Social Science & Medicine 400 (2026) 119305

Contents lists available at ScienceDirect

Social Science & Medicine

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

Do we say what we do? Testing cheap talk to reduce hypothetical bias in moral choice experiments

Nicholas V.R. Smeele [a,b,c,*], Sander van Cranenburgh [d], Bas Donkers [b,e], Esther W. de Bekker-Grob [a,b,c]

[a] Erasmus School of Health Policy & Management, Erasmus University Rotterdam, the Netherlands
[b] Erasmus Choice Modelling Centre, Erasmus University Rotterdam, the Netherlands
[c] Erasmus Centre for Health Economics Rotterdam, Erasmus University Rotterdam, the Netherlands
[d] Department of Engineering Systems and Services, Delft University of Technology, the Netherlands
[e] Erasmus School of Economics, Erasmus University Rotterdam, the Netherlands

A B S T R A C T

Objective: Discrete choice experiments (DCEs) are widely used in healthcare to estimate willingness-to-pay (WTP) but may be affected by hypothetical bias (HB), especially in morally sensitive contexts. While cheap-talk is proposed as a mitigation strategy, its effectiveness in health-related DCEs involving moral trade-offs remains unclear. This study examines how cheap-talk influences WTP in such settings.

Methods: A split-sample DCE on organ transplantation policies was conducted, involving trade-offs between cost and morally salient outcomes: saving lives ("being alive") and improving quality of life ("having a life"). Respondents (N = 651) were randomly assigned to one of three survey arms: control (no manipulation), cheap-talk, or cheap-talk with follow-up question. Multinomial logit model in WTP space with a Taboo Trade-off Aversion (TTOA) specification was used to estimate treatment effects and interactions with religiosity.

Results: Exposure to the cheap-talk script reduced WTP for saving lives, indicating increased attention to financial considerations. WTP for quality-of-life improvements and avoiding taboo trade-offs remained unchanged. Religious respondents reported higher WTP to avoid taboo trade-offs in the control arm, but this gap disappeared under cheap-talk, showing that deliberation moderates monetary expressions among religious individuals without altering underlying convictions.

Conclusions: Cheap-talk promotes more reflective decision-making in morally sensitive health-related choices, particularly among individuals with strong moral or religious convictions. It reduces elevated WTP for taboo trade-offs, while its effect on other respondents is limited. Future research should combine stated and revealed preference data and explore models that account for non-compensatory moral decision rules to better capture complex moral preferences DCEs.

1. Introduction

Discrete choice experiments (DCEs) are widely used in health economics to measure preferences and inform decision-making on healthcare policies, products, and treatments (Whitty et al., 2014; Soekhai et al., 2019; Mott, 2018; Gadjradj et al., 2022). As a stated preference (SP) method, DCEs allow researchers to quantify the importance of various characteristics (attributes) of alternatives, estimate willingness-to-accept (WTA) risks or willingness-to-pay (WTP) for benefits, and forecast participation rates and market shares (Lancsar and Louviere, 2008; Train, 2003; Ben-Akiva and Lerman, 1985; Hensher et al., 2015). Rooted in random utility theory, DCEs assume that individuals choose the alternative that maximizes their utility while trading all attributes characterizing alternatives equally (rational, full-compensatory choice behavior) (McFadden, 1974, 1981; Lancaster, 1966). These behavioral premises enable the elicitation of preferences for a nearly infinite number of (not yet available) alternatives in a controlled environment.

The hypothetical nature of DCEs can result in differences between hypothetical (stated) and actual (revealed) choices, known as hypothetical bias (HB) (Beck et al., 2016; Quaife et al., 2018). A body of literature investigated HB in SP studies (Haghani et al., 2021a, 2021b; Murphy et al., 2005; Penn and Hu, 2018). Most studies suggest that WTP values are often overstated in hypothetical contexts (e.g., Lusk (2003)), with some studies also finding overestimated opt-in/out rates (e.g., Lambooij et al. (2015) and Ready et al. (2010)) and disordering of the attribute (level) importance (e.g., Özdemir et al. (2009)). Factors such as the type of good (public or private) and choice context (moral or non-moral) can influence the extent of HB (Haghani et al., 2021a). For instance, Johansson-Stenman and Svedsäter (2012) found significant HB

* Corresponding author. Erasmus University Rotterdam, Erasmus School of Health Policy & Management, Erasmus Choice Modelling Centre, Erasmus Centre of Health Economics Rotterdam, P.O. Box 1738, 3000 DR, Rotterdam, the Netherlands.
E-mail address: smeele@eshpm.eur.nl (N.V.R. Smeele).

https://doi.org/10.1016/j.socscimed.2026.119305
Received 1 December 2025; Received in revised form 26 February 2026; Accepted 13 April 2026
Available online 15 April 2026

in a moral public goods experiment but not in a morally neutral private goods context. They attribute this to individuals' desire to present a positive self-image, which results in overstated WTP for moral goods in hypothetical settings.

Comparing hypothetical (stated) choices with actual (revealed) behavior is often infeasible in healthcare, particularly for novel or hypothetical interventions. The cheap-talk method, an ex-ante intervention method designed to encourage more reflective decision-making in stated preference surveys, offers one way to address this challenge, which has gained popularity since the work by Cummings and Taylor (1999). The cheap-talk method originates from game theory and describes nonbinding communication from the researcher to respondents prior to the administration of the survey (Haghani et al., 2021b; Penn and Hu, 2018). The script used in the cheap-talk method explicitly discusses HB and urges respondents to treat the hypothetical scenarios as if they were encountered in real life (Cummings and Taylor, 1999). By prompting respondents to consider the actual opportunity costs of their choices, the cheap-talk script encourages more deliberate engagement with the hypothetical scenarios, influencing the expressed WTP.

Although the cheap-talk method has been widely tested in SP surveys for environmental and consumer goods, it has received limited attention in health economics (Haghani et al., 2021b). So far, to our knowledge, only one study has applied the cheap-talk method in a health-related DCE, focusing on patient preferences for rheumatoid arthritis treatments (Özdemir et al., 2009). No studies have, however, examined cheap-talk in health-related DCEs involving moral choice contexts. This limited application likely reflects methodological, practical, and ethical challenges: healthcare choices, particularly morally sensitive decisions such as life-or-death organ allocation, cannot be easily validated against actual behavior; outcomes are emotionally salient and context-dependent, making HB harder to operationalize; and explicit discussion of HB may raise concerns about priming or influencing ethical reasoning (Prentice et al., 2016; Miljeteig et al., 2021; Antiel et al., 2013). Research in moral psychology suggests that real moral decisions often differ from hypothetical choices, which tend to align with pro-social behaviors guided by moral proscriptions (e.g., cultural and religious values) (FeldmanHall et al., 2012a, 2012b; Teper et al., 2011).

To address this gap, this paper investigates the effects of the cheap-talk method in a health-related DCE involving a moral dimension, using the taboo trade-off aversion (TTOA) framework to explicitly model moral choice behavior. Without TTOA, differences in WTP across survey arms would conflate respondents' moral preferences with the effect of the cheap-talk script, making interpretation ambiguous. We conducted a split-sample DCE on organ transplantation policies to raise funds for reviving marginalized organs. By employing two variations of the cheap-talk approach, we examined whether and how cheap-talk influences WTP. As the first application of the cheap-talk in a moral health-related DCE using a theory-driven moral specification, this paper provides novel insights into how cheap-talk affects preferences and WTP in morally sensitive healthcare decisions.

2. Theoretical framework

2.1. Taboo trade-off aversion in choice behaviors

Moral choice behaviors are driven by individuals' beliefs about the right thing to do, guided by moral proscriptions derived from cultural and religious values (Cushman et al., 2006; Haidt, 2007; Smeele et al., 2023; Chorus, 2015). This study focused on taboo trade-off aversion (TTOA), a moral choice behavior that emerges when individuals face trade-offs involving both sacred and secular values. A trade-off becomes taboo when attributes linked to sacred values (e.g., patient deaths) are traded against those associated with secular values (e.g., money) (Fiske and Tetlock, 1997; Tetlock, 2003; Tetlock et al., 2000; Walzer, 2010). Sacred values like human lives or freedom are perceived as absolute and inviolable, whereas secular values like money are viewed as relative, violable, and exchangeable (Tetlock, 2003). The moral psychology literature suggests that individuals have an aversion to making such taboo trade-offs, particularly when sacred values are compromised (Fiske and Tetlock, 1997; Tetlock, 2003; Tetlock et al., 2000).

The TTOA behavior likely varies among individuals, depending on their cultural values (e.g., religious beliefs (Adamczyk et al., 2021; Chandrasekaran et al., 2023)) and underlying moral frameworks (Fiske and Tetlock, 1997; Tetlock et al., 1996). For example, individuals with a utilitarian orientation – who assess moral choices based on the principle of maximizing aggregate welfare – may find it acceptable to allocate scarce healthcare resources using cost-benefit analyses, such as calculating the number of lives saved per Euro spent. In contrast, individuals with a deontological orientation – who prioritize adherence to moral rules and duties irrespective of outcomes – would likely view such trade-offs as inherently impermissible or taboo, particularly when they involve assigning a monetary value to human life. Therefore, differences in moral judgment reflect the extent to which individuals perceive such trade-offs as violations of moral norms – especially when they involve trading sacred attributes (e.g., human life) against secular ones (e.g., money), which are guided by fundamentally different moral and social expectations (Fiske and Tetlock, 1997).

2.2. Taboo trade-off aversion in WTP-space models

Discrete choice models (DCMs), grounded in Random Utility Theory (RUT), provide a rigorous framework for analyzing and predicting choices (Lancaster, 1966; Samuelson, 1948; Luce, 1959). Under RUT, DCMs rely on the principle of Random Utility Maximization (RUM), which assumes that individuals evaluate and trade off all attributes to maximize latent utility. The RUM framework combines a noisy (random) element with the concept of utility maximization (McFadden, 1974, 1981).

Consider an individual n who faces a set of choice tasks S, each containing J alternatives. Each alternative i is defined by K attributes with levels x_{sik}. In a RUM-based DCM, the utility U_{sin} is decomposed into a deterministic and random component as follows:

$$U_{sin} = V_{sin} + \varepsilon_{sin} = \sum_{k=1}^{K} \beta_k \cdot x_{sik} + \varepsilon_{sin} \tag{1}$$

Where V_{sin} is the deterministic part of the utility function, β_k is the estimable preference parameter for the attribute k and ε_{sin} is the random stochastic component. Assuming a known distribution of ε_{sin} (e.g., Gumbel), the probability P_{sin} that alternative i is chosen is:

$$P_{sin} = Pr\left(V_{sin} > V_{sjn}\right) \text{ for all } j \neq i \, \epsilon \, S \tag{2}$$

To incorporate TTOA behavior, Chorus et al. (2018) and Smeele et al. (2025) extended the utility function to penalize trade-offs between sacred and secular attributes that are perceived as morally unacceptable or taboo. The deterministic component V_{sin} can be extended as follows:

$$V_{sin}^{TTOA} = \sum_{k=1}^{K} \beta_k \cdot x_{sik} + \tau_{a \to b} \cdot \max_{(a,b) \in T, j \in J-i} I_{x_{sia} \to x_{sib} \mid x_{sja},\, x_{sjb}} \tag{3}$$

Here, $\tau_{a \to b}$ captures the disutility associated with the aversion to taboo trade-offs (i.e., taboo penalty), T is a set of ordered pairs (a, b) with sacred attribute a (e.g., human lives) and secular attribute b (e.g., costs), and the indicator function $I_{x_{sia} \to x_{sib} \mid x_{sja},\, x_{sjb}}$ returns one if, relative to a comparator alternative j, the sacred attribute x_{sia} deteriorates while the secular attribute x_{sib} improves. The maximum operator ensures that a penalty is applied if at least one such taboo trade-offs exist for alternative i. This extension allows the model to account for aversion to taboo trade-offs. For more details and theoretical explanation of TTOA behavior in choice analyses, see Smeele et al. (2025).

In preference-space models as defined in Eqn. (1) and Eqn. (3), the WTP for a non-monetary attribute is derived indirectly as the ratio of its parameter β_k to the (negative) cost attribute parameter β_c. Alternatively, DCMs can be estimated directly in the WTP space, where parameters represent marginal WTP values – that is, the monetary amounts individuals are willing to pay for changes in non-cost attributes. This approach eliminates the need for post-estimation transformations, thereby reducing the risk of unstable or biased WTP estimates that may arise from imprecise cost parameters in preference-space model (Hensher et al., 2015; Train et al., 2005; Cameron and James, 1987; Cameron, 1988; Louviere, 2004).

WTP-space models are based on a reformulated utility function in which the cost coefficient normalizes the utility expression, allowing the estimated parameters to directly represent marginal WTP values. To estimate WTP-space models that account for TTOA behaviors, the deterministic part of the TTOA utility function, as defined in Eqn. (3) becomes:

$$V_{sin} = \beta_c \left(\sum_{k=1}^{K} \gamma_k \cdot x_{sik} + \gamma_t \cdot \max_{(a,b)\in T, j\in J-i} I_{x_{sia}\rightarrow x_{sib}} \mid x_{sja},\, x_{sjb} - X_{sic} \right) \quad (4)$$

Where V_{sin} is the deterministic utility of individual n for alternative i in choice task s, β_c is the cost coefficient, X_{sic} is a vector of cost attribute levels for the alternative, γ_k is the marginal WTP for each non-cost attribute k, X_{sik} is a vector of all non-cost attribute levels for the alternative, γ_t is the marginal WTP to avoid taboo trade-offs, and the indicator function $I_{x_{sia}\rightarrow x_{sib}} \mid x_{sja},\, x_{sjb}$ indicates the presence of a taboo trade-off in choice task s. This formulation expresses the utility in terms of WTP values: $\gamma_k = \beta_k/\beta_c$ and $\gamma_t = \tau_{a\rightarrow b}/\beta_c$, where γ_t represents the monetary value individuals associate with avoiding taboo trade-offs.

In this context, the estimated WTP for the TTOA penalty γ_t aligns with its interpretation as a disutility term in preference-space models. It reflects the monetary amount individuals are willing to pay to avoid making taboo trade-offs, such as accepting lower health premiums at the expense of more patient deaths. Conceptually, this makes γ_t directly comparable to the WTP for changes in attributes γ_k, enabling a meaningful interpretation of the relative strength of TTOA behaviors versus other attribute preferences in monetary terms.

2.3. Hypotheses development on the effects of the cheap-talk script

Following the theoretical foundation of TTOA and its modelling in the WTP space, we now develop testable hypotheses concerning the impact of the cheap-talk script on choice behaviors in morally salient contexts. Such contexts may involve trade-offs between sacred attributes (e.g., preserving human life or improving quality of life) and secular attributes (e.g., financial costs). These taboo trade-offs often evoke moral aversion, particularly when assigning monetary value to outcomes related to patient deaths or wellbeing.

In these situations, individuals may express pro-social preferences that reflect moral ideals rather than their actual preferences. Prior research has shown that when participants are not prompted to consider real-world consequences of their decisions, they may overstate their WTP for morally important outcomes, especially in hypothetical scenarios involving harm to others or public health goods (FeldmanHall et al., 2012a, 2012b; Teper et al., 2011; Greene et al., 2001, 2004). For example, FeldmanHall et al. (2012a) found that while participants claimed they were unwilling to give up real money to avoid harming others in hypothetical settings, they acted differently when real money was at stake. Similarly, studies show that WTP values in experiments involving moral goods are often inflated (Johansson-Stenman and Svedsäter, 2012).

To encourage more deliberative decision-making, a cheap-talk script can be used. This technique prompts individuals to reflect on opportunity costs and make choices as if real money were involved. In moral choice contexts, this increased realism may lead individuals to express lower WTP for desirable outcomes. Hence, our first hypothesis is:

Hypothesis 1. Individuals exposed to the cheap-talk script will report lower willingness to pay to save patient lives or improve patients' quality of life, due to increased sensitivity to real financial trade-offs.

Moreover, the moral framing of choice scenarios - where individuals must trade personal financial contributions against preventing patient deaths or suffering - makes the moral conflict highly salient. Literature on moral decision-making suggests that when individuals are prompted to consider real costs, they may become more aware of their tendency to overstate their WTP for morally sensitive outcomes due to the psychological discomfort (FeldmanHall et al., 2012b; Haidt, 2007; Smeele et al., 2023; Gigerenzer, 2010). Consequently, we expect the cheap-talk script to reduce the WTP for morally sensitive outcomes, such as avoiding taboo trade-offs, by encouraging more deliberative and less emotional decision-making. This results in more realistic and considered valuations of taboo trade-offs. Therefore, our second hypothesis is:

Hypothesis 2. Individuals exposed to the cheap-talk script will exhibit lower willingness to pay to avoid taboo trade-offs, reflecting reduced overstatement driven by immediate moral judgments rather than increased aversion to taboo trade-offs.

Religious and cultural beliefs play a significant role in shaping how individuals perceive and respond to taboo trade-offs (Fiske and Tetlock, 1997; Tetlock, 2003). Literature on moral psychology suggests that individuals' moral judgment – often shaped by religious tradition – affects their willingness to engage in such trade-offs.

Monotheistic traditions, grounded in a single divine authority, typically promote deontological moral reasoning, favoring adherence to moral rules (e.g., do not harm) regardless of outcomes (Piazza, 2012; Piazza and Landy, 2013; Piazza and Sousa, 2014; Banerjee et al., 2010; Barak-Corren and Bazerman, 2017). In these frameworks, certain actions are categorically impermissible and thus not open to negotiation or exchange. Similarly, polytheistic and virtue-based traditions emphasize moral character and intentions, leading to judgments that reject the commodification of sacred values (Johnson, 2005; McKay and Whitehouse, 2015; Norenzayan and Shariff, 2008). In contrast, non-religious individuals tend to rely more on consequentialist reasoning, where the morality of actions is assessed based on their outcomes rather than adherence to moral absolutes (Piazza and Landy, 2013; Piazza and Sousa, 2014).

In the context of taboo trade-offs – such as assigning monetary value to saving lives or preventing suffering –differences in moral reasoning are likely to produce varying WTP responses. Religious individuals may view placing a monetary value on human life as morally unacceptable, intensifying their aversion to such trade-offs and increasing WTP to avoid them. Non-religious individuals, guided more by consequentialist reasoning, may evaluate these trade-offs in economic terms when outcomes justify the means. This distinction is particularly relevant in health economic decisions, such as organ donation and transplantation, which are directly influenced by religious norms and moral beliefs. This leads to our third hypothesis:

Hypothesis 3. In the absence of the cheap-talk script, religious individuals will demonstrate a higher willingness to pay to avoid taboo trade-offs compared to non-religious individuals, reflecting a deontological or virtue-based rejection of such trade-offs.

However, exposure to the cheap-talk script may prompt religious individuals to reframe their moral judgments by incorporating financial considerations into their choice behaviors. This process may reduce TTOA by enabling morally motivated individuals to justify difficult choices through broader, outcome-oriented reasoning. As a result, the difference in WTP between religious and non-religious individuals may diminish under cheap-talk conditions, reflecting a convergence in WTP between religious and non-religious individuals as the former

incorporate financial considerations into their choice behavior. Hence, the fourth hypothesis is:

Hypothesis 4. When exposed to the cheap-talk script, the difference in WTP values associated with taboo trade-offs between religious and non-religious individuals will decrease, as religious individuals engage in more nuanced evaluations that incorporate financial consequences into their choice behaviors.

2.4. Hypotheses testing in the WTP-space model

To clarify how the hypotheses derived in Section 2.3 are empirically tested, we explicitly link each hypothesis to its operational specification in the WTP-space model. This is achieved by extending the utility function in Eqn. (4) with two- and three-way interaction terms among attributes, survey arms, and respondent covariates. For notational convenience, we introduce Θ to summarize components of the extended utility function; however, the full utility function contains all corresponding parameters explicitly, so that each Θ represents a collection of coefficients and interaction terms rather than a single estimable parameter. The full formulation allows all hypotheses to be tested simultaneously in a single pooled model, while ensuring that the relevant effects are identified through combinations of underlying parameters.

Consider the utility function as defined in Eqn. (4). We simplify its notation as follows:

$$\Theta^{base} = \sum_{k=1}^{K} \gamma_k \cdot x_{sik} + \gamma_t \cdot \max_{(a,b)\in T, j\in J-i} I_{x_{sia}\rightarrow x_{sib}} \mid x_{sja},\ x_{sjb} \tag{5}$$

$$V_{sin} = \beta_c \left(\Theta^{base} - X_{sic}\right) \tag{6}$$

To test Hypothesis 1, two-way interactions between each non-cost attribute (i.e., saving patient lives and improving patients' quality of life) and the two survey arms with manipulations M^m are included in the utility function. The corresponding interaction terms are defined as:

$$\Theta^{H1} = \sum\nolimits_{m=2}^{M} \sum_{k=1}^{K} \lambda_k^m \cdot x_{sik} \cdot M^m \tag{7}$$

Where λ_k^m captures the differences in marginal WTP of each non-cost attribute k between the treatment and control survey arms (M^m versus M^1). Evidence supporting Hypothesis 1 is obtained when λ_k^m is significantly negative, indicating lower WTP for the respective non-cost attribute among individuals exposed to the cheap-talk manipulation.

For Hypothesis 2, two-way interactions between the marginal WTP to avoid taboo trade-offs γ_t and the two survey arms with manipulations M^m are added to the utility function. Using a shorthand for the presence of a taboo trade-off:

$$\rho_{si} = \max_{(a,b)\in T, j\in J-i} I_{x_{sia}\rightarrow x_{sib}} \mid x_{sja},\ x_{sjb} \tag{8}$$

The interaction terms are:

$$\Theta^{H2} = \sum\nolimits_{m=2}^{M} \lambda_t^m \cdot \rho_{si} \cdot M^m \tag{9}$$

Here, λ_t^m captures differences in marginal WTP to avoid taboo trade-offs between individuals exposed to the cheap-talk manipulations and those in the control survey arm. This implies that Hypothesis 2 is supported when λ_t^m is significantly negative; that is, the WTP is lower whenever individuals are exposed to one of the manipulations.

Hypotheses 3 and 4 are evaluated by examining the marginal effect of religiosity on the WTP to avoid taboo trade-offs under each survey arm. For population-level interpretation, the dummy variable of religiosity covariate z_n for an individual n is mean-centred:

$$z_n^c = z_n - \bar{z} \tag{10}$$

This shifts the baseline to an average individual, rather than the non-religious subgroup, ensuring that main effects capture population-level WTP while interactions capture deviations for religious individuals. Using the shorthand in Eqn. (8), the interactions are denoted as:

$$\Theta^{H3-4} = \delta^1 \cdot \rho_{si} \cdot z_n^c + \sum\nolimits_{m=2}^{M} \delta^m \cdot \rho_{si} \cdot z_n^c \cdot M^m \tag{11}$$

Where z_n^c is the mean-centred religiosity covariate for individual n. In the control survey arm, δ^1 captures the difference between religious and average respondents (Hypothesis 3). Under the cheap-talk conditions, the marginal effect becomes $\delta^1 + \delta^m$, allowing us to assess whether exposure to cheap-talk reduces the religiosity-related differences in WTP (Hypothesis 4).

Combining Eqns. (5)–(11), the full utility function of the WTP-space model consists of baseline marginal WTP for non-cost attributes and taboo trade-offs Θ^{base} (see Eqn. (5)), augmented by interaction terms capturing the effects of cheap-talk exposure and individual religiosity (see Eqns. (7)–(11)) . Hence, the full utility function can be written in the notation-friendly form:

$$V_{sin} = \beta_c \left(\Theta^{base} + \Theta^{H1} + \Theta^{H2} + \Theta^{H3-4} - X_{sic}\right) \tag{12}$$

Again, each Θ term is a shorthand representing the underlying parameters and interaction terms explicitly included in the model. In this specification, the baseline parameters reflect the WTP among non-religious respondents in the control survey arm, while the interaction terms allow both cheap-talk and religiosity to systematically shift marginal WTP. This formulation ensures that all hypotheses are tested in a single model using the pooled sample, and that differences in WTP – both across survey arms and between religious and non-religious respondents – are identified through well-defined conditional marginal effects rather than isolated coefficients.

3. Methods

3.1. Discrete choice experiment

We conducted a split-sample DCE originally developed and validated by Smeele et al. (2025). The DCE focused on taboo trade-offs in the context of organ transplantation policies in the Netherlands, with attributes and levels carefully designed to capture morally salient and policy-relevant trade-offs. For a detailed description and justification of the DCE design and its theoretical foundations, see Smeele et al. (2025).

The DCE employed a two-stage choice design: participants first chose between two unlabeled alternatives (Policy A and Policy B), followed by a choice between their selected alternative in the first stage and the status quo (i.e., Current policy). The experiment elicited public preferences regarding increases in health premiums to fund the revival of marginalized organs for transplantation. Table 1 presents the attributes and levels used in the experiment.

Organ transplantation inherently involves morally salient trade-offs, particularly concerning the sanctity of human life, distributive fairness,

Table 1
Attributes and levels in the discrete choice experiment.

Attribute	Levels
Number of deaths on the waiting list (out of 2,000 patients on the waiting list)	200 400 600 800 1,000
Change in patients' quality of life after transplantation	Moderate Good
Change in basic health premium per month	€5 more €10 more €15 more €20 more

and the ethical tension of commodifying human organs. In this context, the use of marginalized organs - although potentially life-saving – does not guarantee full restoration of health or post-transplant quality of life, thereby intensifying ethical concerns.

The experimental design contained two types of taboo trade-offs, reflecting distinct dimensions of human existence: "being alive" (involving life-or-death outcomes) versus "having a life" (involving quality-of-life considerations). In each choice task, participants compared alternatives that varied in terms of health premiums (a secular attribute), the number of patient deaths, and post-transplant quality of life (both sacred attributes). Therefore, a taboo trade-off was defined as occurring when one alternative resulted in (i) higher patient deaths and/or (ii) lower quality of life, while changes in health premiums are lower, compared to the counter-alternative. Fig. 1 shows an example choice task with a taboo trade-off between policies A and B in the domain "being alive" – that is, the taboo trade-off related to life-or-death outcomes.

The choice tasks reflect real-world policy dilemmas, where decision-makers routinely balance health outcomes against financial costs for both patients and the broader public. Therefore, the experiment not only investigates the effect of the cheap-talk script on choice behaviors but also provides insights relevant to healthcare policymakers.

3.2. Cheap-talk strategy

The cheap-talk method was employed to examine how informational prompts influence choice behaviors. Participants were randomly assigned to one of three survey arms. In the first survey arm (control), respondents completed choice tasks without exposure to a cheap-talk script. The second survey arm included the cheap-talk script, providing contextual information on HB and encouraged participants to consider the opportunity costs of their choices. The third survey arm used the same script as the second survey arm, followed by a brief follow-up question. The follow-up question was designed to reinforce engagement with the script's content by increasing attentiveness, rather than inducing normative or policy-oriented framing. Importantly, the script does not advocate any policy position or socially desirable response. We hypothesized that respondents exposed to the cheap-talk script would report lower WTP than those in the control arm, with the effect being more pronounced in the third survey arm, where the follow-up question further encouraged reflection on the trade-offs involved.

The cheap-talk script and follow-up question were developed following the structured approach proposed by Cummings and Taylor (1999) and were pre-tested prior to fielding the survey to ensure clarity, comprehension, and appropriate cognitive priming. Appendix A presents the English translation of the cheap-talk materials as well as the original Dutch version used in the experiment.

3.3. Survey

All survey arms consisted of four sections. Each survey arm started with an introductory text and informed consent. The first section also contained questions regarding respondent characteristics (age, gender, education, religion, household composition, household income). Religiosity was measured using self-reported religious affiliation with eight options: Catholicism, Protestantism, Islam, Judaism, Hinduism, Buddhism, Other, or No religious belief. Those selecting one of the first seven categories were coded as "religious", while those selecting "no religious belief" were coded as "non-religious". This operationalization captures self-identification with a religious tradition, which is a relevant proxy for moral and value-guided decision-making in our experiment.

The second section focused on introducing the DCE; that is, it explained the choice context, attributes and levels, and the format of the choice tasks. The cheap-talk scripts were presented at the end of the second section for the two survey arms that included them. The third section contained the choice tasks of the DCE (see Section 2.2. for more details). The fourth section consisted of additional questions about respondents' post-hoc rationalizations for their choices in taboo choice tasks, attitudes towards risks for their health, as in Chorus et al. (2020) and moral statements (Graham et al., 2013), and other attitudinal and behavioral constructs potentially related to TTOA behavior. Appendix B shows all covariates included in the survey.

3.4. Study sample

Participants were recruited through market research panels and sampled to be representative of the adult population of the Netherlands (18 years or older). The required minimum sample size for estimating the pooled model was approximately 260 participants, calculated using the parametric framework proposed by de Bekker-Grob et al. (de Bekker-Grob et al., 2015). This sample size applies to the pooled model rather than to each survey arm individually. With our research budget, we targeted complete data from 770 participants. For further details on the study sample, see Smeele et al. (2025). To compare respondent

	Current policy	Policy A	Policy B
Number of patients on the waiting list who die prematurely	1,200	600	800
Improvement in quality of life of patients after organ transplantation	Good	Moderate	Good
Your health insurance premium per month	€138	↑ €10 more	↑ €5 more
YOUR CHOICE		☐	☐

Fig. 1. Example choice task with taboo trade-off in the domain of "being alive".

characteristics across survey arms, we used chi-squared tests for categorical variables.

3.5. Model estimation

We estimate the WTP-space model using the Multinomial Logit (MNL) framework, following the approach described by Train and Weeks (Train et al., 2005). The MNL model assumes homogeneous preferences across respondents; however, our primary objective is not to recover individual-level heterogeneity, but to identify average differences in WTP across experimental conditions induced by the cheap-talk interventions.

The treatment effects are identified through interaction terms between survey arms, religiosity, and taboo trade-offs (Eqns. (8)–(11)), which operationalize the theory-driven heterogeneity underlying our hypotheses. In this framework, differences in WTP across survey arms are interpreted as changes in expressed valuations under different informational conditions, rather than changes in underlying moral preferences.

More flexible models, such as Mixed Logit, allow for unobserved preference heterogeneity and could capture additional individual-level variation. However, unobserved heterogeneity primarily affects the precision of estimates rather than the consistency of interaction effects associated with experimentally randomized treatments. Because survey arm assignment is random, the estimated interaction effects identifying the impact of cheap-talk remain unbiased under the MNL specification, while more flexible models would substantially increase model complexity and reduce interpretability.

Accordingly, the MNL model provides a transparent, computationally stable, and theoretically aligned framework for testing sample-level treatment effects of cheap-talk on WTP. All models are estimated in R 4.4.2 using the Apollo 0.3.6 package, based on the utility specification described in Section 2.4.

To formally assess whether the follow-up question had an effect beyond the cheap-talk script itself, we conducted Wald tests of coefficient equality between survey arms 2 and 3. These tests assess whether WTP estimates differ between the two treatment arms. Wald tests are appropriate in this context as both models are nested within a pooled model (i.e., a single estimated specification) and survey arm assignment is randomized.

4. Results

4.1. Comparison of respondent characteristics

The survey was completed by 691 respondents, resulting in 6,910 choice observations after excluding certain respondents during data preprocessing. For further details on the dataset, see Smeele et al. (2025).

To ensure comparability in respondent composition across the three survey arms, we down-sampled respondents in survey arms 2 and 3 (i.e., the cheap-talk and cheap-talk with a follow-up question arms). The initial sub-samples in these survey arms differed significantly from the control arm in terms of educational level. To correct for this imbalance, we randomly removed respondents from overrepresented educational categories while preserving balanced distributions across all other demographic variables.

The final dataset comprises 651 respondents, resulting in 6,510 choice observations distributed across the three survey arms. The control survey arm included 226 respondents (2,260 choice observations), while survey arms 2 and 3 included 216 and 209 respondents (2,160 and 2,090 choice observations), respectively.

Chi-squared tests confirmed that respondent characteristics did not differ significantly across the survey arms. Table 2 summarizes the demographic composition of the study sample across the three survey arms.

Table 2
Demographic composition of the study sample across the survey arms.

	Default		Cheap talk		Cheap talk + Follow-up		P[a]
	#	%	#	%	#	%	
Total	226	100.0	216	100.0	209	100.0	
Gender							0.509
Male	100	44.2	102	47.2	87	41.6	
Female	126	55.8	114	52.8	122	58.4	
Age[b]							0.710
Low	113	50.0	103	47.7	110	52.6	
Med	65	28.8	68	31.5	64	30.6	
High	48	21.2	45	20.8	35	16.8	
Income[c]							0.550
Low	67	29.6	77	35.6	77	36.8	
Med	94	41.6	84	38.9	77	36.8	
High	65	28.8	55	25.5	55	26.4	
Education[d]							0.257
Low	56	24.8	62	28.7	46	22.0	
Med	75	33.2	78	36.1	86	41.2	
High	95	42.0	76	35.2	77	36.8	
Household composition							0.888
Single	50	22.1	64	29.6	53	25.4	
With partner	63	27.9	51	23.6	54	25.8	
With partner and child(ren)	63	27.9	59	27.3	62	29.7	
With partner, child (ren), and others	9	4.0	6	2.8	8	3.8	
With partner and others	3	1.3	3	1.4	1	0.5	
Single parent with child(ren)	10	4.4	11	5.1	8	3.8	
Single parent with child(ren) and others	5	2.2	4	1.9	1	0.5	
Other	23	10.2	18	8.3	22	10.5	
Religious							0.816
Yes	102	45.1	91	42.1	91	43.5	
No	124	54.9	125	57.9	118	56.5	

[a] P-values from Chi-squared tests of independence assessing whether the distribution of each categorical variable differs across the three survey arms.
[b] Age: Low <40 years; Med = 40-60 years; High >60 years.
[c] Income (net household income per month): Low < €2,550; Med = €2,550-€4,250; High > €4,250.
[d] Education (Dutch system): Low = at least VMBO, HAVO/VWO onderbouw, or MBO-1; Med = HAVO/VWO bovenbouw, or MBO-2; High = HBO/WO bachelor or higher.

4.2. WTP estimates

Table 3 presents the estimated WTP values and their standard errors from the WTP-space model. It also reports two- and three-way interaction effects on the WTP estimates to test the hypotheses defined in Section 2.2.

First, we examine the main effects of the WTP estimates in the first (control) survey arm, which serves as the baseline in the full model. As shown in Table 3, all WTP estimates are statistically significant ($p < 0.01$), except for the estimate associated with avoiding taboo trade-offs ($p > 0.1$). Respondents in the control arm are willing to pay an additional 3.72 Euros per month in health premiums to save 100 patient lives (i.e., to avoid an increase in mortality) and 21.09 Euros per month to improve patients' quality of life from moderate to good condition. The WTP estimate for avoiding taboo trade-offs is not statistically significant and is positive. Although not significant, the WTP is very close to zero, suggesting that, on average, respondents in the control arm do not exhibit a strong preference for avoiding taboo trade-offs.

Having established the baseline WTP estimates, we next assess the hypotheses outlined in Section 2.3. Hypothesis 1 posits that exposure to the cheap-talk script reduces WTP for saving patient lives and improving patients' quality of life. The interaction terms between the cheap-talk

Table 3
The Multinomial Logit (MNL) model in WTP-space.

Attribute	Estimate	S.e.
Patient lives saved (x 100 fewer deaths on the waiting list)	3.72***	0.5389
Change in patients' quality of life after transplantation to *Moderate* condition	0	(fixed)
Change in patients' quality of life after transplantation to *Good* condition	21.09***	3.0996
Taboo	0.01	1.5407
Interaction effects	**Estimate**	**S.e.**
Patient lives saved (x 100 fewer deaths on the waiting list) x Survey Arm 1	0	(fixed)
Patient lives saved (x 100 fewer deaths on the waiting list) x Survey Arm 2	−0.93***	0.2766
Patient lives saved (x 100 fewer deaths on the waiting list) x Survey Arm 3	−1.33***	0.3051
Change in patients' quality of life after transplantation to Good condition x Survey Arm 1	0	(fixed)
Change in patients' quality of life after transplantation to Good condition x Survey Arm 2	−0.95	2.0596
Change in patients' quality of life after transplantation to Good condition x Survey Arm 3	0.18	2.1016
Taboo x Survey Arm 1	0	(fixed)
Taboo x Survey Arm 2	0.51	1.7800
Taboo x Survey Arm 3	1.44	1.8214
Taboo x Religious	6.69***	2.4213
Taboo x Religious x Survey Arm 1	0	(fixed)
Taboo x Religious x Survey Arm 2	−8.63**	3.4316
Taboo x Religious x Survey Arm 3	−8.02**	3.4338
Log-likelihood	−6,660.99	
Adj. Rho Sq.	0.0667	
BIC	13,444.91	
Number of observations	6,510	
Number of parameters	12	

* significant at 10%, ** significant at 5%, *** significant at 1%, *S.e.* Standard error.
Note: Religious dummy was mean-centrered (Religion = 0.436). To compute WTP for non-religious respondents, multiply interaction terms by −0.436; for religious respondents, multiply by 1 − 0.436 = 0.564. Baseline parameters reflect an average respondent.

survey arms (arms 2 and 3) and the attribute for saving patient lives are negative and statistically significant ($p < 0.01$), indicating a reduction in WTP relative to the control arm. Specifically, respondents exposed to the cheap-talk script (survey arm 2) exhibit a 0.93 Euros decrease per month in health premium, while those in the cheap-talk with follow-up question arm (survey arm 3) show a 1.33 Euros decrease, compared to the control arm.

In contrast, the interaction effects between cheap-talk arms and the quality-of-life improvement attribute are not statistically significant, with small negative and slightly positive estimates for survey arms 2 and 3, respectively. This suggests that exposure to the cheap-talk script significantly reduces WTP for saving patient lives but does not significantly influence WTP for quality-of-life improvements. Thus, these results provide partial support for Hypothesis 1.

Hypothesis 2 suggests that exposure to the cheap-talk script reduces respondents' WTP to avoid taboo trade-offs by mitigating the tendency to overstate valuations in morally sensitive contexts. This hypothesis is tested via interaction effects between the taboo attribute and the cheap-talk survey arms. Contrary to the hypothesis, the interaction estimates for survey arms 2 and 3 were positive, indicating slightly higher WTP to avoid taboo trade-offs compared to the control arm. However, these differences were not statistically significant ($p > 0.1$). Importantly, the lack of significant differences suggests that respondents in the control condition did not overstate their WTP to avoid taboo trade-offs. Because the cheap-talk scripts are designed to correct for overstated valuations, the findings may imply that respondents at the population level already considered the opportunity costs in the hypothetical scenarios, rendering the intervention largely redundant.

Although Hypothesis 2 is not empirically supported, there may still be meaningful variation in WTP across specific respondent subgroups, such as those differing in religiosity. This leads to Hypothesis 3, which posits that, in the absence of the cheap-talk script (i.e., in the control survey arm), religious respondents will exhibit a higher WTP to avoid taboo trade-offs than non-religious respondents. This expectation follows from the theoretical framework, which suggests that religious individuals are more likely to reject such trade-offs on deontological or virtue-ethical grounds rather than evaluate them in compensatory terms.

Consistent with Hypothesis 3, the two-way interaction between the taboo attribute and religiosity is positive and statistically significant ($p < 0.01$). The WTP of religious respondents to avoid taboo trade-offs is 6.69 Euros higher per month than that of non-religious respondents in the control arm. These findings suggest that religious individuals may be more morally sensitive to taboo trade-offs and express this sensitivity through a stronger economic aversion. Therefore, the results provide empirical support for Hypothesis 3, highlighting the influential role of moral and religious frameworks in shaping valuations of taboo trade-offs.

Finally, Hypothesis 4 predicted that exposure to the cheap-talk script would reduce the difference in WTP to avoid taboo trade-offs between religious and non-religious respondents, as religious individuals incorporate financial considerations into their moral judgments.

The results are consistent with this expectation. In the baseline condition, religious respondents reported a substantially higher WTP to avoid taboo trade-offs (3.79 Euros) compared to non-religious respondents (−2.90 Euros), yielding an absolute difference of 6.69 Euros. This gap, however, markedly reduces in the cheap-talk arms: in Arm 2, the absolute difference falls to 1.94 Euros, and in Arm 3 to 1.33 Euros. The corresponding three-way interaction effects between the taboo attribute, religiosity, and cheap-talk arms are negative and statistically significant (−8.63 Euros and −8.02 Euros, $p < 0.05$). These results indicate that the cheap-talk scripts substantially reduce the inflated WTP among religious respondents while leaving non-religious respondents largely unaffected.

Notably, the WTP of religious respondents decreased in both cheap-talk arms (−0.57 Euros in Arm 2 and 0.70 Euros in Arm 3) compared to the baseline condition (3.79 Euros). This indicates that religious respondents in Arm 2 are no longer willing to pay to avoid taboo trade-offs and would instead require compensation. This change reflects a shift from purely moral reasoning to more deliberative, financially-informed decision-making. When prompted to consider real financial consequences, religious individuals appear to rely more on economic considerations, potentially overriding their initial moral aversion. In line with Hypothesis 4, these results show that the cheap-talk scripts effectively reduce the initial moral aversion that inflated the baseline WTP, leading to a more outcome-oriented evaluation of taboo trade-offs. Table 4 summarizes the support for each hypothesis, detailing the direction and size of the key WTP estimates underlying these results.

To assess the effect of the follow-up question beyond the cheap-talk script, we first analyzed the downsampled sample, finding no statistically significant differences for any attribute or interaction ($p > 0.1$). This indicates that the follow-up question did not meaningfully alter WTP. Robustness checks on the full sample (see Appendix C) confirm that our substantive findings regarding the effects of cheap-talk and the moderating role of religiosity remain largely robust, while acknowledging a small additional effect of the follow-up question on WTP for saving patient lives.

5. Discussion

This paper examined the effect of the cheap-talk method on choice behaviors and expressed willingness to pay (WTP) in a health-related discrete choice experiment (DCE) with a moral dimension. Prior studies suggest that WTP is often overstated in hypothetical contexts,

Table 4
Summary of hypothesis testing results.

Hypothesis	Supported?	Key WTP Estimates (Euros/ month)	Reasoning
H1: Cheap-talk reduces WTP for saving lives and improving quality of life.	Partially supported	Saving lives: 3.72 (Control), −0.93*** (Arm 2), −1.33*** (Arm 3) QoL: 21.09 (Control), −0.95 (Arm 2, n.s.), +0.18 (Arm 3, n.s.)	Significant reduction in WTP for saving lives under cheap-talk conditions; no significant effect on WTP for QoL improvements.
H2: Cheap-talk reduces WTP to avoid taboo trade-offs	Not supported	Taboo: 0.01 (Control), 0.51 (Arm 2, n.s.), 1.44 (Arm 3, n.s.)	Positive but non-significant changes in WTP to avoid taboo trade-offs; no evidence of the cheap-talk effect.
H3: Without cheap-talk, religious individuals have higher WTP to avoid taboo trade-offs.	Supported	Taboo x Religious: 6.69*** (Control)	Significant positive WTP difference, showing that religious individuals have higher WTP to avoid taboo trade-offs when not exposed to the cheap-talk script.
H4: Cheap-talk reduces WTP differences between religious and non-religious individuals.	Supported	Taboo x Religious x Arm 2: −8.63** Taboo x Religious x Arm 3: −8.02** Compared to the Taboo x Religious: +6.69 (Control)	Significant negative interaction terms indicate that exposure to the cheap-talk scripts reduces WTP among religious respondents, suggesting their WTP converges toward that of non-religious respondents relative to the baseline condition.

* significant at 10%, ** significant at 5%, *** significant at 1%, *n.s.* Not significant, *QoL* Quality of Life.

influenced by factors such as the type of good (public or private) and the choice context (moral or non-moral) (Haghani et al., 2021a, 2021b; Murphy et al., 2005; Penn and Hu, 2018; Johansson-Stenman and Svedsäter, 2012). To explore this, we used a split-sample DCE on organ transplantation policies involving taboo trade-offs between money and "being alive" (involving life-or-death outcomes) versus "having a life" (involving quality-of-life considerations). Participants were randomly assigned to one of three survey arms: a control arm, a survey arm with cheap-talk script, and one with the cheap-talk script and a follow-up question. We tested four hypotheses related to the effects of these treatments on WTP estimates. It is important to emphasize that our results pertain to changes in WTP under different informational conditions, rather than direct evidence of changes in underlying moral preferences or ethical commitments.

The first hypothesis suggested that the cheap-talk conditions would reduce individuals' WTP for saving patient lives and quality-of-life improvements. Our findings provide partial support for this hypothesis. We observed significantly lower WTP for saving patient lives in the cheap-talk arms, suggesting that the cheap-talk method may enhance individuals' sensitivity to financial constraints and opportunity costs – thereby prompting more deliberative and budget-conscious choices. In contrast, WTP for improving patients' quality of life was not significantly affected by the cheap-talk treatments.

This asymmetry is somewhat counterintuitive: one might expect that life-or-death decisions – often seen as the most morally charged – would be less influenced by pragmatic reasoning. Yet the observed decline in WTP for saving lives suggests that these decisions may be emotionally salient or intuitively inflated in hypothetical settings, making them more responsive to reflective interventions, such as cheap-talk. In contrast, the stability of WTP for quality-of-life improvements may indicate that individuals approach these choices as more personal judgments, guided by deontological or absolutist intuitions that resist cost-benefit trade-offs (Bartels and Medin, 2007; Baron and Spranca, 1997; Baron, 1994). These patterns may be consistent with, and may be interpreted through, social intuitionist models of decision-making, which propose that moral judgments often arise from intuitive responses and only occasionally shift through reflection or contextual cues (Haidt, 2001, 2007; Baron, 1994).

The second hypothesis predicted that exposure to the cheap-talk script would reduce respondents' WTP to avoid taboo trade-offs by mitigating potential overstatement in morally sensitive contexts. However, the results do not support this hypothesis. The interaction effects for the cheap-talk arms are positive but statistically insignificant ($p > 0.1$), indicating no meaningful difference relative to the baseline condition. Because these differences are small, they are consistent with random variation rather than a systematic effect of the intervention.

The absence of significant changes suggests that respondents' initial valuations were not substantially inflated by the hypothetical choice context. Since cheap-talk scripts are designed to prompt more deliberative decision-making, their failure to shift WTP implies that the potential of HB was limited or absent in the baseline condition. Thus, although Hypothesis 2 is not supported, the findings indicate that respondents' valuations remained stable across conditions, with the cheap-talk interventions exerting no measurable influence on the WTP.

For the third hypothesis, we predicted that, when not exposed to the cheap-talk conditions (i.e., in the control arm), religious individuals would exhibit a higher WTP to avoid taboo trade-offs than non-religious individuals, reflecting a principled moral objection rather than a compensatory valuation. This prediction is supported by our findings. The interaction term between religiosity and taboo trade-offs was positive and statistically significant ($p < 0.01$), indicating that religious individuals are willing to pay 6.69 Euros more per month than non-religious individuals to avoid such choices. This pattern may suggest that religious respondents interpret these trade-offs as morally unacceptable (FeldmanHall et al., 2012a, 2012b; Tetlock et al., 2000), although this interpretation is exploratory.

To better understand the moral underpinnings of the elevated WTP values, future research could employ Latent Class Logit (LCL) models to identify subgroups, such as religious respondents, who may rely on non-compensatory decision rules. In particular, conjunctive or disjunctive rules may better capture categorical moral stances. This approach could help clarify whether high WTP among religious individuals reflects strong valuation or a principled refusal to engage in taboo trade-offs – a distinction that standard RUM models may obscure.

The fourth and final hypothesis suggested that exposure to the cheap-talk script would reduce the gap between religious and non-religious individuals in their WTP to avoid taboo trade-offs. This prediction is supported by our findings. The elevated WTP observed among religious respondents in the control arm disappeared, or at least became non-significant, once they were exposed to the cheap-talk script. As the cheap-talk script encourages respondents to reflect on the opportunity costs embedded in the choice task, the implied WTP for religious respondents becomes negative in both cheap-talk arms, whereas the WTP for non-religious respondents remains largely unchanged.

To assess whether the follow-up question had an effect beyond the cheap-talk script itself, we conducted Wald tests of coefficient equality between survey arms 2 and 3 for all relevant WTP. These tests revealed no statistically significant differences between the two arms ($p > 0.1$). This indicated that the additional follow-up question did not

meaningfully alter expressed WTP relative to the cheap-talk script alone, suggesting that the primary effect operates through the initial cheap-talk interventions rather than through repeated or reinforced prompting.

While the cheap-talk script does not necessarily alter underlying moral convictions, it appears to influence how respondents approach the choice task. In this sense, cheap-talk may operate by shifting the cognitive frame of decision-making – through mechanisms such as framing, priming, or increased deliberation – rather than by correcting a well-defined bias in a narrow statistical sense. One possibility is that it engages a more reflective mode of decision-making, prompting religious respondents to consider opportunity costs or trade-offs they may not have weighed in the control condition. As a result, the choices they make under cheap-talk conditions lead to substantially lower implied WTP estimates. This pattern aligns with dual-process theories of moral judgment (Haidt, 2001; Brand, 2016), where deliberative reasoning can moderate the influence of intuitive moral responses on expressed decisions. Accordingly, reductions in WTP under cheap-talk should not be interpreted as definitive evidence of bias correction but rather as indicative of a shift toward more reflective and consequential decision-making in hypothetical settings, without implying changes in underlying moral convictions.

Nonetheless, our findings should be interpreted in light of several limitations. First, as with all stated preference methods, participants made hypothetical choices, and it is unclear to what extent these responses reflect real-world behavior. Even with the cheap-talk script, decisions may differ when choices have real consequences, which could affect the external validity of our findings. Second, although we observed reductions in WTP for certain outcomes under cheap-talk conditions, we cannot draw definitive conclusions about the method's effectiveness in mitigating hypothetical bias, as our study lacks revealed preference data for validation. Future work combining stated and revealed data would be needed to assess the external validity of cheap-talk as an ex-ante mitigation tool. Third, the cheap-talk script and follow-up question could have primed specific moral or consequentialist frames, potentially shaping how participants interpreted and responded to the trade-offs in our experiment. Finally, although the achieved sample size exceeded the minimum required to estimate the pooled model, detecting interaction effects generally requires larger samples than estimating main effects. As such, the study was powered to identify interaction effects of moderate to large magnitude, whereas smaller differences in preferences across survey arms may not have been detectable. This limitation should be considered when interpreting non-significant interaction estimates.

To address these limitations and extend our work, future studies could vary the framing of taboo trade-offs, test alternative versions of cheap-talk scripts or moral prompts, and draw on moral psychology frameworks such as Moral Foundations Theory (Graham et al., 2013) or Morality-as-Cooperation (Curry et al., 2019). Additionally, exploring how religious belief systems interact with different moral reasoning styles – such as deontological versus consequentialist thinking – may offer deeper insights into the heterogeneity of moral preferences in discrete choice contexts. Using models like LCL models could help uncover non-compensatory or categorical decision rules, particularly among respondents with strong moral or religious convictions.

Moreover, future research could examine additional sources of observable and unobservable preference heterogeneity via Mixed Logit or hierarchical models, including socio-demographic factors (e.g., gender, age, education, income) and personal or family values and beliefs. While such extensions are beyond the scope of the current study, they could provide complementary insights into individual-level variation in WTP and the role of personal values in shaping choice behavior in morally sensitive contexts.

6. Conclusion

This paper is the first to examine the effects of the cheap-talk method in a morally sensitive, health-related DCE. We find that cheap-talk reduces WTP for saving lives, while WTP for quality-of-life improvements or avoiding taboo trade-offs remains largely stable, suggesting that some moral preferences are more responsive to deliberation than others.

A key finding concerns the difference between religious and non-religious respondents: in the baseline condition, religious participants report higher WTP to avoid taboo trade-offs, but cheap-talk removed this gap, indicating that deliberation can moderate monetary expressions of moral sensitivity without altering underlying convictions. The cheap-talk script and follow-up question were designed to increase cognitive engagement; formal tests comparing the cheap-talk arm with and without the follow-up question indicate that the additional question did not meaningfully alter respondents' WTP, confirming that the observed effects are primarily driven by the cheap-talk intervention itself.

These findings demonstrate that carefully designed stated preference studies can elicit meaningful variation in morally sensitive choices while minimizing potential confounds from framing or demand effects. They highlight the importance of cognitive engagement in DCEs and underscore the need to consider heterogeneity in moral decision-making. Future research should further explore the interplay between deliberative interventions, moral framing, and individual preferences, and combine stated and revealed preference data to more directly assess hypothetical bias.

Role of the funders/sponsors

The funders have no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data, preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Funding/support

This study was financially supported by the Erasmus School of Health Policy & Management, the Erasmus initiative 'Smarter Choices for Better Health' Seed Grant, the European Research Council (ERC) Consolidator Grant BEHAVE – 724431, and the Flagship 'Organ Transplantation: making unsuitable donor organs suitable', which has received funding from the Health & Technology Convergence; a public initiative where Delft University of Technology, Erasmus University Rotterdam, and Erasmus Medical Center are joining forces [20600210.003].

CRediT authorship contribution statement

Nicholas V.R. Smeele: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Writing – original draft. **Sander van Cranenburgh:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Bas Donkers:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing. **Esther W. de Bekker-Grob:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

The authors have no conflicts to disclose.

Acknowledgements

The authors acknowledge the Erasmus School of Health Policy & Management, the Erasmus Initiative 'Smarter Choices for Better Health' (Seed Grant obtained by the first author), and the European Research Council (ERC Consolidator Grant BEHAVE – 724431) for the financial support. This project is also part of the Flagship 'Organ Transplantation: making unsuitable donor organs suitable', which has received funding

from the Health & Technology Convergence; a public initiative where Delft University of Technology, Erasmus University Rotterdam, and Erasmus Medical Center are joining forces [20600210.003].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.socscimed.2026.119305.

Data availability

Data will be made available on request.

References

Adamczyk, A., Liu, Y.H., Scott, J., 2021. Understanding the role of religion in shaping cross-national and domestic attitudes and interest in abortion, homosexuality, and pornography using traditional and Google search data. Soc. Sci. Res. 100, 102602. https://doi.org/10.1016/j.ssresearch.2021.102602.

Antiel, R.M., Curlin, F.A., James, K.M., Tilburt, J.C., 2013. The moral psychology of rationing among physicians: the role of harm and fairness intuitions in physician objections to cost-effectiveness and cost-containment. Philos. Ethics Humanit. Med. 8 (1), 13. https://doi.org/10.1186/1747-5341-8-13.

Banerjee, K., Huebner, B., Hauser, M., 2010. Intuitive moral judgments are robust across variation in gender, education, politics and religion: a large-scale web-based study. J. Cognit. Cult. 10 (3–4), 253–281. https://doi.org/10.1163/156853710X531186.

Barak-Corren, N., Bazerman, M.H., 2017. Is saving lives your task or God's? Religiosity, belief in god, and moral judgment. Judgm. Decis. Mak. 12 (3), 280–296. https://doi.org/10.1017/S193029750000588X.

Baron, J., 1994. Nonconsequentialist decisions. Behav. Brain Sci. 17 (1), 1–10. https://doi.org/10.1017/S0140525X0003301X.

Baron, J., Spranca, M., 1997. Protected values. Organ. Behav. Hum. Decis. Process. 70 (1), 1–16. https://doi.org/10.1006/obhd.1997.2690.

Bartels, D.M., Medin, D.L., 2007. Are morally motivated decision makers insensitive to the consequences of their choices? Psychol. Sci. 18 (1), 24–28. https://doi.org/10.1111/j.1467-9280.2007.01843.x.

Beck, M.J., Fifer, S., Rose, J.M., 2016. Can you ever be certain? Reducing hypothetical bias in stated choice experiments via respondent reported choice certainty. Transp. Res. Part B Methodol. 89, 149–167. https://doi.org/10.1016/j.trb.2016.04.004.

Ben-Akiva, M.E., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press.

Brand, C. (Ed.), 2016. Dual-Process Theories in Moral Psychology. Springer, Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-12053-5.

Cameron, T.A., 1988. A new paradigm for valuing non-market goods using referendum data: maximum likelihood estimation by censored logistic regression. J. Environ. Econ. Manag. 15 (3), 355–379. https://doi.org/10.1016/0095-0696(88)90008-3.

Cameron, T.A., James, M.D., 1987. Estimating willingness to pay from survey data: an alternative pre-test-market evaluation procedure. J. Mark. Res. 24 (4), 389. https://doi.org/10.2307/3151386.

Chandrasekaran, S., Key, K., Ow, A., et al., 2023. The role of community and culture in abortion perceptions, decisions, and experiences among Asian Americans. Front. Public Health 10, 982215. https://doi.org/10.3389/fpubh.2022.982215.

Chorus, C.G., 2015. Models of moral decision making: literature review and research agenda for discrete choice analysis. Journal of Choice Modelling 16, 69–85. https://doi.org/10.1016/j.jocm.2015.08.001.

Chorus, C., Sandorf, E.D., Mouter, N., 2020. Diabolical dilemmas of COVID-19: an empirical study into Dutch society's trade-offs between health impacts and other effects of the lockdown. In: Angeli, F. (Ed.), Plos One, 15, e0238683. https://doi.org/10.1371/journal.pone.0238683, 9.

Chorus, C.G., Pudāne, B., Mouter, N., Campbell, D., 2018. Taboo trade-off aversion: a discrete choice model and empirical analysis. Journal of Choice Modelling 27, 37–49. https://doi.org/10.1016/j.jocm.2017.09.002.

Cummings, R.G., Taylor, L.O., 1999. Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method. Am. Econ. Rev. 89 (3), 649–665.

Curry, O.S., Mullins, D.A., Whitehouse, H., 2019. Is it good to cooperate? Testing the theory of morality-as-cooperation in 60 societies. Curr. Anthropol. 60 (1), 47–69. https://doi.org/10.1086/701478.

Cushman, F., Young, L., Hauser, M., 2006. The role of conscious reasoning and intuition in moral judgment: testing three principles of harm. Psychol. Sci. 17 (12), 1082–1089. https://doi.org/10.1111/j.1467-9280.2006.01834.x.

de Bekker-Grob, E.W., Donkers, B., Jonker, M.F., Stolk, E.A., 2015. Sample size requirements for discrete-choice experiments in healthcare: a practical guide. Patient 8 (5), 373–384. https://doi.org/10.1007/s40271-015-0118-z.

FeldmanHall, O., Dalgleish, T., Thompson, R., Evans, D., Schweizer, S., Mobbs, D., 2012b. Differential neural circuitry and self-interest in real vs hypothetical moral decisions. Soc. Cognit. Affect Neurosci. 7 (7), 743–751. https://doi.org/10.1093/scan/nss069.

FeldmanHall, O., Mobbs, D., Evans, D., Hiscox, L., Navrady, L., Dalgleish, T., 2012a. What we say and what we do: the relationship between real and hypothetical moral choices. Cognition 123 (3), 434–441. https://doi.org/10.1016/j.cognition.2012.02.001.

Fiske, A.P., Tetlock, P.E., 1997. Taboo trade-offs: reactions to transactions that transgress the spheres of justice. Polit. Psychol. 18 (2), 255–297. https://doi.org/10.1111/0162-895X.00058.

Gadjradj, P.S., Smeele, N.V.R., de Jong, M., et al., 2022. Patient preferences for treatment of lumbar disc herniation: a discrete choice experiment. J. Neurosurg. Spine 36 (5), 704–712. https://doi.org/10.3171/2021.8.SPINE21995.

Gigerenzer, G., 2010. Moral satisficing: rethinking moral behavior as bounded rationality. Top. Cogn. Sci. 2 (3), 528–554. https://doi.org/10.1111/j.1756-8765.2010.01094.x.

Graham, J., Haidt, J., Koleva, S., et al., 2013. Moral foundations theory. In: Advances in Experimental Social Psychology, 47. Elsevier, pp. 55–130. https://doi.org/10.1016/B978-0-12-407236-7.00002-4.

Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., Cohen, J.D., 2004. The neural bases of cognitive conflict and control in moral judgment. Neuron 44 (2), 389–400. https://doi.org/10.1016/j.neuron.2004.09.027.

Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., Cohen, J.D., 2001. An fMRI investigation of emotional engagement in moral judgment. Science 293 (5537), 2105–2108. https://doi.org/10.1126/science.1062872.

Haghani, M., Bliemer, M.C.J., Rose, J.M., Oppewal, H., Lancsar, E., 2021a. Hypothetical bias in stated choice experiments: part I. Macro-scale analysis of literature and integrative synthesis of empirical evidence from applied economics, experimental psychology and neuroimaging. Journal of Choice Modelling 41, 100309. https://doi.org/10.1016/j.jocm.2021.100309.

Haghani, M., Bliemer, M.C.J., Rose, J.M., Oppewal, H., Lancsar, E., 2021b. Hypothetical bias in stated choice experiments: part II. Conceptualisation of external validity, sources and explanations of bias and effectiveness of mitigation methods. Journal of Choice Modelling 41, 100322. https://doi.org/10.1016/j.jocm.2021.100322.

Haidt, J., 2001. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. Psychol. Rev. 108 (4), 814–834. https://doi.org/10.1037/0033-295X.108.4.814.

Haidt, J., 2007. The new synthesis in moral psychology. Science 316 (5827), 998–1002. https://doi.org/10.1126/science.1137651.

Hensher, D.A., Rose, J.M., Greene, W.H., 2015. Applied Choice Analysis, second ed. Cambridge University Press. https://doi.org/10.1017/CBO9781316136232.

Johansson-Stenman, O., Svedsäter, H., 2012. Self-image and valuation of moral goods: stated versus actual willingness to pay. J. Econ. Behav. Organ. 84 (3), 879–891. https://doi.org/10.1016/j.jebo.2012.10.006.

Johnson, D.D.P., 2005. God's punishment and public goods: a test of the supernatural punishment hypothesis in 186 world cultures. Hum. Nat. 16 (4), 410–446. https://doi.org/10.1007/s12110-005-1017-0.

Lambooij, M.S., Harmsen, I.A., Veldwijk, J., et al., 2015. Consistency between stated and revealed preferences: a discrete choice experiment and a behavioural experiment on vaccination behaviour compared. BMC Med. Res. Methodol. 15 (1), 19. https://doi.org/10.1186/s12874-015-0010-5.

Lancaster, K.J., 1966. A new approach to consumer theory. J. Polit. Econ. 74 (2), 132–157. https://doi.org/10.1086/259131.

Lancsar, E., Louviere, J., 2008. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. Pharmacoeconomics 26 (8), 661–677. https://doi.org/10.2165/00019053-200826080-00004.

Louviere, J.J., 2004. Random utility theory-based stated preference elicitation methods: applications in health economics with special reference to combining sources of preference data. https://api.semanticscholar.org/CorpusID:16555201.

Luce, R.D., 1959. Individual Choice Behavior. John Wiley.

Lusk, J.L., 2003. Effects of cheap talk on consumer willingness-to-pay for golden rice. Am. J. Agric. Econ. 85 (4), 840–856. https://doi.org/10.1111/1467-8276.00492.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. Frontiers in Econometrics. Published online.

McFadden, D., 1981. Econometric models of probabilistic choice. In: Structural Analysis of Discrete Data with Econometric Applications. MIT Press, pp. 198–272.

McKay, R., Whitehouse, H., 2015. Religion and morality. Psychol. Bull. 141 (2), 447–473. https://doi.org/10.1037/a0038455.

Miljeteig, I., Forthun, I., Hufthammer, K.O., et al., 2021. Priority-setting dilemmas, moral distress and support experienced by nurses and physicians in the early phase of the COVID-19 pandemic in Norway. Nurs. Ethics 28 (1), 66–81. https://doi.org/10.1177/0969733020981748.

Mott, D.J., 2018. Incorporating quantitative patient preference data into healthcare decision making processes: is HTA falling behind? Patient 11 (3), 249–252. https://doi.org/10.1007/s40271-018-0305-9.

Murphy, J.J., Allen, P.G., Stevens, T.H., Weatherhead, D., 2005. A meta-analysis of hypothetical bias in stated preference valuation. Environ. Resour. Econ. 30 (3), 313–325. https://doi.org/10.1007/s10640-004-3332-z.

Norenzayan, A., Shariff, A.F., 2008. The origin and evolution of religious prosociality. Science 322 (5898), 58–62. https://doi.org/10.1126/science.1158757.

Penn, J.M., Hu, W., 2018. Understanding hypothetical bias: an enhanced meta-analysis. Am. J. Agric. Econ. 100 (4), 1186–1206. https://doi.org/10.1093/ajae/aay021.

Piazza, J., 2012. "If you love me keep my commandments": religiosity increases preference for rule-based moral arguments. Int. J. Psychol. Relig. 22 (4), 285–302. https://doi.org/10.1080/10508619.2011.638598.

Piazza, J., Landy, J.F., 2013. "Lean not on your own understanding": belief that morality is founded on divine authority and non-utilitarian moral judgments. Judgm. Decis. Mak. 8 (6), 639–661. https://doi.org/10.1017/S1930297500004691.

Piazza, J., Sousa, P., 2014. Religiosity, political orientation, and consequentialist moral thinking. Soc. Psychol. Personal. Sci. 5 (3), 334–342. https://doi.org/10.1177/1948550613492826.

Prentice, T., Janvier, A., Gillam, L., Davis, P.G., 2016. Moral distress within neonatal and paediatric intensive care units: a systematic review. Arch. Dis. Child. 101 (8), 701–708. https://doi.org/10.1136/archdischild-2015-309410.

Quaife, M., Terris-Prestholt, F., Di Tanna, G.L., Vickerman, P., 2018. How well do discrete choice experiments predict health choices? A systematic review and meta-analysis of external validity. Eur. J. Health Econ. 19 (8), 1053–1066. https://doi.org/10.1007/s10198-018-0954-6.

Ready, R.C., Champ, P.A., Lawton, J.L., 2010. Using respondent uncertainty to mitigate hypothetical bias in a stated choice experiment. Land Econ. 86 (2), 363–381. https://doi.org/10.3368/le.86.2.363.

Samuelson, P.A., 1948. Consumption theory in terms of revealed preference. Economica 15 (60), 243. https://doi.org/10.2307/2549561.

Smeele, N.V.R., Chorus, C.G., Schermer, M.H.N., De Bekker-Grob, E.W., 2023. Towards machine learning for moral choice analysis in health economics: a literature review and research agenda. Soc. Sci. Med. 326, 115910. https://doi.org/10.1016/j.socscimed.2023.115910.

Smeele, N.V.R., Van Cranenburgh, S., Donkers, B., Schermer, M.H.N., De Bekker-Grob, E.W., 2025. Taboo trade-off aversion in choice behaviors: a discrete choice model and application to health-related decisions. Soc. Sci. Med. 386, 118606. https://doi.org/10.1016/j.socscimed.2025.118606.

Soekhai, V., de Bekker-Grob, E.W., Ellis, A.R., Vass, C.M., 2019. Discrete choice experiments in health economics: past, present and future. Pharmacoeconomics 37 (2), 201–226. https://doi.org/10.1007/s40273-018-0734-2.

Teper, R., Inzlicht, M., Page-Gould, E., 2011. Are we more moral than we think?: exploring the role of affect in moral behavior and moral forecasting. Psychol. Sci. 22 (4), 553–558. https://doi.org/10.1177/0956797611402513.

Tetlock, P.E., 2003. Thinking the unthinkable: sacred values and taboo cognitions. Trends Cognit. Sci. 7 (7), 320–324. https://doi.org/10.1016/S1364-6613(03)00135-9.

Tetlock, P.E., Kristel, O.V., Elson, S.B., Green, M.C., Lerner, J.S., 2000. The psychology of the unthinkable: taboo trade-offs, forbidden base rates, and heretical counterfactuals. J. Pers. Soc. Psychol. 78 (5), 853–870. https://doi.org/10.1037/0022-3514.78.5.853.

Tetlock, P.E., Peterson, R.S., Lerner, J.S., 1996. Revising the value pluralism model: incorporating social content and context postulates. In: The Psychology of Values, 8. Lawrence Erlbaum Associates, Inc, pp. 25–51. *The Ontario Symposium, Vol. 8.* The Ontario symposium on personality and social psychology.

Train, K.E., 2003. Discrete Choice Methods with Simulation. Cambridge University Press. https://doi.org/10.1017/CBO9780511753930.

Train, K., Weeks, M., 2005. Discrete choice models in preference space and willingness-to-pay space. In: Scarpa, R., Alberini, A. (Eds.), *Applications of Simulation Methods in Environmental and Resource Economics*. Vol 6. the Economics of Non-market Goods and Resources. Springer-Verlag, pp. 1–16. https://doi.org/10.1007/1-4020-3684-1_1.

Walzer, M., 2010. Spheres of Justice: a Defense of Pluralism and Equality. Nachdr. Basic Books.

Whitty, J.A., Lancsar, E., Rixon, K., Golenko, X., Ratcliffe, J., 2014. A systematic review of stated preference studies reporting public preferences for healthcare priority setting. Patient 7 (4), 365–386. https://doi.org/10.1007/s40271-014-0063-2.

Özdemir, S., Johnson, F.R., Hauber, A.B., 2009. Hypothetical bias, cheap talk, and stated willingness to pay for health care. J. Health Econ. 28 (4), 894–901. https://doi.org/10.1016/j.jhealeco.2009.04.004.