Desire, moral evaluation or sense of duty: The modal framing of stated preference elicitation
Eva Wanek, Sacha Bourgeois-Gironde, Alda Mari

To cite this version:

HAL Id: halshs-04354354
https://shs.hal.science/halshs-04354354
Submitted on 19 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Desire, moral evaluation or sense of duty: the modal framing of stated preference elicitation

Eva Wanek\(^1\)*, Sacha Bourgeois-Gironde\(^{1,2,3}\), Alda Mari\(^1\)

\(^1\)Ecole normale supérieure – PSL University, Institut Jean-Nicod, Paris, France
\(^2\)Department of Economics, Université Paris 2 Panthéon-Assas, Paris, France
\(^3\)Faculty of Law, University of Haifa, Israel

*Corresponding author: eva.wanek@ens.psl.eu

ABSTRACT

Contingent valuation surveys generally elicit stated preferences by asking how much a respondent would be willing to pay for an environmental improvement. By drawing on linguistic theory, we propose that the modal phrasing of this question establishes a particular type of commitment towards a hypothetical payment, namely a subjective want or desire. Based on the idea that beyond subjective desires, considerations about what is morally adequate may guide expressed values and that elicitation of these can be linguistically facilitated, we employ an experimental framework to investigate the effects of different modals (willing, should, and appropriate) in the elicitation question on stated preferences. We find that elicited amounts with appropriate are higher than those elicited with willing and should for environmental improvements more associated with use values, while differences are non-significant for environmental improvements more associated with non-use values. We discuss the implications of our findings for stated preference studies, as well as the potential broader theoretical implications that our study entails regarding linguistic representations of the moral entrenchment of environmental values.

KEYWORDS
INTRODUCTION

Economic environmental valuation is generally performed in order to produce values that can be used for environmental policy-making, and, ideally, to improve environmental protection (Hansjürgens et al., 2017). However, the topic is contentious, notably due to the questionable capacity of economic valuation methods to capture the multitude of value dimensions inherent to nature besides instrumental values, such as intrinsic, shared and social values (Kallis et al., 2013; Vatn, 2000).

Classical contingent valuation method (CVM) studies represent the most widely used stated preference methods for estimating economic values of non-market goods, such as environmental goods and services (Carson, 2000; Hanemann, 1994). In these surveys participants typically read a description of a good or service and a proposed intervention and are then asked how much they would be willing to pay for an environmental improvement (willingness-to-pay, WTP) or how much they would ask in compensation for an environmental degradation (willingness-to-accept, WTA). The stated amounts are taken as proxies for individual economic preferences that can be aggregated to obtain the total public valuation for a given good or service. The underlying definition of economic preferences is in the neoclassical tradition, i.e. preferences are seen as choices that maximise the satisfaction of subjective wants and desires (von Neumann and Morgenstern, 1944).

Stated preference studies constitute value articulating institutions (VAI) (Farrell, 2007; Jacobs, 1994; Vatn, 2005). VAI define a set of rules concerning the valuation process, which do not only reveal, but also heavily influence the results of a valuation process, through the types of values that are elicited (monetary values, preference rankings or weights,…), and the
content and form of the information given about the environmental issue at stake. Restricting environmental VAI to wants and desires expressed in WTP may not be reflective of the multitude of values associated with environmental public goods and services, which include broader moral and society-oriented considerations (Gómez-Baggethun and Ruiz-Pérez, 2011; Sen, 1995; Vatn and Bromley, 1994).

The different understandings of economic value articulations have been analysed in the context of the WTP/WTA gap (Kahneman et al., 1990), a phenomenon in which WTA structurally exceeds WTP for the same goods or services, even when income effects are compensated for. Kahneman et al. (1990) attributed the WTP/WTA gap to the endowment effect, referring to the observation that people value things more once they own them, which in turn is based on loss aversion. In their experimental investigation, the endowment effect was instantaneously detectable for consumption goods that were handed out and described as belonging to participants, but not for tokens with merely induced financial value, a finding which highlights the importance of both the goods’ properties and the structural context of valuation for preference construction.

This interpretation was supplemented by Hart and Latacz-Lohmann (2001) with regard to economic environmental valuation, who argued that the WTP/WTA gap may be indicative of a feeling of moral responsibility towards a valued entity and towards others who value it.

Sagoff (1988; 1998) underlined the distinction between consumer and citizen preferences: Consumer preferences refer to what the individual thinks is good for them in terms of the satisfaction of personal, individual-level desires. This notion of preferences thus corresponds to the mainstream idea of economic preferences, which is derived from a utilitarian conception of rational choice. Citizen preferences, on the other hand, pertain to considerations about fairness and justice, about people’s own predicaments and those of others – both now and in the future (Sen, 1995). The notion of citizen preferences is derived
from a deontological understanding of rational choice and thereby conceptually diverges from
the mainstream definition of economic preferences.

Massenberg et al. (2023) proposed a conceptual framework of how social values enter
economic environmental valuation, which consists of three spheres: social, natural and
contextual. The social sphere, in their understanding, comprises social factors that influence
the valuing individual(s) independently of the concrete valuation exercise, such as
transcendental values, beliefs, dispositions, and cultural aspects. The natural sphere relates
to the valued ecosystem and its intrinsic value, defined as a weak anthropocentric intrinsic
value, meaning valuing an entity for itself, in a non-instrumental way, albeit inevitably from a
human perspective. The social and natural spheres are linked by human-nature relationships
and a sense of connectedness, which are all highly heterogeneous among different groups
and individuals. The contextual sphere, finally, contains the valuation context which defines a
set of rules based on specific ontology and epistemology. Notably, the valuation context leads
from abstract values to concrete (economic) value articulations. Massenberg et al. (2023)
highlighted the role of information for the likelihood of considering specific aspects of the
natural and social spheres, alluding to the possibility that tweaks in the descriptions of
environmental goods and services can give indications as to which value dimensions are
deemed relevant in a given valuation context.

The idea of linguistic correspondents of preferences regarding different value
dimensions was introduced by Sagoff (1998, p. 214-215), who contended that "[w]hile the
words 'I want' are likely to introduce a consumer preference, a statement that begins with
'society should...' is likely to express a citizen preference". Influences of linguistic features on
consumer behaviour have also been examined in the field of marketing research, where
structural features of language as well as lexical-semantic features have been shown to affect
preferences (Zhang et al., 2003; Zhang and Schmitt, 1998). We take Sagoff’s proposed link
between linguistic representations of different preference orderings and the empirical findings
from marketing research as a conceptual basis, which we aim to refine by focusing on specific modal structures to assess economic environmental values.

The notion of linguistically accessed preferences is highly pertinent in the context of stated preference methods, as they strongly rely on the verbal construction of counterfactual markets: To account for the hypothetical nature of the payment, it is necessary to make use of linguistic terms that prompt participants to take a specific stance towards it.

The most common elicitation question in CVM studies roughly has the following structure: *What is the maximum amount you would be willing to pay (for the proposed intervention)?* The specific stance that the participant is asked to take towards the payment is here indicated by the phrase *would be willing*. This phrase is what linguists call a modal (Kratzer 1981, Portner 2009, Giannakidou and Mari 2021), because it refers to a hypothetical situation. In this case, the modal expresses a want or desire, as indicated by *willing*.

In this study, we experimentally investigate the effect of different linguistic modals on stated preferences. In particular, we ask not only what people would be willing to pay, but also what they think they should pay and what they think is appropriate to pay for environmental interventions. This allows us to isolate the influence of different types of modals on stated preferences, which, to our knowledge, has not yet been done. Our aim is to contribute to the debate on the expression of concepts underlying economic environmental valuation that has been ongoing for many years. Arias-Arévalo et al. (2018) linked metaphors used for describing human-nature relationships to different value dimensions. The expression “gaining from nature” is taken to relate to the view that human welfare and economic productivity depend on the benefits derived from ecosystems, and therefore to the instrumental value domain. Conversely, the metaphor “living for nature” is associated with the intrinsic value domain, here defined as the value that ecosystems have in and of themselves. “Living in nature”, finally, is linked to two value domains: fundamental and eudaimonistic values. The fundamental value domain encompasses all systems of relations and processes that are necessary to sustain life...
in general. Eudaimonistic values relate to entities and processes that are conditions for “a
good human life”, yet not driven by merely self-interested preferences, but extend to notions
of what is considered meaningful, including altruism and environmental justice. How people
think and talk about nature is thus indicative of the values ascribed to it. Instead of employing
different metaphors or using different kinds of information to describe environmental goods
and services, our experimental investigation is focused only on the elicitation question of a
stated preference study. This is a highly restricted approach to accessing the plurality of
environmental values, as we remain in a stated preference framework eliciting individual
monetary values, and do not present different kinds of information or different metaphors to
describe the scenarios. This differentiates our approach from a number of studies in which
alternatives to WTP have already been applied, notably in the domain of what has become
known as "deliberative monetary valuation" (DMV) (Spash, 2007). This methodology emerged
following the conceptualization of environmental preferences as more inclusive than the
traditional consumer view and as a response to calls for the democratisation of valuation
processes (Kenter, et al., 2016a; Niemeyer and Spash, 2001; Pelletier et al., 1999; Sagoff,
1998). In contrast to traditional stated preference methods, these studies feature deliberative
activities. Group discussions and learning sessions are often part of these activities (Álvarez-
Farizo et al., 2009; Lienhoop and Völker, 2016). To account for the community-based nature
of the economic values they aim to elicit, some DMV studies employ alternatives to the
standard WTP elicitation question. For example, they ask for a “fair price” to pay for an
environmental good or service (Szabó, 2011; Kenter, 2016). However, the alternative
elicitation questions are implemented along with deliberative activities, and so their effects are
likely confounded with the effects of deliberation. The relationship between WTP and fair
prices is not unequivocal. Szabó (2011) observed higher elicited amounts for group-
deliberated fair prices than for individual WTP, in contrast to Kenter (2016) and Kenter et al.
(2016b), who measured lower deliberated fair prices than non-deliberated individual WTP.
Concepts like fair price, how much one should pay and what would be appropriate to pay tend
to be used interchangeably in these studies. For example, Szabó (2011) asked the participants
in his study which amount they would find "acceptable/approvable", and later referred to their answers as "fair price", or "how much society [should] pay".

By contrast, in this study we concentrate on the effect of the subtle differences between linguistic modals on stated preferences, which constitutes a focused way to access different moral conceptions and evaluative tendencies underlying stated preferences. The study might then bear potential broader implications and help understand how deontic and other morally-based values are entertained by the participants during CVM experiments depending on the linguistic format of the elicitation format they face. By contrasting linguistic, and in particular modal formats of the elicitation mechanism, we take a very minimalistic approach, which could be seen as a “lower bound” for the expression of different values related to the environment, i.e. values that may not correspond (only) to the definition of consumer preferences, but that can also be based on deontological ethics, considerations of justice and norms. Those kinds of value motivations have been recognized as being integral to environmental values, as discussed above (Arias-Arévalo et al., 2018; Hart and Latacz-Lohmann, 2001; Massenberg et al., 2023; Sagoff, 1988; 1998). We assume that if a small change in linguistic modals used in establishing practical commitments (as defined below) can lead to differentiated stated preferences, even when measured only on a monetary scale, then this is indicative of the richness of environmental values, which too linguistically rigid experimental standards in traditional CVM studies may conceal or flatten out. The paper thus contributes to advancing the understanding of how values, manifested in linguistic expressions, are articulated with preferences and how preferences vary according to the nature of the value that the linguistic expression conveys.

The rest of this paper is organised as follows: In section 2 we review literature on the concepts relevant for our study, based on which we formulate our hypotheses in section 3. Section 4 lays out our experimental design and the results. We discuss the results and their practical and, especially, theoretical implications, as well as the limits imposed by our experimental design, and conclude in section 5.
2 CONCEPTS

2.1 Preferences, attitudes and practical commitments

CVM was developed to measure economic preferences, meaning how much people would actually pay to “buy” the environmental good or service in question. This classical view was termed by Kahneman et al. (1993) the "purchase model"; CVM respondents are assumed to be able to perform an evaluation of their utility levels in a state of the world where the good is provided, as well as in a state of the world where it is not provided and assign a monetary value to the difference in utility. This view has been challenged for a long time and the analysis of stated preferences has notably been enriched with insights from psychology and philosophy (Gregory, Lichtenstein, and Slovic, 1993; Sagoff, 1988; Sen, 1995). It is clear today that stated preferences are not unidimensional, and that they have the potential to be influenced by many factors, such as the level of information that is provided, the valuation context and social interactions. What stated preferences truly represent is still not without controversy. In the Review on the economics of biodiversity (Dasgupta, 2021), CVM estimates were referred to as "opinions", a notion which is not further defined. However, it seems to be interpreted as an epistemic concept, since Dasgupta (2021) went on to argue that opinions which are not sufficiently informed about the valuation object should not be used for valuation.

The much-voiced criticism of CVM that stated preferences often rest on an insufficient information base has been one of the driving forces behind the recent development of deliberative monetary valuation (DMV) (Bunse et al., 2015; Spash, 2007; Zografos and Howarth, 2008). Studies that apply this methodology include elements of discussion and deliberation among participants and often allow for exchange between participants and experts on the topic. Deliberation is supposed to fulfil two main functions: First, to allow for preferences to be constructed, since they might not be pre-existing (especially for unfamiliar goods and services) and second, to prompt participants to adopt a less self-centred perspective through the exchange with others. In some of these studies, the latter is reinforced
by using alternative economic value elicitation questions, such as "What do you think is the maximum increase in the price [...] that is acceptable?" (Szabó, 2011, p. 40), labelling this a fair price, or how much an average individual should pay for the good. In a study by Kenter (2016) a group-based fair price is elicited in addition to individual WTP in order to shift the conceptualization of payments from a purchase model to a public policy model, based on the considerations of the linguistic correspondents of citizen preferences (Sagoff, 1998). Kenter (2016) highlighted, however, that a firm theoretical and empirical basis is still missing for the fair price concept. By employing an approach anchored in linguistic theory, we hope to facilitate understanding of language independently of deliberative processes, i.e. to identify the capacity of different modals to access values that may transcend consumer preferences. By doing so, our goal is not to propose an alternative way to formulate those questions, but to demonstrate that we may reveal underlying moral stances indicative of the heterogeneity of environmental values. Our driving assumption, then, is that a linguistic format for a preference elicitation mechanism actually reveals a moral attitude, assimilable to a practical commitment which can eventually expressed through a hypothetical payment or other environmental value indicators. Below we outline the ongoing debate on how stated preferences should be interpreted, and notably how they relate to attitudes. We then go on to explain the understanding of attitudes that underlies the design of our study.

A branch of literature that investigates the motives for stated WTP of laypeople was established by Kahneman et al. (1993); Kahneman and Ritov (1994) and Kahneman et al. (1999). They proposed that stated WTP estimates are better understood as an expression of attitudes than as economic preferences, with attitudes being defined as "an evaluative tendency, which can be favourable or unfavourable" based on Eagly and Chaiken (1993). This was justified by the observation that WTP measures were strongly correlated with elicited attitudes towards briefly described environmental and public health interventions (Kahneman and Ritov, 1994). One group of participants to the study was asked to state their WTP and another group was asked to state their attitudes, measured on Likert scales in three domains,
with one question respectively: support for government intervention, satisfaction derived from
contribution, and an importance rating of the issue. The strong correlations between monetary
amounts and attitude measures were taken as an indication that WTP is actually an
expression of attitudes that is forced onto a monetary scale.

Ryan and Spash (2011) extended this idea in a WTP study for environmental
interventions. In their study, attitudes were defined in line with Kahneman and Ritov (1994). In
contrast, they were not measured with a single item, but with psychometric scales featuring
several items. Attitude items included for example how good or bad the proposed intervention
was perceived to be and how effective it was. In addition to attitudes, Ryan and Spash (2011)
included measures for subjective norms and perceived behavioural control in the framework
of the theory of planned behaviour. These were also measured on multi-item psychometric
scales. The authors found that all of the psychometric scales, attitudinal and non-attitudinal
ones, predicted whether participants stated a positive WTP amount. From this, they postulate
that what underlies WTP is a complex process of psychological apprehension which is not
only, but also driven by attitudes.

In the present study we also pursue an attitudinal approach to stated preferences, but
we define attitudes more narrowly as propositional attitudes (Schwitzgebel, 2019), referring to
the mental state of having some stance, take or opinion about a proposition or about the
potential state of affairs in which that proposition is true. The following sentence can be taken
as an example of this understanding of attitudes:

(1) Ahmed (the subject) hopes (the attitude) that Alpha Centauri hosts
intelligent life (the proposition).

This focused approach allows us to have more control over the attitudes we are testing than
do Kahneman and Ritov (1994) and Ryan and Spash (2011). While they investigated the
correlation between attitudes and stated WTP, we focus on the effect of different propositional
attitudes (willing, should, appropriate) on hypothetical payments (HP), based on the idea of
the survey as a value articulating institution which mediates the form of the elicited values through its own structure (Farrell, 2007).

We will use the expression *hypothetical payment* rather than *willingness to pay* to be able to refer to all of the elicited economic values. This reflects our approach of situating WTP in a wider realm of modals and associated stances. To avoid confusion with different understandings of attitudes and to underline the associated behavioural implications, we will favour the term *practical commitment*, based on Bratman et al. (1987) and Bratman (2012), to refer to the specific attitudes that we look at. Bratman suggests that people’s practical commitments are elements of stable plans of action which are usually grounded in practical normative reasons. We will employ the term practical commitment in a slightly different manner, namely as referring to the propositional attitudes, which are conveyed by a certain modal (*should*, *appropriate* and *willing*, in our study), and that are related to an associated action. The action to which the practical commitments in our experiment are related is the hypothetical making of a payment. The modal lexically conveys norms and the attitudes or practical commitments that the modals convey are stances towards hypothetical payments grounded in the norms that are linguistically encoded. We manipulate the practical commitments by linguistically manipulating the modals that encode these commitments, to gain insights into how the assigned economic value is construed. In the next section, we will spell out the basic linguistic theories of modality which underlie this approach.

2.2 **Linguistic expressions of practical commitments**

CVM questionnaires that directly aim to elicit a value (i.e. that are not designed as a referendum) usually include a WTP elicitation question such as *What is the maximum amount you would be willing to pay (for the proposed intervention)*? This question features the linguistic concept of modality. The notional category ‘modality’ and its linguistic reflex in modal auxiliaries (*must/might/can/should* …) refers to situations which need not be real (Portner, 2009, p. 1), typically in the realm of possibility and necessity. As an example, the sentence
John must be sick.

is a modal sentence, because it does not imply that John is actually sick, but that based on what the speaker knows about John, for example that he has been sneezing all day, they believe that he is sick. With modals, truth cannot be evaluated objectively, as in I am typing this paper now. This sentence is either true or false objectively, but, with modals, truth is relative to non-actual or hypothetical possibilities (Portner, 2009, ch. 2).

Many different categories of modality have been identified by both logicians and linguists (see Fitting and Mendelson, 1998, for the logic approaches and Portner, 2009, for an overview of the linguistic approaches), but there is an agreement that these can be classified into two main classes: `non-root' modals for epistemic modals and `root' modals for the large category encompassing deontics and priority modals.

Epistemic modals (e.g. must/might), convey the speaker's epistemic stance towards truthfulness of the state of affairs described in the modal sentence. In (2) the word must conveys that, given the information that the speaker has, John is sick (Giannakidou and Mari 2016). Truth is thus relative to a body of information, and, for this reason it has been called subjective (Giannakidou and Mari 2021). When using an epistemic modal, the speaker signals that she only has partial information and that she cannot be fully committed to the truthfulness of the state of affairs that is being described. If one utters ´It must be raining now' and not ´It is raining now', it is because one is not entirely certain that it is raining, given the information available in the context (one might have seen a wet umbrella, and this can only lead to the hypothesis that it is raining but not to full knowledge that it is raining). With a bare assertion ´It is raining', the speaker is fully committed to truth, but with an epistemic modal, it presents the state of affairs in which it is raining as a possibility or a necessity, but not as an established

---

1 The class of root modals also comprises ability modals, describing capacities of individuals as in John can play the piano.
fact (Giannakidou and Mari ibid.). In other terms, the state of affairs in which it is raining is presented as non-actual.

In this study, we situate the modal expressions in the realm of root modals (e.g. should/must), with a specific attention to deontic and priority modals.

Root modals (e.g. must/should/ought) are used to describe possibilities and necessities that depend on some state of the world and not on the knowledge that an individual has in a given context. Deontic and priority modals are subcategories of root modals. Deontic modals convey notions of permission and obligation (see (3a) and discussion below) and priority modals convey individuals' preferences (3b).

(3) a. You must pay the taxes every year.

b. I'd rather buy a new pair of shoes.

(3a) conveys that, given what the laws state, the taxes must be paid. The state of affairs in which taxes are paid is relativized to a body of laws. Note that non-actuality is also the hallmark of deontic modals. Indeed, the sentence is compatible with the fact that, in the actual world, the one we live in, one might not pay the taxes and be in a situation of financial evasion.

In (3b), the modal `would rather' describes a preference of the speaker, and the buying of the shoes is relativized to the worlds in which the preferences of the speaker are satisfied. Once again, at the time at which the preference is formed, the shoes have not been bought yet, and the sentence describes a potentiality rather than an actuality.

If the hallmark of modals is non-actuality, then, the category modal encompasses several grammatical categories beyond modal auxiliaries. It is today commonly accepted in the linguistic literature (see Giannakidou and Mari, 2021, for an overview) that the empirical spectrum of the notional category modal encompasses lexical expressions of propositional
attitudes (such as want, believe) and adjectives (such as appropriate) across both the root and the epistemic domains (see Portner, 2009).

This extension to other grammatical categories leads to a revisitation of the strict understanding of the notion of modality as pertaining to possibilities and necessities into a broader view that conceives it as attitudes towards a possible state of affairs (as desired, morally (un)suitable, epistemically conceivable, …), in line with Giannakidou and Mari (2021).

In the question generally used in CV studies What is the maximum amount you would be willing to pay?, modality is grammatically expressed through the use of the conditional mood in would be willing, which refers to the hypothetical (non-actual) nature of the market for the good or service in question. The set of information which serves as the basis of evaluation is called the "modal base" (Kratzer, 1991), and in our case it is the information given about the environmental good and the hypothetical market (the same across our different elicitation questions). In order to define the type of modal we are dealing with, we need another parameter, which Kratzer (ibid.) called the "modal ordering source". This describes how the possibility induced by the modal is realised. For example, in our What is the maximum amount you would be willing to pay?, the modal ordering source is indicated by "willing": It is about a will or desire, what linguists call "bouletics" (Palmer, 2001). The modality expressed by willing is therefore deeply anchored in the subject and their individual goals.

In our experiment the different modal expressions take the role of practical commitments towards the hypothetical payment with the respondent being the subject. In our framework, the term practical commitment replaces the term "root modals" in a perspective that emphasises the behavioural component of propositional attitudes and conveys that the modal describes an attitude towards a possible action on the part of the speaker.

---

2 The expression “hypothetical” as referring to a payment is ambiguous. Under a strict understanding, the payment is presented as hypothetical by the very fact of being syntactically embedded under the modal (as in willing to pay). Under a broader understanding, the payment is hypothetical, as in the experiments, subjects are not actually going to pay.
The gist of our study is to modify the modal expression typically used in stated preference elicitation for contingent valuation to confront respondents with different practical commitments towards the hypothetical payment. In addition to the "classical" bouletic, priority modal question *What is the maximum amount you would be willing to pay?*, we therefore also propose the elicitation questions *What is the maximum amount you think you should pay?* and *What is the maximum amount you think is appropriate to pay?*, which possess particular linguistic properties.

Since *should* and *appropriate* deal with how the world ought to be, both the questions with *should* and *appropriate* can, *prima facie*, be classified in linguistic works in the realm of deontic modality. However, there are important differences between the two. Traditionally, deontic modality has been associated with permission and obligation (Kratzer, 2001; Verstraete, 2005; Von Wright, 1951). In a redefinition of deontic modality, Nuyts et al. (2010) said that it should cover only cases where the moral desirability of a state of affairs is evaluated and under this definition, deontic modality does not involve granting a permission or imposing an obligation. Examples of deontic modality from Nuyts (2008):

(3)  

a. The way you are behaving is intolerable for a civilised person.  

b. This fund raising agency is a very laudable initiative.

In these examples, no permission is granted, nor is any obligation imposed. Rather, a speaker evaluates the state of affairs against some ideal. For expressions of permission and obligation, on the other hand, Nuyts et al. (2010) proposed the term "directive" meaning, which is associated with a clear action plan.³ Example from Nuyts (2008):

³ A related idea was brought forward by Cariani, Kaufmann, and Kaufmann (2013) in their discussion of "deliberative modality", which they characterize as referring to "the thing to do" (Cariani et al., 2013, p. 225), an action-oriented proposition that depends on the agent's information state. Schroeder (2011) contrasts this deliberative modality that establishes a clear agent-action relationship with what he calls, similarly to Nuyts (2008), "evaluative", meaning a broadly normative claim.
For our question with *should*, both the deontic-evaluative as well as the directive-deliberative reading are possible, depending on whether it is interpreted as being more associated with an individual action plan or, inversely, with a general evaluation of an "ideal" behaviour.

With *appropriate*, we explicitly ask participants to perform an evaluation of the appropriateness or normative acceptability of a state of affairs, thus labelling it clearly as an evaluative modal. We do not engage here with whether moral norms pertaining to what is appropriate are absolute or relative, nor are we committed to a set of moral value categories (see Vincent and Koessler, 2019). In our experiments, we intend to investigate what stance speakers adopt with respect to a specific situation of consideration, with reference to the character of which the appropriateness of one or another behaviour is adjudicated.4

To conclude, based on the linguistic notions introduced above we define three categories of practical commitments: evaluative (refers to our elicitation question with *appropriate*), bouletic (*willing*) and normative (*should*). Note that we allow for ambiguity in the interpretation of *should* as more evaluative or more directive.

2.3 *Use and non-use value of environmental goods and services*

With the linguistic base for the choice of the practical commitments in our experiment laid out, we will now turn to our second experimental parameter. It concerns the type of environmental good or service under valuation.

We assume that there may be differences between types of goods with regard to their sensitivity towards modal modification of the valuation question. As the bouletic commitment evokes a desire, it is conceivable that it facilitates the elicitation of value components that are associated with the satisfaction of personal wants and needs, since desires are mainly

4 We thank an anonymous reviewer who suggested this formulation of what attitude the modal "appropriate" is hypothesized to realize in the participant.
associated with personal motivation and "are directly connected only to intentions and mental
states in the pre-intention formation phase of decision making" (Perugini and Bagozzi, 2004,
p.72), excluding factors such as the social implications of an action.

Correspondingly, it is possible that deontic commitments are related more easily to
concerns for entities outside of the self, since moral considerations rest on other-involving
norms or rules (Gillett, 1993).

In order to refine our understanding of these value components, we will employ the
concept of Total Economic Value (TEV) of environmental goods to humans, as first developed
by Pearce and Moran (1994). They make the principal distinction between use values and
non-use values. Use values are connected to the use of natural resources and comprise for
example consumptive values like food or indirect values like climate regulation. Non-use
values are generated without personal use. In general, they encompass values arising from
the knowledge about benefits to nature or to other humans, both in the present and in the
future. Notably, existence value, as conceptualised by Krutilla (1967), is understood to be a
non-use value. It refers to the satisfaction one derives simply from knowing that a natural good
(e.g. an ecosystem) exists. These aspects of value tap into different deontic, moral and
metaphysical considerations. Importantly, use values are anchored in the individual and the
satisfaction of their wants or needs, without taking into account implications for other entities.
Non-use values, on the other hand, are defined as being created without the satisfaction of
personal wants or needs, rather they are assumed to be anchored in the individual’s
perception of the world that surrounds them and the well-being of others.

Nyborg (2000) suggested that the type of good or service in question and notably the
value components it encompasses will trigger a certain mode of evaluation. One is more likely
to adopt a consumer perspective when asked to assess market goods such as coffee mugs.
Conversely, when asked about ethically complex issues usually at stake in CVM studies such
as biodiversity protection it might be more natural for people to take a citizen point of view.
For our experiment we bridge these considerations about the influence of the valuation object on processing mode and about the suitability of modal cues in determining the characteristics of preferences that are expressed. We hypothesise and test in our first experiment that environmental public goods which people cannot directly interact with or derive directly noticeable benefits from, such as the Arctic Ocean and the Amazon rainforest, are subjectively perceived as generating relatively more non-use value than environmental public goods which have directly observable benefits, like clean air and pesticide-free produce.

In our second experiment, we test whether HP differ according to the stance conveyed by the respective modal expression, which could be due to different modes of processing triggered by certain modal expressions, thus leading to diverging HP. Conversely, it is possible that certain modals facilitate the accessibility of specific types of value, which could lead to differences between HP for use and non-use value goods in their susceptibility to modal modification. As a concrete manifestation of the former mechanism, we could imagine, in applying Sagoff’s (1998) theory of language correspondents of preference orderings, that deontic modals are more likely to trigger a citizen processing mode, while bouletic modals are more likely to activate a consumer processing mode. The latter mechanism would imply that certain value dimensions are more or less easily accessible, depending on the use of specific modals. Deontic modals could be assumed to facilitate the expression of non-use values, while bouletic modals could facilitate the expression of use values. Our design does not allow us to differentiate precisely between the two mechanisms\(^5\), and they need not be exclusive. However, we may expect that if there is a difference between the environmental goods more associated with use values and those more associated with non-use values in terms of how sensitive the elicited values are to different modals, then a specific modal might help to access non-use values in particular.

\(^5\) In order to do this, we would need to collect qualitative data and develop a definition of the ways consumer and citizen processing are expressed (including both economic and non-economic measures) and in what ways those value expression might be different from facilitation of the mental accessibility of certain value dimensions (Kenter et al., 2016c).
In our explorative approach, we simply take differentiated elicited values as an indication for underlying preferences that do not correspond to the standard consumer model, but are instead more multifaceted.

3 HYPOTHESES

We assume that the two scenarios which describe directly noticeable benefits to participants, i.e. the air pollution and the pesticides scenarios are perceived as having (relatively) more use value components than the two scenarios which do not describe directly noticeable benefits, but instead focus on benefits to other people and species. Our first hypothesis is thus:

H1: The protection of the Arctic Ocean and the Amazon rainforest are perceived as generating relatively more non-use value than the provision of clean air and the abolishment of harmful pesticides in the UK.

Further, i) appropriate, as a deontic-evaluative modal, evokes a mental state towards what is represented as counting as a moral ideal, whereas the bouletic willing evokes a subjective desire. It seems reasonable that what one is willing to pay is equal to or less than the perceived moral ideal. ii) Appropriate has an evaluative anchoring, which may prompt the consideration of a wider range of interests (analogously to the concept of citizen preferences) than willing, which is anchored in the individual (and would therefore correspond more to consumer preferences). As outlined above, should could be interpreted as a deontic-evaluative or as a deontic-directive modal. We assume that its effect on stated preferences depends on this interpretation. If should is interpreted mainly as deontic-evaluative, it is expected to yield similar amounts of hypothetical payments as appropriate. If should is interpreted mainly as deontic-directive, but grounded in the individual, it is expected to yield similar amounts of hypothetical payments as willing. Consequently, we assume that:
H2: Other things being equal, HP elicited with *appropriate* are higher than those elicited with *willing* for any given environmental good or service.

H3: If H2 holds, then HP elicited with *should* lie in between those elicited with *appropriate* and those elicited with *willing*.

As discussed above, *use* values are defined here to be created based on the satisfaction of the individual’s wants and needs. The bouletic modality indicated by *willing* is, similarly, embedded in a highly personal, desire-oriented context. Given the anchoring in the individual that is present in the generation of *use* values and expressed by bouletic modals, we assume that there is a matching process which may facilitate the consideration of *use* values when the elicitation question is asked with *willing*. Conversely, *non-use* values are grounded in the well-being of entities other than the self. Given the deontic- evaluative anchoring of *appropriate*, we expect that there is a match which may facilitate the expression of *non-use* values when elicited with *appropriate*.

Since *should* is ambiguous between a deontic-directive interpretation and a deontic-evaluative interpretation, we expect a less unequivocal matching effect.

Overall, it is difficult to predict whether there is a difference in the effect of the use of different modals between *use* and *non-use* value goods. In the scenarios that we refer to as *use* value scenarios, value is also generated for entities other than the self – other people as well as other species derive benefits. This *non-use* value aspect of the *use* value goods could be made salient by the use of a deontic- evaluative modal, while it may be more evident from the outset in the case of *non-use* value goods. Conversely, the difference in the perceived associated *non-use* and *use* values may have the opposite effect if the deontic-evaluative modal acts as a reinforcer, rather than a spotlight, on the consideration of *non-use* values. We therefore formulate the following directionless and exploratory hypothesis:
H4: If H2 and H3 hold, then there may be a difference in the strength of the effects between use and non-use value goods.

4 EXPERIMENTS

We conducted two separate experiments. The first one was the value perception study, which we included in order to verify that the scenarios that we termed non-use value scenarios were indeed perceived as having relatively more non-use value components than the scenarios that we termed use value scenarios, and vice versa.

The second experiment was the main study, in which we elicited HP. We separated the experiment into two studies, because we did not want participants’ answers (use vs. non-use value evaluation and HP) to be influenced by the other evaluation, respectively (see for example Schwarz and Strack, 1991).

Demographic information in terms of gender, age, income and household size for the participants of the two studies as well as information on sampling and recruitment is in the supplementary material.

Since the distribution of demographics do not differ significantly between the two samples, we assume that the main results from the value perception study are generalizable also for the sample of the second experiment. We do not pool the two datasets, but instead simply use the first study to corroborate our interpretation of the air pollution and pesticides as use value scenarios and the Arctic Ocean and the Amazon as non-use value scenarios.

All analyses were conducted in R (v.4.0.4), using R Studio (v.1.1.463).

4.1 Study 1

4.1.1 Participants
The value perception study was conducted in English among British citizens via the online platform Prolific (N = 101, 74 female, mean age = 37.1).

4.1.2 Design

This study was designed to categorise environmental goods as use or non-use value goods, depending on which value components were perceived to be most prevalent in them by participants. In total, four environmental scenarios were considered: i) air quality improvement, ii) toxic pesticide abolition, iii) protection of a part of the Arctic Ocean, iv) protection of a part of the Amazon rainforest. Participants first read a description of the problem and then an intervention to address the problem was proposed. In the description of the former two scenarios, the focus was on adverse public health effects of indoor air pollution and the consumption of especially toxic pesticide residue on produce, respectively. In the description of the other two scenarios, the focus was on the well-being of the respective ecosystems and indigenous groups living in them. The following interventions were proposed: i) the installation of air ventilation and filtration systems to improve indoor air quality, ii) the abolition of particularly toxic pesticides and preferential treatment of pesticide-free crops, iii) the creation of a special marine protected area, an “Arctic sanctuary” zone, that would be off limits to all extractive and destructive uses, iv) the establishment of a special protected zone in the Amazon.

Since the former two scenarios entail directly consumable goods – clean air and pesticide-free produce – we assume that they are perceived as having relatively more use value components. As for the latter two scenarios, we assume that they are perceived as having relatively more non-use value components, precisely because they do not involve such directly consumable goods.

After having read a description of the respective problem and the proposed intervention, participants were asked to evaluate the partition of its usefulness for themselves and for the world in general.
This evaluation was performed on an 11-point scale, with 0 representing usefulness only for themselves and none for the rest of the world and 10 representing the opposite. This approach unifies the variables "perceived utility for oneself" and "perceived utility for other entities outside of oneself" on the same scale to allow for a straightforward comparison of people’s value perceptions. It permitted participants to evaluate the intervention even if they thought it had little to no usefulness, since we did not elicit the absolute levels of usefulness on the two dimensions, but only its perceived partition between use and non-use components.

4.1.3 Analysis and results

The mean rating for use value scenarios was 6.44 and for non-use value scenarios 7.67. This difference was significant (t(96) = -4.11, p-value = 0.00). There were no significant differences between the ratings for the two use value scenarios and the two non-use value scenarios, respectively.

There were no significant differences between the answers of men and women to study 1 (details of the analysis are in the supplementary information).

A summary of the results is provided in table 1.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Mean (SD)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air pollution</td>
<td>6.09* (1.67)</td>
<td>21</td>
</tr>
<tr>
<td>Pesticides</td>
<td>6.70*</td>
<td>27</td>
</tr>
</tbody>
</table>

Table 1: Value ratings (study 1)
Values with the same superscript are non-significantly different from each other (p > 0.05)

Both use and non-use value scenarios were perceived as having relatively more non-use than use value, but non-use value scenarios significantly more so than use value scenarios. We retain the expressions use and non-use value scenarios for simplicity.

### 4.2 Study 2

#### 4.2.1 Participants

The study was conducted in English among British citizens (N = 756, 533 female, mean age = 35.8), recruited via the online platform Prolific.

#### 4.2.2 Design

We considered the same four environmental scenarios as in study 1. Participants read a description of the environmental issue and a proposed intervention. In order to elicit hypothetical payments, we added financing mechanisms to all interventions.
Since coercive payment vehicles (i.e. binding mechanisms such as taxes) have been shown to generally produce more realistic results in CVM studies than non-binding vehicles like donations (Schläpfer, 2006), we chose coercive payment vehicles for all scenarios for which it was feasible. Payment vehicles in stated preference studies should also be credible in the context of the environmental good or service at stake (Johnston et al., 2017). The choice of a coercive payment vehicle was hence a more straightforward choice for the two use than for the two non-use value scenarios, as the former were concerned with interventions inside the UK, whereas the latter were related to ecosystem protection outside of the UK. However, we did not want any differences between the effects of language on HP for use and non-use value scenarios to be attributable solely to different payment vehicles. Therefore, we framed one of the payment vehicles of a non-use value scenario as a tax (the "Amazon rainforest tax") and the other one as a voluntary contribution (donation to the "Arctic Ocean sanctuary"), which may be perceived as being more realistic for a good outside of the realm of domestic governance. Since payment vehicles can fundamentally influence people’s apprehension of the scenarios and, for instance, trigger protesting behaviour if they disagree with them (Meyerhoff et al., 2014), we analysed potential differences in the perception of the scenarios by focusing on the share of zero bids as well as importance ratings.

In a between-subjects design, each participant was presented with one scenario, and HP were elicited with one of three questions:

i) What is the maximum amount you would be willing to pay to support this cause? ii) What is the maximum amount you think you should pay to support this cause? iii) What is the maximum amount you think is appropriate to pay to support this cause?

In a statement preceding the elicitation question, it was specified that HP were to be indicated as yearly payments on the household level, for a period of 10 years. HP can be elicited via different question modes, such as open-ended formats, single or multiple-bounded dichotomous choice questions, or payment cards. While close-ended format has been
recommended by Arrow et al. (1993), it has been found to be subject to specific biases like yea-saying (Brown et al., 1996). We consider that for our purpose, a payment card style elicitation mechanism, where people choose from a range of possible amounts (see for example OECD, 2018), is adequate. This minimises hypothetical bias induced by yea-saying in dichotomous choice formats which tends to produce inflated results, and we consider that the anchoring that occurs by presenting participants with a range of answers does not corrupt our goal, which is to capture evaluative tendencies, not to produce reliable CVM estimates for policy use.

Participants were subsequently asked to indicate how important the respective environmental good or service was to them personally on a 5-point Likert scale ranging from “not important at all” (1) to “very important” (5).

Finally, the socio-demographic variables age, gender, household size and annual household income category were elicited.

4.2.3 Analysis

In our study there was a very low ratio of zero bids (under 6% in all scenarios).\(^6\) Therefore, in the data analysis, we restricted the sample to positive answers (positive HP). We transformed the positive bids using the formula \(\log(\text{positive HP} + 1)\) (LNHP) because of the large positive skew of the data, which is typical for CVM estimates. This improved the normality of the data (as can be seen in the estimated Kernel densities in figures 1 and 2 in the supplementary material).

To investigate the effect of language on HP for the different goods, we conducted simple and multiple ordinary least squares (OLS) regression analyses for each scenario.

\(^6\) Details on zero bids are in the supplementary information.
separately. In the simple regressions, only the modal used in the elicitation question was
included as an explanatory variable. We transformed the language variable into 3 dichotomous
variables and included appropriate and should, making willing the base category. In the
multiple regressions, we added the importance measures as well as the socio-demographic
variables age, gender household size and annual household income category.

Finally, we performed Welch independent sample t-tests between the LNHP elicited
with willing and with one of the other two elicitation questions, respectively, and reported 95%
confidence intervals (CIs) of the difference between means. This procedure is akin to the
simple regressions, and gives us a more refined idea of the strength of the effects.

In order to detect possible differences in protesting behaviour between the scenarios,
we compared the ratio of zero bids by means of Chi-squared tests.

4.2.4 Results

Table 2: Means of positive HP and LNPH

<table>
<thead>
<tr>
<th></th>
<th>Positive HP</th>
<th>LNHP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(SD)</td>
<td>(SD)</td>
</tr>
<tr>
<td>Air pollution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>willing</td>
<td>165.15</td>
<td>4.62</td>
</tr>
<tr>
<td></td>
<td>(149.98)</td>
<td>(1.15)</td>
</tr>
<tr>
<td>should</td>
<td>156.85</td>
<td>4.65</td>
</tr>
<tr>
<td></td>
<td>(134.05)</td>
<td>(1.02)</td>
</tr>
<tr>
<td>appropriate</td>
<td>280.27</td>
<td>5.15</td>
</tr>
<tr>
<td></td>
<td>(269.75)</td>
<td>(1.06)</td>
</tr>
<tr>
<td>Pesticides</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>willing</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>Arctic Ocean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>willing</td>
<td>123.49</td>
<td>4.07</td>
</tr>
<tr>
<td>should</td>
<td>169.66</td>
<td>4.39</td>
</tr>
<tr>
<td>appropriate</td>
<td>152.22</td>
<td>4.49</td>
</tr>
<tr>
<td>Amazon rainforest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>willing</td>
<td>156.58</td>
<td>4.55</td>
</tr>
<tr>
<td>should</td>
<td>123.32</td>
<td>4.24</td>
</tr>
<tr>
<td>appropriate</td>
<td>230.89</td>
<td>4.59</td>
</tr>
</tbody>
</table>

Table 2 provides a summary of the means of positive HP and LNHP. The results of the regression analyses for the individual scenarios are summarised in the supplementary material. We do not observe a significant effect of *should* on the LNHP, neither in the simple
nor in the multiple regressions, meaning that it does not lead to amounts that are significantly different from those elicited with *willing*.

*Appropriate*, on the other hand, has a significant positive effect in the two *use* value scenarios – HP were higher when they were elicited with *appropriate* than when they were elicited with *willing*. For the two *non-use* value scenarios, *appropriate* did not have a significant effect.

The importance ratings had a significant positive effect on LNHP in all but the air pollution scenario.

Table 3 summarises the results of Welch independent t-tests between LNHP with *willing* and the other elicitation questions, respectively. We can see that *should* did not significantly change the elicited LNHP for any of the environmental goods as compared to *willing*. *Appropriate* had an unequivocal positive effect for *use* value goods, with similarly wide CIs, the furthest from zero for the pesticides scenario. It did not have a significant effect for *non-use* value goods.

<table>
<thead>
<tr>
<th>Practical commitment</th>
<th>Use value goods</th>
<th>Non-use value goods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Air pollution</td>
<td>Pesticides</td>
</tr>
<tr>
<td>Willing</td>
<td>4.62</td>
<td>4.53</td>
</tr>
<tr>
<td>Should</td>
<td>[-0.39,0.45]</td>
<td>[-0.49,0.45]</td>
</tr>
<tr>
<td>Appropriate</td>
<td><strong>[0.11,0.95]</strong></td>
<td><strong>[0.37,1.24]</strong></td>
</tr>
</tbody>
</table>

Table 3: LNHP with "willing" as baseline, pairwise comparisons with "should" and "appropriate"
There were no significant differences in the ratios of zero bids between the scenarios, \( \chi^2 = 1.04, \text{df} = 3, \text{p-value} = 0.79 \). Therefore, we do not assume that protesting behaviour and differences therein between the scenarios are of particular concern for our study.

The means of the importance ratings ranged from 3.42 for the pesticide scenario to 4.14 for the Amazon rainforest scenario (measured on a 1-5 scale, 5 corresponding to "very important"). A Kruskal-Wallis test indicated that importance ratings differed between the scenarios \( H(3) = 69.53, \text{p-value} = 0.00 \). Subsequent pairwise comparisons showed that both non-use value scenarios were perceived as having significantly more personal importance for the participants in our study than both use value scenarios. There were no significant differences in importance ratings between the two use and between the two non-use value scenarios, respectively (see supplementary information).

Pairwise comparisons between men's and women's LNHP revealed only one significant difference: For the Arctic Ocean scenario, when values were elicited with *should*, women’s answers were significantly higher than men's, \( t(39) = 2.09, \text{p-value} = 0.04 \). In additional regressions of LNHP on gender, elicitation questions and their interactions, gender did not have a significant effect (see supplementary information).

6 DISCUSSION

6.1 Sample characteristics

There was an overrepresentation of women in our samples relative to the UK population, which must be considered when making inferences from the results. However, response patterns did not significantly differ between women and men, as indicated by the insignificant
effects of gender and gender/language interactions in all specified models (see supplementary information).

In our sample, the percentages of positive HP were very high (over 94% in all scenarios), as compared to the studies conducted by Kahneman and Ritov (1994), where positive bids ranged from 24% to 79%, and Ryan and Spash (2011), where positive bids accounted for 42%. We assume that this is due to our sample which we recruited using an online recruitment tool. Participants recruited via such platforms tend to be more experienced in study participation (Palan & Schitter, 2018) and might therefore be less sceptical towards the idea of a hypothetical payment for natural goods and services.

6.2 Discussion of results

In light of our results, we find support for the hypothesis that the scenarios we termed “non-use value goods” are perceived as generating relatively more non-use value than those that we termed “use value goods” (H1). However, all of the scenarios were perceived as generating more non-use than use value, so the value perception was generally skewed towards non-use values, albeit with significantly different shares for the two types of scenarios.

With regard to H2, we find evidence that, other things being equal, HP elicited with appropriate are higher than those elicited with willing, but the difference was significant only for use value, and not for non-use value goods.

H3 concerned the elicitation question with should - we expected that, conditionally on H1 holding, HP elicited with should would lie in between those elicited with willing and those elicited with appropriate. This hypothesis was not corroborated by the results – values elicited with should did not differ significantly from those elicited with willing for any of the scenarios.

Linguistic theory proved to be a valuable source of information for the characterization of the value articulating institutions represented by the scenarios. The bouletic practical commitment established with willing led to lower elicited HP than the deontic-evaluative
appropriate. We assumed that willing prompts a narrower practical commitment, focused on individual desires, while appropriate demands a comparison of a state of affairs against what is considered by the speakers as a moral standard. Should, as a deontic modal that is associated with some degree of ambiguity between a directive and an evaluative interpretation, produced HP that were not significantly different from those elicited with willing, hinting at an interpretation of should that is more directive than evaluative.

In H4, we considered the possibility that a modal that is explicitly not grounded in the self, i.e. appropriate, could act as a spotlight, making non-use value components more salient in contexts where they may be less obvious upon first glance. Appropriate would then be expected to highlight non-use values in use value scenarios, thereby increasing elicited values. In the non-use value scenarios, the overall focus is already on non-use values, so the spotlight effect of appropriate may be attenuated. Another possibility is that appropriate acts as a reinforcer for non-use values. In that case, the relatively higher prevalence of non-use values would be expected to lead to a stronger effect of appropriate on elicited values for non-use value scenarios, and a weaker effect for use value scenarios. Our results hint at an interpretation of the modal as a spotlight, since we observed a significant effect of appropriate on elicited values only for use, but not for non-use value goods.

Both non-use value scenarios were rated as having more personal importance for participants than the two use value scenarios and for non-use value scenarios importance had a significant effect on HP in the regression analyses, which suggests that moral considerations likely played a crucial role in the valuation process.

We want to underline the possibility that confounding factors related to study limitations may have played a role in producing the differentiated results for use and non-use value goods. In particular, the two non-use value scenarios in our experiment were set outside of the UK’s national jurisdiction, while the use value scenarios were set inside. Participants may have moral considerations about environmental issues in their own country that reflect
differently in economic valuation as compared to things that are further away. A possible reason for this could be a lack of perceived behavioural control for non-use value scenarios as compared to use value scenarios, associated with geographical and psychological distance. Future research could look at how perceived behavioural control, which was shown by Ryan and Spash (2011) to correlate with WTP measures, relates to different practical commitments in environmental values.

Another possibility is that appropriate, contrarily to our expectations, actually matches less with considerations about ideals than with considerations about individual well-being. This seems somewhat counter-intuitive, but a related effect has been observed by Kenter et al. (2016b). They compared individual WTP for marine ecosystems elicited through a survey to deliberated group-based "fair prices" and found that the latter were significantly better explained by measures of subjective well-being associated with the good in question than the former. Thus, it seems possible that personal utility or well-being is the main concern of respondents when the question concerns fairness or appropriateness.

Elicitation questions that syntactically differ from the ones we tested, such as "fair prices" to pay for society, may entail other effects on HP. In the current study, we focused on syntactically similar questions that only differ in the type of modal, to be able to trace back observed effects to one linguistic phenomenon. Future research could test a larger variety of questions and notably include linguistic indicators of society-level aggregation, to account for the type of questions that are typically used in innovative stated preferences studies.

6.3 Conclusions

Our study contributes to a better understanding on how values grounding preference are articulated and reflected in their linguistic implementations.

We presented the idea that the question *How much would you be willing to pay?* that is standardly used in CVM studies implies an important implicit presupposition, namely that
underlying economic preferences can best be accessed by linguistically establishing a bouletic or desire-oriented practical commitment towards a hypothetical payment. We suggested that the choice of elicitation question wording may give rise to different moral stances, thereby influencing directly the elicited values.

We found that hypothetical payments elicited with the question *What is the maximum amount you think is appropriate to pay?* were higher than those elicited with *What is the maximum amount you would be willing to pay?* for environmental goods more associated with use values, while there was no significant effect for environmental goods more associated with non-use values. We did not find significant differences between the amounts elicited with willing and those elicited with *What is the maximum amount you think you should pay?*.

*Willing* prompts a narrow practical commitment, concerned with individual wants, while *appropriate* is associated with a deontic-evaluative practical commitment, implying a general moral standard. *Should* is ambiguous regarding its reading as deontic-directive (i.e. referring to an individual action plan) or deontic-evaluative (i.e. referring to a more general evaluation of a moral ideal). We take our results as an indication of an interpretation of *should* in the given context as more deontic-directive and therefore anchored in the individual. With regard to the contrasted results for *use* and *non-use* value goods we noted that the deontic-evaluative *appropriate* may have acted as a spotlight, making *non-use* value components more salient and decisive in the *use* value scenarios, while in the non-use value scenarios they are more likely to be of primary concern regardless of the elicitation modal. We discussed possible confounding factors that differentiate *use* and *non-use* value scenarios besides the *use* and *non-use* value composition.

A practical conclusion that can be drawn from our findings is the importance of language sensitivity in environmental valuation. Especially now that researchers are starting to employ alternative elicitation formats which target values that go beyond individual WTP (Szabó, 2011; Kenter, 2016; Kenter, et al., 2016), it is important to have a good understanding
of what these represent, and insights from the semantic analysis of modality are a promising starting point. In taking into account the informational basis for evaluation, or the modal base, as well as the modal ordering source, i.e. how the individual relates to the payment, we are able to understand nuanced differences, for example between seemingly similar deontic modals such as the questions with *appropriate* and *should*, which in our experiment have been shown to produce differentiated results.

Our study demonstrates that it is (relatively) easy to highlight the variety and also the variability of moral stances and of their associated practical commitments that can emerge through preference elicitation mechanisms in their role as value articulating institutions. By that we do not mean to say that purely monetary environmental valuation suffices to capture the variety and variability of environmental values (which is the subject of animated debates in the field, see for example Gómez-Baggethun and Ruiz-Pérez, 2011; Farley, 2012, Vatn and Bromley, 1994), but rather that even with purely monetary valuation, we can detect indications for differentiation created by such practical commitments.

Without a conception of the connection between the linguistic substrate of environmental valuation based on hypothetical scenarios and the axiological dimensions that depend on it, one could be misled in the interpretation of the elicited values. We would like to conclude this article by insisting on two broad implications of our study with respect to our understanding of the nature of environmental valuation.

The first one is that values elicited with stated preference methods may not be stable and final and, thus, different from individual preferences as they are defined in mainstream economics (Carson, 2012). This is the case not only because they are morally dependent, but also because they are dependent on the contextual actualisation of a particular moral stance at the moment of the valuation procedure, a value articulating institution. In that way, preferences, as revealed through a hypothetical monetary scale, are rather to be conceived of as indirect measures of the extent to which the respondent engages in a certain practical
commitment. This hints at a substantial difference from traditional stated preference studies, where valuation is forced onto a monetary scale, which constrains respondents to a default practical commitment. The applicability of the conventional logic of the economic sphere of reasoning is therefore put into question for environmental goods and services (see also: Farrell, 2007; Light, 2002; O’Connor, 2006).

Second, we note that our study allowed us to question the degree of adequacy, through the stated preference procedures, between the practical commitment solicited from the participants and the presumed nature of the environmental change described in the scenario. Our results point to differences in how different environmental goods and services are valued in terms of the responsiveness to linguistically activated practical commitments. It could also be the case that asking to evaluate an environmental scenario, from the angle of a particular evaluative term or phrase, does not simply correlate with an objective feature (use or non-use aspects) of the considered environmental entity but influences its perception. Language encodes and shapes the construction and expression of values. This has deep implications on the scope of stated preferences methods, revealing, in particular, their sensitivity to linguistic framing. Our present study, although standardly eliciting monetary stated preferences, sheds light on the issue of the sensitivity of expressed values to the linguistic framing of the elicitation methods. This raises the question of how, if at all, it is possible to elicit supposedly underlying values in an undistorted way.

ACKNOWLEDGEMENTS

We want to thank two anonymous reviewers whose comments helped us to improve this paper. We gratefully acknowledge funding from the French National Research Agency (grant number ANR-17-EURE-0017 FrontCog).


Table 4: Sample demographics studies 1 and 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>Study 1</th>
<th>Study 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male: 33</td>
<td>Male: 218</td>
</tr>
<tr>
<td></td>
<td>Female: 66</td>
<td>Female: 533</td>
</tr>
<tr>
<td></td>
<td>Other/NA: 5</td>
<td>Other/NA: 5</td>
</tr>
<tr>
<td>Age</td>
<td>Mean: 37.1</td>
<td>Mean: 35.82</td>
</tr>
<tr>
<td></td>
<td>SD: 13.45</td>
<td>SD: 13.78</td>
</tr>
<tr>
<td>Income category</td>
<td>£13,300 - £26,800: 18</td>
<td>£13,300 - £26,800: 170</td>
</tr>
<tr>
<td></td>
<td>£26,801 - £35,700: 24</td>
<td>£26,801 - £35,700: 146</td>
</tr>
<tr>
<td></td>
<td>£35,701 - £54,000: 26</td>
<td>£35,701 - £54,000: 188</td>
</tr>
<tr>
<td></td>
<td>&gt;£54,000: 25</td>
<td>&gt;£54,000: 172</td>
</tr>
<tr>
<td>Household size</td>
<td>1: 8</td>
<td>1: 85</td>
</tr>
<tr>
<td></td>
<td>2: 31</td>
<td>2: 240</td>
</tr>
<tr>
<td></td>
<td>3: 27</td>
<td>3: 179</td>
</tr>
<tr>
<td></td>
<td>4: 18</td>
<td>4: 174</td>
</tr>
<tr>
<td></td>
<td>5+: 15</td>
<td>5+: 78</td>
</tr>
</tbody>
</table>

Categorial variables were compared by means of Chi-squared tests, continuous variables by means of Welch independent sample t-tests. For all comparisons p > 0.05.
Table 5: Simple and multiple OLS regressions, DV: LNHP, robust standard errors in parentheses

<table>
<thead>
<tr>
<th></th>
<th>Air pollution</th>
<th>Pesticides</th>
<th>Arctic Ocean</th>
<th>Amazon rainforest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Simple</td>
<td>Multiple</td>
<td>Simple</td>
<td>Multiple</td>
</tr>
<tr>
<td>Should</td>
<td>0.03</td>
<td>0.03</td>
<td>-0.02</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>(0.24)</td>
<td>(0.22)</td>
<td>(0.23)</td>
<td>(0.23)</td>
</tr>
<tr>
<td>Appropriateness</td>
<td>0.53</td>
<td>0.49</td>
<td>0.81</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
<td>(0.21)</td>
<td>(0.21)</td>
<td>(0.22)</td>
</tr>
<tr>
<td>Importance</td>
<td>0.03</td>
<td>0.29</td>
<td>0.34</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.09)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.02</td>
<td>-0.00</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Gender: male</td>
<td>-0.08</td>
<td>0.12</td>
<td>-0.31</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
<td>(0.19)</td>
<td>(0.20)</td>
<td>(0.20)</td>
</tr>
<tr>
<td>Gender: other</td>
<td>-0.14</td>
<td>0.93</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
<td>(0.32)</td>
<td>(0.41)</td>
<td></td>
</tr>
<tr>
<td>Household size</td>
<td>0.11</td>
<td>0.06</td>
<td>0.09</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>Income category</td>
<td>0.04</td>
<td>-0.04</td>
<td>-0.07</td>
<td>0.02</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.05)</td>
<td>(0.05)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Constant</th>
<th>4.62</th>
<th>4.23</th>
<th>4.53</th>
<th>3.46</th>
<th>4.07</th>
<th>3.14</th>
<th>4.55</th>
<th>3.52</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.16)</td>
<td>(0.55)</td>
<td>(0.16)</td>
<td>(0.55)</td>
<td>(0.16)</td>
<td>(0.51)</td>
<td>(0.14)</td>
<td>(0.65)</td>
</tr>
<tr>
<td></td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>161</th>
<th>160</th>
<th>160</th>
<th>160</th>
<th>204</th>
<th>204</th>
<th>190</th>
<th>190</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>R²</th>
<th>0.05</th>
<th>0.09</th>
<th>0.09</th>
<th>0.19</th>
<th>0.02</th>
<th>0.16</th>
<th>0.02</th>
<th>0.05</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Adjusted R²</th>
<th>0.04</th>
<th>0.04</th>
<th>0.08</th>
<th>0.15</th>
<th>0.01</th>
<th>0.13</th>
<th>0.01</th>
<th>0.01</th>
</tr>
</thead>
</table>

*Note:* *p<0.1; **p<0.05; ***p<0.01

Figure 1: Kernel density estimation – positive HP
Participants were recruited in 2021 using the online panel Prolific. We posted a task under the neutral name “household study” that required participants to complete a survey hosted by Qualtrics. Participants were paid £0.63. The study was open to panelists who were British citizens and residing in the UK. Participants selected themselves into the study group.

The share of female participants to study 1 was 66%. Compared to the UK population, there was a higher percentage of women in our sample (percentage of women in the UK: 51%\(^7\)). The overrepresentation of women needs to be taken into account when considering the applicability of our results to the general population. In order to gauge whether men and women responded differently, we compared their answers. Table 1b below shows averages of use and non-use value ratings for men and women separately.


(Access: 16.08.2022)
Table 1b: Value ratings by gender

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Men</th>
<th></th>
<th>Women</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD) N</td>
<td></td>
<td>Mean (SD) N</td>
<td></td>
</tr>
<tr>
<td>Air pollution</td>
<td>6.43a (1.90)</td>
<td>7</td>
<td>6.00a (1.56)</td>
<td>15</td>
</tr>
<tr>
<td>Pesticides</td>
<td>6.50a (1.24)</td>
<td>12</td>
<td>6.87a (1.30)</td>
<td>15</td>
</tr>
<tr>
<td>Average use scenarios</td>
<td>6.47 (1.47)</td>
<td>19</td>
<td>6.43 (1.48)</td>
<td>30</td>
</tr>
<tr>
<td>Arctic Ocean</td>
<td>7.50b (1.31)</td>
<td>8</td>
<td>7.53b (1.23)</td>
<td>17</td>
</tr>
<tr>
<td>Amazon rainforest</td>
<td>8.67b (1.21)</td>
<td>6</td>
<td>7.63b (1.98)</td>
<td>19</td>
</tr>
<tr>
<td>Average non-use scenarios</td>
<td>8.00 (1.36)</td>
<td>14</td>
<td>7.58 (1.65)</td>
<td>36</td>
</tr>
</tbody>
</table>

Values with the same superscript are non-significantly different from each other (p > 0.05).

Pairwise t-tests do not detect significant differences between men and women’s value ratings for any of the scenarios with all p-values > 0.05.

Sample study 2
Participants were recruited in 2021 using the online panel Prolific. We posted a task under the neutral name “household study” that required participants to complete a survey hosted by Qualtrics. Participants were paid £0.63. The study was open to panelists who were British citizens and residing in the UK. Participants selected themselves into the study group. Panelists who had participated in study 1 were excluded from participation to study 2.

The share of female participants to study 2 was 71%. Compared to the UK population, there was a higher percentage of women in our sample.

We compared answers between men and women by different means (five participants who answered “other / prefer not to specify” in the gender question were excluded for these analyses):

- We fitted an additional regression model with only the elicitation question, gender and their interactions as independent variables
- We conducted pairwise comparisons of men’s and women’s answers

Table 3b: OLS regressions with elicitation question, gender and interactions, DV: LNHP, robust standard errors in parentheses

<table>
<thead>
<tr>
<th></th>
<th>Air pollution</th>
<th>Pesticides</th>
<th>Arctic Ocean</th>
<th>Amazon rainforest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Should</td>
<td>0.18</td>
<td>-0.01</td>
<td>0.59**</td>
<td>-0.30</td>
</tr>
<tr>
<td></td>
<td>(0.23)</td>
<td>(0.29)</td>
<td>(0.26)</td>
<td>(0.24)</td>
</tr>
<tr>
<td>Appropriate</td>
<td>0.57**</td>
<td>0.73***</td>
<td>0.43**</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>(0.24)</td>
<td>(0.29)</td>
<td>(0.24)</td>
<td>(0.26)</td>
</tr>
<tr>
<td>Gender: male</td>
<td>0.24</td>
<td>0.02</td>
<td>0.02</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>(0.45)</td>
<td>(0.33)</td>
<td>(0.37)</td>
<td>(0.35)</td>
</tr>
</tbody>
</table>
Neither the gender coefficients nor the interaction terms between gender and the elicitation question are significant in this analysis.

Tables 2b below show the means of positive HP and log-normalized HP for men and women, respectively.

**Table 2b: Means of LNHP, men and women**

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th></th>
<th>Women</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LNHP</td>
<td>N</td>
<td>LNHP</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>(SD)</td>
<td></td>
<td>(SD)</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** *p<0.1; **p<0.05; ***p<0.01
<table>
<thead>
<tr>
<th>Scenario</th>
<th>Type</th>
<th>Value1</th>
<th>Value2</th>
<th>Value3</th>
<th>Value4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air pollution</td>
<td>willing</td>
<td>4.82</td>
<td>10</td>
<td>4.58</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.37)</td>
<td></td>
<td>(1.11)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>should</td>
<td>4.25</td>
<td>11</td>
<td>4.76</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.23)</td>
<td></td>
<td>(0.94)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>appropriate</td>
<td>5.16</td>
<td>15</td>
<td>5.15</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.11)</td>
<td></td>
<td>(1.06)</td>
<td></td>
</tr>
<tr>
<td>Pesticides</td>
<td>willing</td>
<td>4.54</td>
<td>23</td>
<td>4.52</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.26)</td>
<td></td>
<td>(1.24)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>should</td>
<td>4.49</td>
<td>18</td>
<td>4.51</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.03)</td>
<td></td>
<td>(1.24)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>appropriate</td>
<td>5.48</td>
<td>19</td>
<td>5.26</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.77)</td>
<td></td>
<td>(1.26)</td>
<td></td>
</tr>
<tr>
<td>Artic Ocean</td>
<td>willing</td>
<td>4.04</td>
<td>21</td>
<td>4.02</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.37)</td>
<td></td>
<td>(1.26)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>should</td>
<td>3.90</td>
<td>24</td>
<td>4.61</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.40)</td>
<td></td>
<td>(1.23)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>appropriate</td>
<td>4.28</td>
<td>18</td>
<td>4.55</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.49)</td>
<td></td>
<td>(1.04)</td>
<td></td>
</tr>
<tr>
<td>Amazon rainforest</td>
<td>willing</td>
<td>4.61</td>
<td>21</td>
<td>4.51</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.26)</td>
<td></td>
<td>(1.11)</td>
<td></td>
</tr>
</tbody>
</table>
We compared men's and women's log-normalized HP by means of Welch independent sample t-tests and found one significantly different mean:

In the Arctic Ocean scenario, when the elicitation question was "What is the maximum amount you think you should pay?", log-normalized HP of women were significantly higher than for men, $t(39) = 2.09$, $p$-value = 0.04.

Table 5: Means and SD of importance ratings

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Importance</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air pollution</td>
<td>3.61&lt;sup&gt;a&lt;/sup&gt;</td>
<td>(0.89)</td>
</tr>
<tr>
<td>Pesticides</td>
<td>3.42&lt;sup&gt;a&lt;/sup&gt;</td>
<td>(0.98)</td>
</tr>
<tr>
<td>Arctic Ocean</td>
<td>3.99&lt;sup&gt;b&lt;/sup&gt;</td>
<td>(0.89)</td>
</tr>
<tr>
<td>Amazon rainforest</td>
<td>4.14&lt;sup&gt;b&lt;/sup&gt;</td>
<td>(0.83)</td>
</tr>
</tbody>
</table>

Values with the same superscript are non-significantly different from each other ($p > 0.05$).
3. Zero bids

The shares of zero bids for the different scenarios are given in table 6 below.

Table 6: Shares of zero bids

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Share of zero bids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air pollution</td>
<td>0.036</td>
</tr>
<tr>
<td>Pesticides</td>
<td>0.059</td>
</tr>
<tr>
<td>Arctic Ocean</td>
<td>0.047</td>
</tr>
<tr>
<td>Amazon rainforest</td>
<td>0.046</td>
</tr>
</tbody>
</table>

A Chi-squared test between the four scenarios indicated that there were no significant differences in terms of the share of zero bids, Chi-squared = 1.04, df = 3, p-value = 0.79.