bounded (SB) CV question with an open-ended WTP question as a follow-up.

Before being asked whether they were willing to pay a specified amount the risk scenario was described (US sample):

> Various air pollutants from power plants, industries, vehicles and residential sources enter the air and eventually turn into atmospheric pollutants such as fine particulate matter (PM2.5), ground level ozone and so forth. The harmful substances of these pollutants will enter the human body through breathing and cause various health problems.
>
> According to research PM2.5 is responsible for 107,000 annual premature deaths in the United States. The number of premature deaths can be reduced by, for instance, government action but it will come at a cost. We will therefore now ask you to make choices in situations where there is a tradeoff between the risk of dying and the cost.
>
> We ask you to assume that new stricter government policies on PM2.5 emissions will reduce the number of deaths. The policy will last for 5 years and the payment will be in form of an annual tax during 5 years that will be earmarked for the specific policy. Keep in mind that besides the reduction in the number of deaths and the costs, everything else stays the same. That is, there are no additional benefits or costs of the policies. There are no right or wrong answers, you decide according to your own opinion, and we want to remind you that a higher cost means that you have less money left to consume other goods and services.

The UK sample got the same scenario except from having the information about the number US deaths replaced by the correct number for the UK which was 33,000. The respondents were then asked to state whether they would agree to pay for a policy implemented to reduce the number of annual deaths from atmospheric pollutants such as fine particulate ($PM_{2.5}$) as a SB yes/no question,

> Assume a government policy to reduce PM2.5 that results in 3000 fewer individuals who die (per year). Would you support this policy if the annual cost to you was USD 50?

Table 9
Training part characteristics.

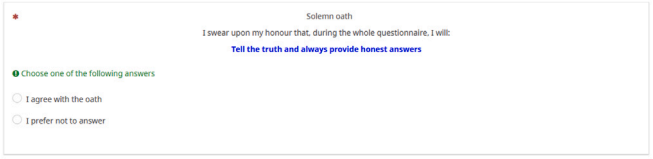
	US		UK	
Group	Number of fewer deaths	Cost	Number of fewer deaths	Cost
1	3000	$50	650	£40
2	3000	$250	650	£200
3	9000	$50	2000	£40
4	9000	$250	2000	£200

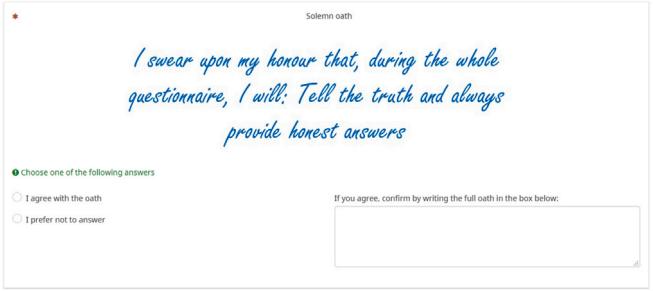
Two different risk reduction levels as well as two different bid levels were used for the US and UK samples, respectively. This means that there were four different versions in each sample and respondents were randomly assigned to one out of the four, with levels and combinations shown in Table 9.

As described, as a follow-up to the SB question above respondents were asked about their maximum WTP for the risk reduction. The main objective of this two step procedure was to make the respondents think twice about the money-risk tradeoffs and, hence, familiarize them with the scenario, but it also allows us to identify inconsistent answers. For instance, if a respondent answered yes to the question above on whether he/she was willing to pay USD 50 and then stated that his/her maximum WTP was USD 35 it could be used as an indication that the respondent either did not understand the choice situation, or did not pay attention. Respondents identified as inconsistent respondents were excluded when analyzing the DCE data.

Appendix B. Treatments

B.1. Oath and oath script treatments





B.2. Priming scenario

According to OECD, air pollution will cost around $330/person (£250/person) in 2060 in terms of annual healthcare costs.

[Think question] Do you think that you do enough to protect the environment? (Yes, No)

[Suffer question] Do you, or does anyone you know, suffer from air pollution? (Yes, No)

[Action question] Would you be willing to undertake actions now to protect the future generations? (Yes, No)

References

Aadland, D., & Caplan, A. J. (2006). Cheap talk reconsidered: New evidence from cvm. *Journal of Economic Behaviour and Organization*, *60*(4), 562–578.

Adamowicz, W., Dupont, D., Krupnick, A., & Zhang, J. (2011). Valuation of cancer and microbial disease risk reductions in municipal drinking water: An analysis of risk context using multiple valuation methods. *Journal of Environmental Economics and Management*, *61*(2), 213–226.

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, *95*(9–10), 1082–1095.

Andersson, H., Hole, A. R., & Svensson, M. (2016). Valuation of small and multiple health risks: A critical analysis of sp data applied to food and water safety. *Journal of Environmental Economics and Management*, *75*, 41–53.

Andersson, H., Hole, A. R., & Svensson, M. (2019). Valuation of health risks. In *Oxford research encyclopedia of economics and finance*. Oxford University Press.

Andersson, H., Levivier, E., & Lindberg, G. (2019). Private and public willingness to pay for safety: A validity test. *Accident Analysis and Prevention*, *123*(3), 170–175.

Arechar, A. A., Gächter, S., & Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics*, *21*(1), 99–131.

Baca-Motes, K., Brown, A., Gneezy, A., Keenan, E. A., & Nelson, L. D. (2013). Commitment and behavior change: Evidence from the field. *Journal of Consumer Research*, *39*(5), 1070–1084.

Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., & Trötschel, R. (2001). The automated will: Nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology*, *81*(6), 1014.

Beatty, T. K., & Shimshack, J. P (2014). Air pollution and children's respiratory health: A cohort analysis. *Journal of Environmental Economics and Management*, *67*(1), 39–57.

Blumenschein, K., Blomquist, G. C., Johannesson, M., Horn, N., & Freeman, P. (2008). Eliciting willingness to pay without bias: Evidence from a field experiment. *The Economic Journal*, *118*(525), 114–137.

Brent, D. A., Cook, J. H., & Olsen, S. (2015). Social comparisons, household water use, and participation in utility conservation programs: Evidence from three randomized trials. *Journal of the Association of Environmental and Resource Economists*, *2*(4), 597–627.

Burnett, R., Chen, H., Szyszkowicz, M., Fann, N., Hubbell, B., Pope, C. A., et al. (2018). Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proceedings of the National Academy of Sciences*, *115*(38), 9592–9597.

Callen, M., Isaqzadeh, M., Long, J. D., & Sprenger, C. (2014). Violence and risk preference: Experimental evidence from Afghanistan. *American Economic Review*, *104*(1), 123–148.

Cameron, T. A., DeShazo, J., & Stiffler, P. (2010). Demand for health risk reductions: A cross-national comparison between the us and canada. *Journal of risk and Uncertainty*, *41*(3), 245–273.

Carlsson, F., Frykblom, P., & Lagerkvist, C. J. (2005). Using cheap talk as a test of validity in choice experiments. *Economics Letters*, *89*(2), 147–152.

Carlsson, F., Kataria, M., Krupnick, A., Lampi, E., Å, Löfgren., Qin, P., et al. (2013). The truth, the whole truth, and nothing but the truth—a multiple country test of an oath script. *Journal of Economic Behaviour and Organization*, *89*, 105–121.

Carson, R. T., & Groves, T. (2007). Incentive and informational properties of preference questions. *Environmental and Resource Economics*, *37*(1), 181–210.

Champ, P. A., Boyle, K., & Brown, T. C. (Eds.), (2017). *A primer on nonmarket valuation* (2nd ed.). Dordrecht, The Netherlands: Springer Netherlands.

Chay, K. Y., & Greenstone, M (2005). Does air quality matter? Evidence from the housing market. *Journal of Political Economy*, *113*(2), 376–424.

Cialdini, R. B. (2009). *Influence: Science and practice, Vol. 4*. MA: Pearson education Boston.

Coppock, A., Leeper, T. J., & Mullinix, K. J. (2018). Generalizability of heterogeneous treatment effect estimates across samples. *Proceedings of the National Academy of Sciences*, *115*(49), 12441–12446.

Corso, P. S., Hammitt, J. K., & Graham, J. D. (2001). Valuing mortality-risk reduction: Using visual aids to improve the validity of contingent valuation. *Journal of Risk and Uncertainty*, *23*(2), 165–184.

De-Magistris, T., Gracia, A., & Nayga, R. M., Jr. (2013). On the use of honesty priming tasks to mitigate hypothetical bias in choice experiments. *American Journal of Agricultural Economics*, *95*(5), 1136–1154.

De-Magistris, T., & Pascucci, S. (2014). The effect of the solemn oath script in hypothetical choice experiment survey: A pilot study. *Economics Letters*, *123*(2), 252–255.

Global Burden of Disease Study (2017). GBD 2017 results. URL: https://vizhub.healthdata.org/gbd-compare/.

Donfouet, H., Mahieu, P., & Macha, R. (2013). A comparison of oath and certainty calibration in contingent valuation method: An application to community health fund. In *Document de travail du CREM*.

Ehmke, M. D., Lusk, J. L., & List, J. A. (2008). Is hypothetical bias a universal phenomenon? A multinational investigation. *Land economics*, *84*(3), 489–500.

Ferraro, P. J., & Price, M. K. (2013). Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment. *The Review of Economics and Statistics*, *95*(1), 64–73.

Festinger, L. (1957). *A theory of cognitive dissonance, Vol. 2*. Stanford University Press.

Fiebig, D. G., Keane, M. P., Louviere, J., & Wasi, N. (2010). The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. *Marketing Science*, *29*(3), 393–421.

Fifer, S., Rose, J., & Greaves, S. (2014). Hypothetical bias in stated choice experiments: Is it a problem? And if so, how do we deal with it? *Transportation Research Part A: Policy and Practice*, *61*, 164–177.

Freeman, M. A., Herriges, J. A., & Kling, C. L. (2014). *The measurement of environmental and resource values* (3rd ed.). New York, NY, USA: RFF Press, Routledge.

Gandullia, L. (2019). The price elasticity of warm-glow giving. *Economics Letters*, *182*, 30–32.

Gandullia, L., Lezzi, E., & Parciasepe, P. (2020). Replication with MTurk of the experimental design by Gangadharan, Grossman, Jones, and Leister (2018): Charitable giving across donor types. *Journal of Economic Psychology*, *78*, Article 102268.

Gentry, E. P., & Viscusi, W. K. (2016). The fatality and morbidity components of the value of statistical life. *Journal of Health Economics*, *46*, 90–99.

Greene, W. H., & Hensher, D. A. (2010). Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models. *Transportation*, *37*, 413–428.

Gu, Y., Hole, A. R., & Knox, S. (2013). Fitting the generalized multinomial logit model in stata. *The Stata Journal*, *13*(2), 382–397.

Hauser, D. J., & Schwarz, N. (2016). Attentive turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*(1), 400–407.

Hess, S., & Train, K. (2017). Correlation and scale in mixed logit models. *Journal of Choice Modelling*, *23*, 1–8.

Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online Readings in Psychology and Culture*, *2*(1), 8.

Hole, A. R. (2007). Fitting mixed logit models by using maximum simulated likelihood. *The Stata Journal*, *7*(3), 388–401.

Hole, A. R. (2008). Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. *Journal of Health Economics*, *27*(4), 1078–1094.

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, *14*(3), 399–425.

Hunt, N. C., & Scheetz, A. M. (2019). Using mturk to distribute a survey or experiment: Methodological considerations. *Journal of Information Systems*, *33*(1), 43–65.

Jacquemet, N., James, A. G., Luchini, S., Murphy, J. J., & Shogren, J. F. (2021). Do truth-telling oaths improve honesty in crowd-working? *PLoS One*, *16*(1), Article e0244958.

Jacquemet, N., Joule, R.-V., Luchini, S., & Shogren, J. F. (2013). Preference elicitation under oath. *Journal of Environmental Economics and Management*, *65*(1), 110–132.

Jacquemet, N., Luchini, S., Rosaz, J., & Shogren, J. F. (2019). Truth telling under oath. *Management Science*, *65*(1), 426–438.

Jin, Y., Andersson, H., & Zhang, S. (2020). Do preferences to reduce health risks related to air pollution depend on illness type? Evidence from a choice experiment in Beijing, China. *Journal of Environmental Economics and Management*, *103*, Article 102355.

Johnston, R. J., Boyle, K. J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T. A., et al. (2017). Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists*, *4*(2), 319–405.

Jones-Lee, M. W. (1974). The value of changes in the probability of death or injury. *Journal of Political Economy*, *82*(4), 835–849.

Jones-Lee, M. W. (2003). Safety and the saving of life: The economics of safety and physical risk, in R. In R. Layard, & S. Glaister (Eds.), *Cost-Benefit Analysis* (2nd ed.). (pp. 290–318). Cambridge, UK: Cambridge University Press.

Joule, R.-V., & Beauvois, J.-L. (2010). *La soumission librement consentie*. Presses Universitaires de France.

Joule, R.-V., Bernard, F., & Halimi-Falkowicz, S. (2008). Promoting ecocitizenship: In favour of binding communication. *International Scientific Journal for Alternative Energy and Ecology*, (6), 214–218.

Kemper, N. P., Popp, J. S., & Nayga, R. M. (2020). A query theory account of a discrete choice experiment under oath. *European Review of Agricultural Economics*, *47*(3), 1133–1172.

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., & Pozzer, A. (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, *525*(7569), 367–371.

Lerner, J. S., Gonzalez, R. M., Small, D. A., & Fischhoff, B. (2003). Effects of fear and anger on perceived risks of terrorism: A national field experiment. *Psychological Science*, *14*(2), 144–150.

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology*, *66*, 799–823.

Lokhorst, A. M., Werner, C., Staats, H., van Dijk, E., & Gale, J. L. (2013). Commitment and behavior change: A meta-analysis and critical review of commitment-making strategies in environmental research. *Environment and Behavior*, *45*(1), 3–34.

Lu, J. G. (2020). Air pollution: A systematic review of its psychological, economic, and social effects. *Current Opinion in Psychology*, *32*, 52–65.

Mahieu, P.-A., Andersson, H., Beaumais, O., Crastes dit Sourd, R., Hess, S., & Wolff, F.-C. (2017). Stated preferences: A unique database composed of 1657 recent published articles in journals related to agriculture, environment or healthes. *Review of Agricultural, Food and Environmental Studies*, *98*(3), 201–220.

Malone, T., & Lusk, J. L. (2018). A simple diagnostic measure of inattention bias in discrete choice models. *European Review of Agricultural Economics*, *45*(3), 455–462.

Mamkhezri, J., Thacher, J. A., Chermak, J. M., & Berrens, R. P. (2020). Does the solemn oath lower wtp responses in a discrete choice experiment application to solar energy? *Journal of Environmental Economics and Policy*, *9*(4), 447–473.

Mariel, P., Hoyos, D., Meyerhoff, J., Czajkowski, M., Dekker, T., Glenk, K., et al. (2021). *Environmental valuation with discrete choice experiments: Guidance on design, implementation and data analysis*. Springer Nature.

McFadden, D. (1974). Conditional logit analysis of discrete choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105–142). New York: Academic Press, (Chapter 4).

McFadden, D., & Train, K. (2000). Mixed mnl models for discrete response. *Journal of Applied Econometrics*, *15*(5), 447–470.

Molden, D. C. (2014). *Understanding priming effects in social psychology*. Guilford Publications.

Murphy, J. J., Allen, P. G., Stevens, T. H., & Weatherhead, D. (2005). A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics*, *30*(3), 313–325.

Ouvrard, B., Abildtrup, J., & Stenger, A. (2020). Nudging acceptability for wood ash recycling in forests: A choice experiment. *Ecological Economics*, *177*, Article 106748.

Ouvrard, B., Préget, R., Reynaud, A., & Tuffery, L. (2020). Nudging and subsidizing farmers to foster smart water meter adoption, CEE-M Working Paper 2020-14.

Reisch, L. A., Sunstein, C. R., & Gwozdz, W. (2017). Beyond carrots and sticks: Europeans support health nudges. *Food Policy*, *69*, 1–10.

Robinson, L. A., & Hammitt, J. K. (2011). Behavioral economics and the conduct of benefit-cost analysis: Towards principles and standards. *Journal of Benefit-Cost Analysis*, *2*(2), 1–51.

Ruchiraset, A., & Tantrakarnapa, K. (2020). Association of climate factors and air pollutants with pneumonia incidence in lampang province, Thailand: Findings from a 12-year longitudinal study. *International Journal of Environmental Health Research*, 1–10.

Sandorf, E. D. (2019). Did you miss something? Inattentive respondents in discrete choice experiments. *Environmental and Resource Economics*, *73*(4), 1197–1235.

Sunstein, C. R., & Reisch, L. A (2019). *Trusting nudges: Toward a bill of rights for nudging*. Routledge.

UK HM Treasury (2020). *The green book: Central government guidance of appraisal and evaluation*: *Guidance report*, HM Treasury, http://www.gov.uk/official-documents.

Vossler, C. A., Doyon, M., & Rondeau, D. (2012). Truth in consequentiality: Theory and field evidence on discrete choice experiments. *American Economic Journal: Microeconomics*, *4*(4), 145–171.

Vossler, C. A., & Evans, M. F (2009). Bridging the gap between the field and the lab: Environmental goods, policy maker input, and consequentiality. *Journal of Environmental Economics and Management*, *58*(3), 338–345.

Vossler, C. A., & Zawojska, E. (2020). Behavioral drivers or economic incentives? Toward a better understanding of elicitation effects in stated preference studies. *Journal of the Association of Environmental and Resource Economists*, *7*(2), 279–303.

Zhang, X., Zhang, X., & Chen, X. (2017). Happiness in the air: How does a dirty sky affect mental health and subjective well-being? *Journal of Environmental Economics and Management*, *85*, 81–94.