Journal of Health Economics 85 (2022) 102674

Contents lists available at ScienceDirect

Journal of Health Economics

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

A discrete choice model implementing gist-based categorization of alternatives, with applications to patient preferences for cancer screening and treatment

J. Swait *, E.W. de Bekker-Grob

Erasmus School of Health Policy and Management and Erasmus Choice Modeling Centre, Erasmus University Rotterdam, P.O. Box 1738, Rotterdam 3000 DR, the Netherlands

ARTICLE INFO

JEL Codes:
C25
D83
D91

Keywords:
Categorization
Multi-stage choice
Gist-based judgment
Antecedent volition
Choice modeling
Discrete choice experiment

ABSTRACT

The rational microeconomic decision model is hard-coded into usual econometric specifications such as the Multinomial Logit and Probit models, *inter alia*. There is a very tight link between utility maximization and the apparatus of welfare theory that underlies economic policy analysis, which creates a tension around the possibility of representing other decision rules. We propose a less restrictive model of choice, built on the concept of gist-based categorization judgments that are assumed to precede (thus, condition) the maximization-driven selection process in decision making. This categorization facilitates decision making by allowing adoption of certain simpler decision rules under appropriate conditions, the drivers of which are endogenously determined. We demonstrate that the proposed model provides better fit than traditional choice models, using cancer screening and treatment choice data from two discrete choice experiments. In addition, we show that the model provides a deeper, more nuanced and insightful perspective on (healthcare) decision making.

Introduction

The behavioral underpinnings of the commonly used econometric models of choice assume that (1) decision makers are fully rational, (2) make use of all information presented to them to value the options, (3) have well-formed and retrievable preferences, and (4) choose the best valued alternative among those available. (For an elucidation on these topics from an econometric perspective, please see Ben-Akiva and Lerman, 1985).These assumptions are usually encapsulated in the characterization of decision makers (e.g., patients, healthcare professionals, policy makers) as fully informed utility maximizers with compensatory preferences underlying value formation, in resonance with traditional theories of choice in both economics and psychology (e.g., Lancaster, 1966, Luce, 1959). This characterization lies at the heart of the workhorse choice model forms such as the Multinomial Logit (MNL), Nested MNL, and Multinomial Probit (MNP); in turn, this implies that basically all work in discrete choice experiments (DCEs) explicitly endorses this paradigm of and perspective on decision making.

Health preference/choice data has also been analyzed using these model forms embodying the utility maximizing decision rule (e.g., see reviews in Soekhai et al., 2019, Clark et al. (2014) and de Bekker-Grob et al. (2012)). The imposition of this strong rationality assumption has been questioned in the literature of applied economics (e.g., McFadden, 2006), but it seems to us that this assumption

* Correspondence author.
E-mail address: swaitjr@eshpm.eur.nl (J. Swait).

https://doi.org/10.1016/j.jhealeco.2022.102674
Received 9 September 2021; Received in revised form 28 March 2022; Accepted 16 August 2022
Available online 17 August 2022

may be even more questionable concerning healthcare decision making. Specifically, patient decision making around health conditions, treatments and diagnostic tests can be fraught with the likelihood of very negative outcomes that generate strong emotions, have associated time pressures, give rise to social group tensions, exhibit information source credibility issues, and present multiple other factors that put everyone involved in distressed circumstances. Thus, the tenability of the utility maximization representation is perhaps lower in healthcare than in other fields of applied economics such as marketing, transportation, and resource and recreation economics. Representing the outcomes of such decisions, say in circumstances such as cancer diagnosis and treatment, suggests a need to recognize these distressing contexts in which patients are asked to make them.

In health and health economics, there has been significant use of DCEs as tools for characterizing preferences and decision making in multiple domains and decision types (see, e.g., reviews in Soekhai et al. 2019, Clark et al., 2014 and de Bekker-Grob et al. 2012), among them those surrounding cancer (specifically, consider studies such as Collacott et al. 2021). The vast majority of DCEs pose to respondents two or more options, described by multiple attributes or properties, among which they are asked to choose which one they most prefer. For example, there may be two or more screening options available to promote early detection of colorectal cancer. One option may be more invasive but more diagnostic than the other, which in its turn is less invasive but also less accurate in its detection capability, thus inviting a decision by the patient that analysts surmise is based on risk/benefit tradeoffs (see, e.g., Hol et al., 2010, van Dam et al. 2010).

The rational decision maker assumption of economics has longstanding empirical and theoretical work suggesting alternative processes for decision making in general (e.g., Simon 1955, Kahneman and Tversky 1979, Thaler 1985, Payne et al., 1993, Rubenstein 1998, Luce et al., 2001), and health decision making in particular (e.g., Taylor 2000, McCaughey and Bruning 2010, Marewski and Gigerenzer 2012); and qualitative empirical work pointing to quite different depictions of how patients make health-related decisions (e.g., Sutkowi-Hemstreet et al., 2015, Lie et al., 2019, Taylor et al., 2021). In psychology and consumer behavior, and more recently in behavioral economics, this tension between the regnant paradigm of utility maximization (again, inherited from both microeconomics and psychology) and a significant body of accumulated evidence that suggests decisions can and are made in very different ways by different people and in different situations, has led to development of a heuristics and biases (H&B) literature (e.g., Payne et al., 1993), and another of a fast-and-frugal (F&F) orientation (e.g., Gigerenzer 1996; Gigerenzer and Goldstein 1996, Gigerenzer and Gaissmaier 2011; in healthcare, Marewski and Gigerenzer 2012). From an econometric perspective, these literatures are not focused on deriving empirically tractable models of choice that can be tested against traditional analysis tools such as the MNL and its variants on large samples of decision makers for the purposes of conducting economic policy analysis. This lacuna has led to some skepticism on the part of economic practitioners about the usefulness and practicality of incorporating alternative processes of decision making in models.

This latter point is of crucial importance in developing practical and practicable tools for the health econometrician. To illustrate this point, consider the Elimination-by-Aspects (EBA, Tversky 1972) decision rule, which assumes knowledge of aspects (attributes); their rejection/acceptability thresholds; and most importantly, for our purposes, the order in which these aspects are considered. The econometrician's (not the decision maker's!) effort to compute the likelihood of choice under EBA is primarily driven by the need to compute choice probabilities for every possible sequence of attribute consideration, because this order is assumed unknown since we observe only outcomes. The probability of choice in an EBA decision rule would then be given by $P(j) = \sum_{l=1}^{L} P(j|l)Q(l)$, where $P(j|l)$ is the survival probability of alternative j given sequence $l = 1,\ldots,L$ of attribute consideration, and $Q(l)$ is the likelihood of sequence l being employed by the decision maker. With K attributes, the inference machinery built around this description implies a relatively onerous set of calculations to predict outcomes, since L is given by the factorial of K: to make this concrete, consider $K = 2$, for which $L = 2$; $K = 4$, $L = 24$; $K = 8$, $L = 40{,}320$; $K = 16$, $L = 20{,}922{,}789{,}888{,}000$. Since most health DCEs have K in the range of 4 to 12 aspects, we can conclude that implementation of the EBA would be an onerous (even prohibitive) computational effort, in the setting of parameter inference making. In contrast, implementation of the utility maximization (UM) decision rule requires computational effort that is linear in K.

This contrast of the EBA with the UM decision process illustrates how the econometrician's view of a decision process differs broadly (and, perhaps, markedly) from that of the psychologist. The EBA model, while intuitively appealing in a range of decision domains, presents a much greater computational challenge for the econometrician because of the probabilities Q and the need to repeat the calculations of the conditional probabilities. Other decision rules, such as those mentioned above from the H&B and F&F literatures, all imply somewhat or very burdensome calculation procedures for the econometrician, even if the decision rules are relatively straightforward from the decision maker's perspective. The lack of information on the analyst's part implies the need for a compensating probabilistic complexity, with a consequent level of complexity for parameter inference making.

In addition, and central to our motivation, the focus of applied economics is almost always on the evaluation of policy impacts in a decision context, using a welfare theoretic apparatus to conduct these analyses (a seminal reference is Small and Rosen 1981; see also Ben-Akiva and Lerman 1985). At the core of this apparatus is a crucial assumption of utility maximizing behavior on the part of decision makers. As a result, it follows that the wide scope of application of the UM decision rule can be seen from a perusal of the applied choice modeling literatures in transportation (e.g., Ben-Akiva and Lerman 1985, Louviere et al., 2000), resource economics (Lupi et al., 2020), marketing (e.g., Louviere et al. 2000) and healthcare (see, e.g., reviews in Soekhai et al., 2019, Clark et al., 2014 and de Bekker-Grob et al. 2012).

Thus, the applied economist is caught between the desire to allow for the possibility that decision makers aren't always utility maximizing (i.e., fully rational) and the welfare machinery for policy analysis that demands a rational consumer to make its inferences about policy impacts (see, e.g., Layard and Walters 1978, or other graduate level microeconomic theory texts).

This dilemma leads to the first objective of our paper, which is to demonstrate that we can point the applied health economist towards an improved representation of decision making by putting the focus on meta-processes as precursors to rational evaluation of alternatives and tradeoffs between them, while still maintaining the link to policy evaluation through welfare theory. As noted earlier, there exists a wide range of non-UM decision rules that could be proposed, but the resulting testable empirical models of choice (1) generally suffer from high computational burdens for the econometrician, and (2) compromise the link to welfare theory central to an economic framework of policy evaluation. So rather than focusing on specific non-UM decision rules, we argue that working at a more aggregate level of decision process characterization, and specifically, that focusing on categorization as a functional building block of decision processes, avoids both these undesirable consequences. Secondly, we present and develop a specific testable model that builds on this parsimonious conceptualization, with the econometric advantage that it nests the utility maximizing assumption. Thus, what is an imposed assumption (UM) in standard choice models such as the MNL, now becomes one of several decision modalities that will be endogenously determined from the observed choices. And thirdly, we test the model using choice data involving cancer-related decision making: a colorectal cancer (CRC) DCE on screening options (e.g., de Bekker-Grob et al. 2021, 2019); and a prostate cancer (PCa) DCE on treatment options. Our main contribution is to suggest and demonstrate the great advantage obtained from extending existing models of choice to include decision meta-process (categorization) process representation, but to do so with an eye to practical impact with improved preference inference making and policy analysis.

Methods

Standard choice models and decision rule implementation

In the spirit of setting the background to the topic of decision processes and choice modeling, we first present a very brief description of the relationship between Data Generation Processes (dgp's) and choice models.

Econometric models of choice behavior posit a Data Generation Process (dgp) that links a set of observed outcomes (say, choice indicators $Y_j=1$ if alternative j is chosen from choice set M - with J alternatives - $Y_j=0$ otherwise) to some attributes X_j and decision maker and context characteristics Z. The succinct statement

$$Y \sim G(D(U(X,Z|\theta),M)), \tag{1}$$

summarizes that the dgp not only has the statistical properties embodied in the multivariate distribution function G(.), which is a function of X, Z and parameters θ, but also embodies a narrative about how the observed choices Y are produced from choice set M using decision rule D. This narrative involves latent variables U (for utility), capturing the attractiveness of the J alternatives through the additive form

$$U = V(X|\beta) + \varepsilon(\mu) \tag{2}$$

where V is a systematic (i.e., from known sources) component of utility (usually, $V = X\beta$), and ε is a stochastic counterpart to V that captures any unobserved sources of utility (with parameters μ that describe the distribution of unknown utility sources). The common decision rule, familiar to choice modelers, embodies this narrative associated with the dgp: the chosen alternative arises from (1) to (2) by being the alternative with highest utility, thus:

$$Y_j = 1 \text{ iff } D(U,M) = j, Y_j = 0 \text{ otherwise}, \forall j \in M, \tag{3a}$$

where iff is "if and only if", and

$$D(U,M) = j \quad \text{iff } U_j = \max(U \in M). \tag{3b}$$

This characterizes the decision maker (DM) as a maximizer (i.e., the decision rule is utility maximization), using full information (the entire X vector) to arrive at (usually compensatory) valuations ($U=V+\varepsilon$, V linear-in-parameters in β with X) of alternatives using pre-existing preferences (β). The particular assumptions made about the joint distribution of ε's will generate different functional forms; to illustrate,

1 ε ~ independent and identically distributed Gumbel(μ), leads to the very familiar Multinomial Logit (MNL) choice model (see, e.g., Ben-Akiva and Lerman 1985);
2 ε ~ MVN(0,Σ), where MVN is the multivariate normal distribution with mean 0 and covariance matrix Σ, leads to the Multinomial Probit model (see, e.g., Daganzo 1979).

Both these exemplars, though different in details, nonetheless implement the same decision rule, utility maximization (expression 3).

Decision rule extensions to standard choice models

Microeconomic (e.g., Lancaster 1966, Stigler and Becker 1977) and psychological (Simon 1955; Luce, 1959) theories of choice postulate value maximization as a quasi-axiomatic view about decision rules, though both fields, but particularly psychology, have suggested a myriad of alternatives to utility maximization (UM) (e.g., elimination-by-aspects, other lexicographic decision rules, satisficing, prospect theory, heuristics, fast-and-frugal methods – among others, see Simon 1955, Tversky 1972, Kahneman and

Tversky 1979, Payne et al., 1993, Gigerenzer 1996, Gigerenzer and Goldstein 1996, Gigerenzer and Gaissmaier 2011).

So, on the one hand we have an incumbent decision rule (UM) that is commonly adopted in microeconomic applied work through the widely available standard choice model forms, such the MNL and MNP models. But on the other hand we have competing arguments that UM may perhaps not pertain in certain domains, contexts or occasions, as per critiques within economics itself (such as McFadden 2006), as well as the availability of a plethora of alternative decision rules which might be deemed more plausible in any given decision domain. As noted earlier, these alternative decision rules generally imply onerous computational burdens when expressed in econometric model form, making their use beyond most applied economists whose primary interest is in policy analysis, not the particularities of individual decision making. In addition, these alternative model forms may create inconsistencies within the policy evaluation framework of welfare economics, which to an economist constitutes a significant barrier to adoption. Ultimately, this places the empirical choice modeler in something of a dilemma: use the standard UM assumption and the accepted policy analysis framework? Or explore a multitude of (potentially plausible) decision rules, which may create the need for idiosyncratic and debatable extensions to welfare theory, if they are even possible? Such questions constitute enough of a challenge in the context of model development and policy analysis that one can hardly fault the practitioner for falling back on the standard UM model forms.

A gist-based choice model extension

This dilemma about decision rule representation calls for a reframing of the problem, by appealing to meta-processes that humans are quite efficient at executing. Specifically, we know that one of the basic cognitive functions that people utilize to manage and facilitate decision making is categorization (see reviews in Loken et al. 2008, Loken 2006, Medin and Smith 1984, Rosch 1978). Categories are components of mental models, which guide our decision making in a given domain. People categorize objects, options, other people, for the purpose of understanding how to handle/treat/process the objects. Categories have associated properties, e.g., preferences, ensuing actions, handling precautions, how to select between alternatives within and between categories, etc., which enable people to process the objects without the need to construct process mechanisms from scratch.

The health and medical decision making literature has recognized that categorization plays a fundamental role in human decision making. In an overview article, Reyna et al. (2015) – see also Reyna et al. (2001), Reyna (2008, 2012), Brewer et al. (2012), Dawson et al. (2012), Hutton et al. (2009), Zikmund-Fisher (2013) – describe a dual-process concept called Fuzzy-Trace Theory (FTT), which postulates there are two kinds of representations of options that undergird mental models used in decision making: verbatim and gist representations. In decision instantiations, these representations are co-developed and co-exist in DM's minds, but according to FTT they fulfill different roles in the decision process. Verbatim representations capture the concrete form of information – numbers, words, pictures – that are useful for steps/processes that require or are benefited by computations and other forms of analysis, e.g., numerical summaries, comparisons. Gist representations, on the other hand, reflect emotion, culture, knowledge/experience, context, world-view, etc., and capture the essential meaning of information. Gists are of different kinds (Reyna et al., 2015): categorical (e.g., safe vs risky, painless vs painful) and ordinal (e.g., not at all risky, somewhat risky, very risky) are simple examples of gists. As noted by Reyna et al. 2015, pp. 111–112), "… gist captures a functionally significant bottom line that integrates and interprets information, often through causal inferences, as opposed to being a list of arbitrary facts." This distinction between verbatim and gist representations does not mean to imply that gists are somehow based on no information whatsoever. Instead, option information, decision context, as well as background knowledge and experience, all combine to form the gist about an alternative.

Swait et al. (2014) propose a model that employs mental models, or schema, to explain choice outcomes. Their model has schema be the basis for defining preference heterogeneity, because knowing the category of an object permits the DM to value an instance of that category. We propose instead that gist-based categorization is a choice-preceding meta-process engaged by the DM to decide how to process individual alternatives and, ultimately, how to choose between them. Thus, categorization leads to the tailoring of different decision rules to the context. Specifically, we propose that when facing a choice between the discrete objects in choice set M, decision makers first categorize each alternative $j \in M$ into one of three "buckets" (see Fig. 1), depending upon a gist judgment about the alternative:

1. **Dominance**[1] **state**: Alternative j effectively has infinite utility, so verbatim representations don't need to be accessed for choice: valuation is unnecessary. Since all alternatives in this state are highly desirable, the decision rule is random choice.
2. **Tradeoff state:** The alternative has finite utility constructed from attribute information, so its verbatim representation will need to be accessed and preferences retrieved. An eventual choice will arise from comparisons between alternatives in this bucket to pick the best, i.e., the decision rule is utility maximization.[2]
3. **Rejection state:** The alternative effectively has utility$\rightarrow -\infty$, so again verbatim representations are not needed. This state normally leads to the alternative's rejection from among the feasible alternatives, with the sole exception being the condition when all alternatives in M are in this bucket. In the latter case, involving all "bad" alternatives, the decision rule is also random choice.

This action and states were proposed by Swait (2009), in what he called the 3-Mix model of choice. He did not, however, provide a

[1] This term is not to be confused with the idea that an alternative is "dominant" in a DCE task, which refers to the possibility that an alternative is "better in attribute" in all dimensions of X than all other alternatives in the task.

[2] Though presented in terms of utility maximization in the Tradeoff state, it would be wholly consistent to develop the model from the perspective of random regret theory (e.g., Chorus 2010).

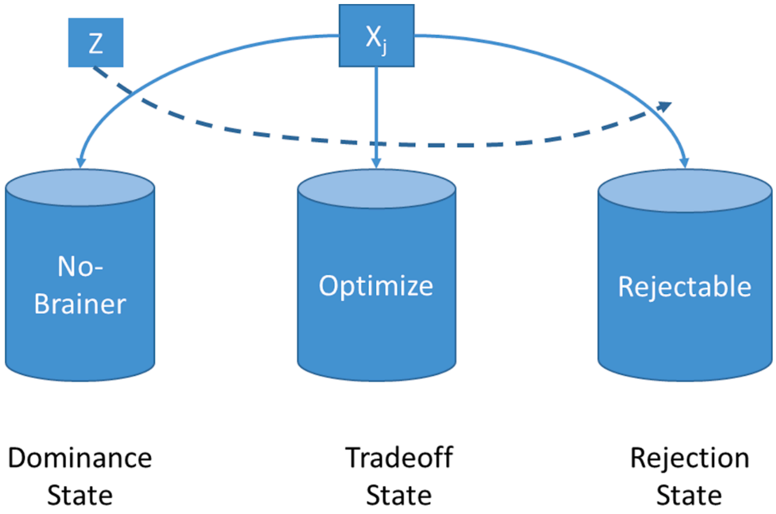

Fig. 1. Gist-based Categorization Phase: Classification of alternatives into decision states (Z=individual difference, X_j=option attributes).

behavioral rationale for these actions and states, nor for their deployment in the decision process. We motivate the action of categorization and the categories themselves as arising from gist-based judgments on the part of DMs, and extend Swait's framework by adding heterogeneity in preferences, now nested within (equivalently, limited to) the Tradeoff state.

Following this classification of each alternative, from the econometrician's perspective the DM has produced a state vector

$$\Psi = \{s_1, s_2, \ldots, s_j, \ldots, s_J\}, \tag{4}$$

which describes how each alternative will be processed during choice. The state variable s_j for alternative j can take on one of three values, corresponding to the state descriptions above:

$$\begin{array}{ll} s_j = +1 & \text{if j in Dominance state} \\ s_j = 0 & \text{if j in Tradeoff state} \\ s_j = -1 & \text{if j in Rejection state} \end{array} \tag{5}$$

Following the Gist-based Categorization Phase (Fig. 1) and production of the ensuing state vector Ψ, the Choice Phase can be undertaken according to Fig. 2. Define the following three-set decomposition of M as a function of Ψ:

$$\begin{array}{l} C_{D|\Psi} = \text{set of alternatives with} s_j = +1 \\ C_{T|\Psi} = \text{set of alternatives with } s_j = 0 \\ C_{R|\Psi} = \text{set of alternatives with } s_j = -1 \end{array} \tag{6}$$

If set $C_{D|\Psi}$ is non-empty, choice is very straightforward because all alternatives in the set have infinite utility (i.e., they are very, very attractive), and so a selection can be made at random among these alternatives. This is a low-cost choice mechanism, where some effort went into the categorization activity, but no valuation of preferences or comparison among alternatives is needed.

If set $C_{D|\Psi}$ is empty, it is then necessary to process set $C_{T|\Psi}$. If $C_{T|\Psi}$ is non-empty, the DM needs to make valuations (based on verbatim representations) and compare among alternatives according to the utility maximization rule. $C_{T|\Psi}$ may not contain all the alternatives in M, since some may have been rejected, which potentially reduces the cost of decision making by reducing the number of alternatives over which optimization must occur. An interesting point to make about this structure is that the proposed model nests within it the UM rule. If we adopt the MNL model for this sub-decision, we would be able to contrast the UM-only model with this UM-in-Tradeoff-only model and thereby infer the impact of the antecedent categorization stage on preferences.

Finally, if $C_{T|\Psi}$ is empty, choice must be made from $C_{R|\Psi}$, the set of alternatives in the Rejection state. This is an unusual feature of the proposed model, that choice can occur from alternatives that are all deemed to be bad outcomes. Arguably, this is a useful feature for certain critical health decisions where all alternatives have unattractive outcomes (e.g., cancer treatments, end-of-life decisions), but one of the actions must nonetheless be undertaken. We will assume, per Fig. 2, that choice is random among the alternatives in $C_{R|\Psi}$.

The prior paragraphs described a hierarchical decision process: first the Gist-based Categorization Phase, when all alternatives are classified into the Dominance, Tradeoff or Rejection states; second, the Conditional Choice Phase, which operates on the categorized alternatives to choose a single option. To formalize this process as a probabilistic choice model, we must recognize that the state vector Ψ is unobserved; we acknowledge its latency by associating with it the probability distribution $Q(\Psi_l)$, where there are $l = 1,\ldots,3^J$ possible state vectors, J=|M|. Formally, the probability of choosing alternative j∈M is

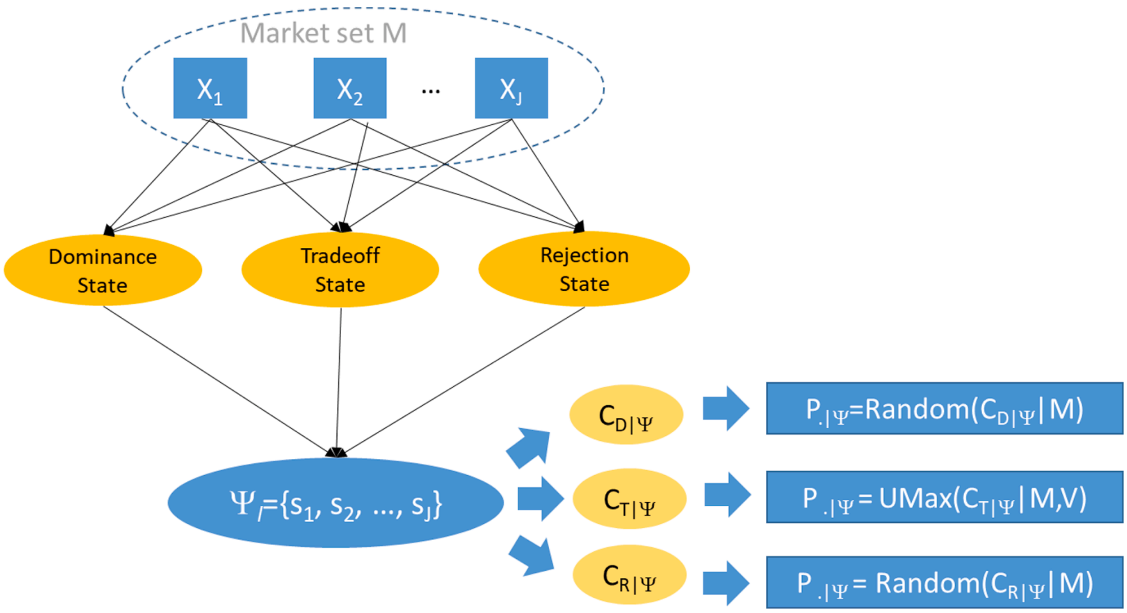

Fig. 2. Conditional Choice Phase, dependent on prior gist-based categorization.

$$P_j = \sum_{l=1}^{3^J} P_{j|\Psi_l} \cdot Q_{\Psi_l}, \forall j \in M. \tag{7}$$

Let us proceed to parse and define the components of expression (7), beginning with the conditional choice probabilities $P_{j|\Psi_l}$, $\forall j \in M$. As indicated in Fig. 2, this probability is defined as a function of sets $C_{D|\Psi}$, $C_{T|\Psi}$, and $C_{R|\Psi}$, as follows:

If $C_{D|\Psi} \neq \emptyset$ Then

$$P_{j|\Psi_l} = \begin{cases} 1/|C_{D|\Psi_l}| & \text{if } j \in C_{D|\Psi_l} \\ 0 & \text{otherwise} \end{cases} \tag{8a}$$

Else if $C_{T|\Psi} \neq \emptyset$ Then

$$P_{j|\Psi_l} = \begin{cases} \dfrac{\exp(\mu_T V_j)}{\sum\limits_{i \in C_{T|\Psi_l}} \exp(\mu_T V_i)} & \text{if } j \in C_{T|\Psi_l} \\ 0 & \text{otherwise} \end{cases} \tag{8b}$$

$$V_j = \beta' X_j$$

Else

$$P_{j|\Psi_l} = \begin{cases} 1/|C_{R|\Psi_l}| & \text{if } j \in C_{R|\Psi_l} \\ 0 & \text{otherwise} \end{cases} \tag{8c}$$

To reiterate, conditional choice probabilities (8a–8c) express the decision rules of random choice among alternatives categorized into Dominance (8a) and Rejection (8c) states, and utility maximization in the Tradeoff state, as embodied in the MNL model (8b). The latter arises from the assumption that the corresponding error terms are IID Gumbel with scale factor μ_T, for all alternatives in $C_{T|\Psi}$. This Tradeoff model differs from the usual MNL expression in that the alternatives are not all those in M, but only those categorized into the Tradeoff state ($C_{T|\Psi}$). It is possible to assume other joint distributions for the error terms in the Tradeoff state, which will result in different conditional choice models replacing (8b). But assuming the MNL for the purposes of this paper permits us to nest the workhorse MNL model within the gist-based framework, allowing a formal test of the hypothesis that the MNL data generating process is a sufficient empirical characterization of a given data set.

We have explicitly allowed for scale factor (or stochastic utility) heterogeneity in (8b). The assumption built into (8b) is that the stochastic utility is constant across the alternatives in set ($C_{T|\Psi}$) – this is fundamental and necessary to justify the MNL form – but can vary across individuals. Thus, we adopt the specification below for the scale factor:

$$\mu_T = \exp(\phi' Z), \tag{9}$$

where Z is a vector of socio-demographic and context characteristics and ϕ is a conforming parameter vector.

To complete the formulation of the choice model, we need to specify the probability of each state Ψ_l. We first define the likelihood of classifying an alternative $j \in M$ into each possible state s_j={Dominance (+1), Tradeoff (0), Rejection (−1)}. Let p_j be the likelihood of categorizing j into Dominance, q_j the likelihood of categorizing j into Tradeoff, and r_i the likelihood of rejection, and note the requirements $0 \le p_j, q_j, r_j \le 1$ and $p_j + q_j + r_j = 1$. For the identity to hold, it is necessary that we know any two of the probabilities, say p_j and q_j, to describe the categorization outcome for an alternative. Now, the probability that the l-th state is observed at the end of Categorization Phase is the likelihood that each alternative is categorized into one of the three individual states. Or, using the notation in Fig. 2 and expression (6), that alternatives in $C_{D|\Psi}$ are categorized into the Dominance state, that the alternatives in $C_{T|\Psi}$ are categorized into Tradeoff, and that the remaining alternatives in $C_{R|\Psi}$ are categorized into Rejection. Assuming these categorizations are probabilistically independent of one another, we arrive at the expression

$$Q(\Psi_l) = \left(\prod_{j \in C_{D|\Psi_l}} p_j\right)\left(\prod_{j \in C_{T|\Psi_l}} q_i\right)\left(\prod_{k \in C_{R|\Psi_l}} (1 - p_k - q_k)\right), \ l = 1, \ldots, 3^J. \tag{10}$$

In our implementation for this paper, we have specified the categorization probabilities p_j, q_j and r_j using a polytomous logit specification. Let the propensity (or factor, or latent gist) Γ_{js} to categorize alternative j into state s={(D)ominance,(T)radeoff,(R)ejection}, or equivalently, $s \in \{+1, 0, -1\}$, be given by

$$\Gamma_{js} = \gamma_s' X_j + \eta_{js}, \quad j \in M, \ s \in \{+1, 0, -1\}, \tag{11}$$

where X_j is the set of attributes of alternative j, γ_s is a state-specific vector of factor loadings, and η_{js} is a stochastic term expressing other (unspecified) variables driving propensity. As noted before, Γ_{js} is the latent factor representation of the s-th gist formed by the decision maker about alternative j. If we assume that stochastic terms η_{js} are IID Gumbel with unit scale, and recognize that one of the parameter vectors γ_s cannot be identified due to the summing-up identity, we conclude that one of these vectors must be normalized to a constant (zero is a convenient choice). Therefore fix $\gamma_R \equiv 0$. We thus arrive at the following polytomous logit model for the categorization probabilities:

Dominance

$$p_j = \frac{\exp(\gamma_D' X_j)}{1 + \exp(\gamma_D' X_j) + \exp(\gamma_T' X_j)}, \ \forall j \in M, \tag{12a}$$

Tradeoff

$$q_j = \frac{\exp(\gamma_T' X_j)}{1 + \exp(\gamma_D' X_j) + \exp(\gamma_T' X_j)}, \ \forall j \in M, \tag{12b}$$

Rejection

$$r_j = 1 - p_j - q_j = \frac{1}{1 + \exp(\gamma_D' X_j) + \exp(\gamma_T' X_j)}, \ \forall j \in M. \tag{12c}$$

We have made the assumption that the categorization process is based on gists constructed by DMs through, *inter alia*, the information known about each alternative, which in the usual DCE is embodied in the attribute vector as well as alternative-specific constants and socio-demographic (Z vector in Fig. 1) intercept shifts and moderator (interaction) terms. As noted earlier, the gists Γ_{js} are not quality or utility assessments, but rather guide the latent categorization process based on information about the alternatives, the context and the decision maker (e.g., knowledge about and experience in the domain). Thus, the variables included in the specifications (12a-c) are meant to be suggestive and could include other possibilities, specific to the application domain and population being studied. The linear-in-parameters functional form used for the gist functions (11) expresses a holistic implementation of gist formation, implying that potentially all information in vector X is determinative in forming the gist. However, once the model is applied to a particular data set, one may find that only subsets of X are statistically significant in forming gist judgments, or even that non-compensatory functional forms are warranted (e.g., a gist is formed by an evaluation that the risk is above a personal comfort threshold or that the benefit is below a desired threshold).

Thus, option attributes can play a dual role in the proposed model: (1) they can drive categorization, enabling the identification of which attributes make an alternative likely to be in Dominance, Tradeoff or Rejection states; (2) they can drive utility, or attractiveness, of the alternatives in the Tradeoff state, within which alternatives without extremal utilities can be chosen. In addition, it should be noted that the decision rule (i.e., random choice among a set of alternatives) adopted for choice among alternatives in the Dominance or Rejection state do not use attribute information for utility valuation, but only for the antecedent categorization. It is only in the Tradeoff mode that the usual concepts of valuation and comparison across alternatives apply. To characterize this property further, preferences as usually understood (the β weights that will ultimately express the partworth formation in functions $V_j = \beta' X_j$, see expression 8b) are only defined and interpretable in the Tradeoff decision modality.

Due to the normalization of the coefficient vector for one gist (which gist is selected is arbitrary since any one of the three vectors can be normalized), interpretations of the marginal impact of a given variable on a gist function must be relative to the normalized gist. For example, if one of the X's is efficacy level and the corresponding γ_D parameter is positive, that would indicate that the Dominance gist increases with increasing efficacy more than the reference Rejection gist does. If γ_D is estimated to be zero, this would be indicative that efficacy is not relatively more of a driver of the Dominance gist, so does not drive categorization into that state any more than it does into the Rejection state.

Finally, we extend the basic model framework by adding stochastic preference heterogeneity to the Tradeoff modality (8c). Specifically, we assume that preference heterogeneity is described by a Latent Class representation with $G \geq 1$ classes to which DMs must be assigned, each class characterized by preferences β_g, $g = 1,\ldots,G$. We do not assume that the gist-function coefficients are person-specific other than from systematic sources (socio-demographic interactions), hence stochastic parameter heterogeneity resides only in the conditional utility functions. We replace (8b) by this expression:

Else if $C_{T|\Psi} \neq \varnothing$ Then

$$P_{j|g,\Psi_l} = \begin{cases} \dfrac{\exp(\mu_T V_{j|g})}{\sum\limits_{i \in C_{T|\Psi_l}} \exp(\mu_T V_{i|g})} & \text{if } j \in C_{T|\Psi_l} \\ & \text{otherwise} \\ 0 & \end{cases} \tag{8b}$$

$$V_{j|g} = \beta_g' X_j$$

The likelihood for $t = 1,\ldots,T$ choice tasks answered by DM n is given by

$$\Lambda_{n|G} = \sum_{g=1}^{G} \left[\prod_{t=1}^{T} \sum_{l=1}^{3^J} \left(P_{ni_t^*|g,t,\Psi_{tl}} Q(\Psi_{ntl}) \right) \right] \cdot \pi_{ng}, \tag{13}$$

where the subscripts for person n and task t have been added throughout, $P_{ni_t^*|g,t,\Psi_{tl}}$ is the probability of choice given by 8b') for the chosen alternative i_t^* for task t, $Q(\Psi_{ntl})$ is given by ((9), and π_{ng} is the probability of n belonging to group/class g. We have conditioned the likelihood on the number of latent classes, which *a priori* is not known and so must be estimated; more on this below. To implement this classification model, we assume this straightforward model:

$$\pi_{ng} = \frac{\exp(\theta_g)}{\sum\limits_{g'=1}^{G} \exp(\theta_{g'})}, g = 1, \ldots, G. \tag{14}$$

This formulation introduces unbounded parameters θ_g, one of which must be normalized (say, $\theta_G \equiv 0$). Maximum likelihood estimates are obtained by optimizing the log likelihood function over the random sample of size N,

$$\ln L_G = \sum_{n=1}^{N} \ln \Lambda_{n|G}, \tag{15}$$

over the full parameter vector ($G, \beta_1, \ldots, \beta_G, \gamma_D, \gamma_T, \gamma_R, \phi, \theta_1, \ldots, \theta_G$), subject to identification restrictions made explicit earlier. Note that G, a discrete quantity, is a parameter to be determined since we do not know a priori the number of latent classes. Maximum likelihood theory does not comport discrete parameters, but this can be circumvented by estimating the conditional parameter set ($\beta_1, \ldots, \beta_G, \gamma_D, \gamma_T, \gamma_R, \phi, \theta_1, \ldots, \theta_G|G$). Under the usual regularity assumptions, including that the model's data generation process is correct, estimates will be asymptotically consistent and unbiased using (15), conditional on G. Parameter estimates will be obtained for a range of G, and goodness-of-fit and other measures (e.g., AIC and BIC information criteria) will be examined to select the optimal G*. This outer estimation process is partly science and partly art: we have augmented the goodness-of-fit measure comparison (where the decision rule is to continue increasing G as long as AIC and/or BIC is decreasing) with an examination of estimated cluster size (as estimated by probabilities π_g in expression 14), stopping the search when the smallest cluster is sized below 100 sample members. This magnitude is in part determined by experience, judgment and the sample sizes we are dealing with (around 1200 respondents). See discussions in Swait (2006) for the use of other than information criteria for model selection.

To summarize, the proposed model (which we term the GBCatL – Gist-Based Categorization Logit) includes the following features/characteristics:

1 Choice is preceded by a gist-based categorization of each alternative into three states: Dominance, Tradeoff and Rejection. See Fig. 1.
2 The underlying latent gist functions are estimated, allowing identification of drivers of categorization. See expression 11.
3 This generates a state vector Ψ that captures all possible classification of alternatives, which is then used as the basis for a three-component decision rule implementation (see expressions 8a–8c, 8b'):
   a If there are any alternatives in the Dominance state (set $C_{D|\Psi}$), choice is random among them since they are equally excellent.
   b Otherwise, if $C_{D|\Psi} = \varnothing$ and $C_{T|\Psi} \neq \varnothing$, choice is obtained by utility maximization from set $C_{T|\Psi}$. Preferences are only relevant in this condition.
   c If $C_{T|\Psi} = \varnothing$, then necessarily all alternatives are in Rejection state, implying that $C_{R|\Psi} = M$. Choice is obtained again using the random choice rule among all alternatives in M.
4 We allow for preference heterogeneity in the Tradeoff modality in two ways: (a) through the inclusion of systematic sources through socio-demographics and context variables, as intercept shifts and moderator interactions; (b) through the stochastic

representation of discrete mass distributions with G taste components. Point (b) implies that GBCatL nests within it the Latent Class (LC) MNL model, which further implies that the log likelihood of the LC class will necessarily be worse than that of the GBCatL.

5 Stochastic utility heterogeneity is allowed by including systematic sources of individual differences and context characteristics in the scale function. This heterogeneity is identical across clusters for the same DM.

Case studies

As illustrative applications of the GBCatL, we employed two data sets encompassing different types of decisions concerning two cancers: colorectal cancer (CRC) screening and prostate cancer (PCa) treatment. The rationale for using these DCE data sets for testing GBCatL is that in daily practice patients receive on the one hand concrete information (numbers, words, pictures) about the different CRC screening and PCa treatment options available (which are mimicked by the DCE tasks), while on the other hand the decision itself is subject to the impacts of information filtered through emotion, culture, knowledge/experience, etc. (see Denberg et al. 2006, Flynn et al. 2011, Clarke et al. 2021). In addition, focusing on the decision of CRC screening and PCa treatment involves different diseases and consequences, which is important for the generalizability of our study findings.

The CRC screening data comprises a sample of $N = 1219$, and the prostate cancer study has $N = 1239$. The CRC DCE is an unlabeled design, whereas the PCa study is labelled. For further details we direct the reader to the Appendix, where we give specific information about survey design, experimental design for the DCEs, and further sample facts.

Results

Colorectal cancer screening DCE

We present in Table 1 the statistics obtained by fitting the GBCatL model to the CRC data, for different number of clusters $G = 1, \ldots, 6$. On the basis of the model selection criteria mentioned before, we select the model with $G = 4$ as the most reliable and valid representation of the data generation process (selected model shown in bold in Table 1). Table 1 reinforces the message that the choice data reflects a strong measure of heterogeneous preference driving choice behavior, but it must be noted that this heterogeneity is nested within the Tradeoff gist in the GBCatL models. From this point on, we shall be discussing the $G = 4$ model.

For comparison purposes, we provide the goodness-of-fit information for two baseline models: the standard MNL and Scaled MNL models. The rationale for selecting these baseline models is that they constitute standard selections for the applied economist undertaking policy analysis work, and embody the UM hypothesis in their formulations. (This rationale stands in contrast to selecting baseline models to test the categorization-then-choice concept against other decision rules, such as EBA, different forms of F&F, etc.) Another advantage to using these baseline models is that both are nested within the GBCatL($G = 1$): the standard MNL when all alternatives are categorized into the Tradeoff category and the scale factor is unity for all respondents; and the Scaled MNL when all alternatives are categorized into the Tradeoff category, but scales vary as in the GBCatL. Both baseline models have homogeneous preferences. The information criteria, with focus on BIC but consistently reinforced by Deviance and AIC, indicate that both these models are inferior characterizations of the data compared to the GBCatL($G = 1$). The difference in BIC for the MNL is 404.6 points, and for the Scaled MNL is 289.8 points, in favor of the GBCatL($G = 1$) model. This is strong support for the empirical plausibility of the antecedent gist-based categorization that underlies the proposed model, compared to the Tradeoff-only assumption built into the baseline models. Compared to the selected GBCatL($G = 4$) model, the homogeneous preference MNL model has a large disadvantage of 4781.9 BIC points, militating strongly against the latter's depiction of the data generating process.

Table 2 presents the parameter estimates for the GBCatL($G = 4$) model. Following the overall logic of the model, we first examine the Gist functions. All the predictors and their socio-demographic interactions in the Gist functions are highly significant, emphasizing that the categorization process is clearly driven by the attributes of the alternatives; further, that the categorization of the opt-out alternative is strongly impacted by the characteristics of the decision maker (higher education, health literacy, numeracy and decision style), prior experience with cancer, family association with cancer and having a positive attitude towards CRC screening. This is in strong agreement with the expectations of Reyna et al. (2015) concerning gist formation.

As explained before, understanding the relative influence of the different gist drivers requires evaluating their impact relative to the base gist (in our case, Rejection). We therefore present an analysis in Fig. 3 based on probabilities of categorization. We initially set an

Table 1
Goodness-of-fit statistics for Colorectal cancer screening DCE model selection.

G	Deviance(=−2*LL)	AIC	BIC	Smallest Cluster (%)	Smallest Cluster (#)
MNL	28,559.4	28,577.4	28,623.4	1.0	1219
MNL Scale	28,387.8	28,421.8	28,508.6	1.0	1219
1	27,899.0	27,989.0	28,218.8	1.0	1219
2	25,951.1	26,061.1	26,341.9	0.387	472
3	24,028.8	24,158.8	24,490.6	0.095	116
4*	**23,308.6**	**23,458.6**	**23,841.5**	**0.085**	**104**
5	22,989.1	23,159.1	23,593.1	0.047	57
6	22,779.7	22,969.7	23,454.8	0.028	34

* Selected model.

Table 2
Colorectal cancer screening DCE: selected GBCatL model ($G = 4$).

Goodness-of-fit				
Log likelihood	−11,654.30		# Choice Sets	19,504
Number of parameters	75		N	1219
Deviance	23,308.60		Tasks/subject	16
AIC	23,458.60			
BIC	23,841.53			
Rho-squared	0.3330			
Rho-squared bar	0.3287			
Utility Function (Tradeoff Modality)	Segment 1	Segment 2	Segment 3	Segment 4
Opt-Out Constant	−15.7976***	−6.5347***	11.6936***	−4.3378***
LN(Effectiveness+1)	3.2576***	0.5841***	1.6815***	0.2813***
LN(Prob. of Not Finding+1)	−4.6396***	−1.1336***	−3.1282***	−0.4412***
Wait for test results/3	0.1276***	−0.0682***	−0.0361	−0.5063***
Wait for followup/8	−0.3128***	−0.5105***	−0.2646***	−1.2314***
Screening Frequency/3	−2.2052***	1.1733***	4.4713***	−3.5085***
Opt-Out x No cancer before	7.0718***	−0.476***	−5.5778***	−0.5222***
Opt-out x CRC in family	0.7446***	1.1557***	−2.5154***	−1.3569***
Opt-out x CRC scrng positive att	−2.2436***	2.3784***	−0.8754***	−4.0421***
Scale function: ln(μ_T)				
Age 65+	−0.0999**			
Higher education	−0.013			
No cancer before	−0.0509			
CRC in family	−0.0002			
CRC scrng positive att	0.3245***			
Health literacy sufficient	0.0128			
Numeracy sufficient	0.098**			
Deliberative decision style	0.055			
Gist Functions	Dominance	Tradeoff	Rejection	
Opt-Out Constant	−0.4138***	19.1091***	−0-	
LN(Effectiveness+1)	−1.8321***	0.4332***	−0-	
LN(Prob. of Not Finding+1)	2.332***	1.7742***	−0-	
Wait for test results/3	3.3694***	4.0136***	−0-	
Wait for followup/8	22.6415***	1.5423***	−0-	
Screening Frequency/3	−15.9385***	−15.4495***	−0-	
Opt-out x CRC in family	−8.5911***	−1.1328***	−0-	
Opt-out x Higher education	−1.2517***	6.4344***	−0-	
Opt-Out x No cancer before	3.1266***	−5.4301***	−0-	
Opt-out x CRC in family	−9.4041***	−3.1076***	−0-	
Opt-out x CRC scrng pos att	1.0267***	−4.9501***	−0-	
Opt-out x Health literacy suff	−10.9853***	−5.0817***	−0-	
Opt-out x Numeracy suff	−3.69***	−2.9905***	−0-	
Opt-out x Delib. decision style	−5.4293***	−6.0984***	−0-	
Segment Coefficients (MNL)			Size (%)	
Segment 1	0.0705		28.2%	
Segment 2	0.3456***		37.1%	
Segment 3	−1.1227***		8.5%	
Segment 4	−0-		26.2%	

Parameter significance coding: Blank:$0.10<p$; *: $0.05<p \leq 0.10$; **: $0.01<p \leq 0.05$; ***: $0<p \leq 0.01$.
LN=natural logarithm.

alternative, labelled Alt1 for exposition, at lowest levels on all attributes; we then systematically change one attribute at a time for Alt1, from lowest to highest level. (For this analysis, all socio-demographic moderators are set to zero.) The consequent positive or negative changes in probability of categorization into the Dominance, Tradeoff and Rejection Gists are presented in Fig. 3. The graph forcefully drives home the point that Frequency of Screening and Wait Time for Follow-up are the main influences on categorization, but these attributes work differently from one another. Decreasing Frequency of Screening from 1- to 3-years influences categorization by making the Tradeoff state less likely for Alt1 and the Rejection state more likely. In contrast, increasing Wait Time for Follow-up from 2 to 8 weeks leads to the Tradeoff state becoming less likely and the Dominance state more likely. Both these impactful attributes make it less likely the alternative is kept in the Tradeoff state, but the gist categorization is different across them. The other three attributes are far less impactful, despite the statistical significance of the individual attribute coefficients. Effectiveness has an interesting pattern, nonetheless: as Effectiveness increases from minimum to maximum, the Dominance state becomes less likely, to the benefit of the Tradeoff state. It is possible that higher Effectiveness introduces the possibility that its generated benefit will offset disadvantages in other attributes during Tradeoff modality, so that increases the likelihood of categorizing the alternative into the latter modality.

Table 3 presents a second perspective on the gist categorization process. Table 3a shows the average decision state probabilities over all scenarios and respondents in the DCE, by alternative (two screening alternatives, plus an opt-out alternative). Thus, on average, the Dominance state had an average probability of about 0.013 for the screening alternatives, but 0.001 for opt-out. Thus, the likelihood of the Dominance state was extremely low, in this experiment (more on this below). The Tradeoff state was very likely, with

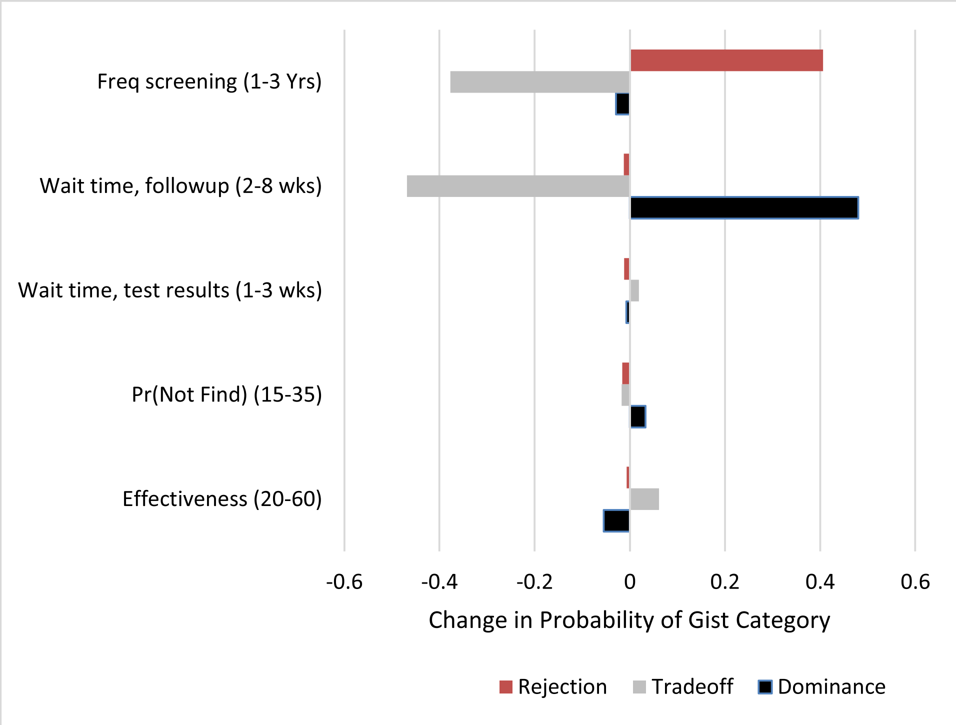

Fig. 3. Impact of individual attributes on probability of categorization into gists (Colorectal cancer screening DCE).

Table 3
Colorectal cancer screening DCE: aggregate 3-alternative scenario state probabilities.

a) Aggregate decision state probabilities over DCE runs			
	Decision State Average Probabilities		
Alternative	Dominance	Tradeoff	Rejection
Screening 1	0.014	0.961	0.024
Screening 2	0.012	0.961	0.026
Opt-out	0.001	0.772	0.227
Average	0.011	0.924	0.066
b) Aggregate gist-based categorization combinations, 3 alternatives			
Alternatives			Aggregate State Probability
Screening 1	Screening 2	Opt-out	
Tradeoff	Tradeoff	Tradeoff	0.714
Tradeoff	Tradeoff	Rejection	0.210
Tradeoff	Rejection	Tradeoff	0.037
Rejection	Tradeoff	Tradeoff	
	All other states		0.039

average probabilities of 0.961 for the screening options, but a much lower 0.772 for the opt-out alternative.

The likelihood of categorization into the Rejection state was ~0.025 for screening alternatives, but a substantial 0.227 for opt-out. Thus, in aggregate, the behavioral picture that emerges is that the Tradeoff category predominates in this experiment, but Rejection was substantial for the opt-out alternative. Additionally, insights about the gist categorization states for multiple alternatives are laid out in Table 3b, which shows the likelihood of multi-alternative categorizations for 3-alternatives scenarios (Scrn1, Scrn2, Opt-out), based on the average probabilities from Table 3a. This analysis shows that on average the most likely state of scenario categorization was all alternatives in Tradeoff state, with a probability of 0.714. The next most likely scenario categorization has Scrn1 and Scrn2 in Tradeoff state, but Opt-out in Rejection, with a probability of 0.210. The two scenario combinations in which either Scrn1 or Scrn2 is in Rejection, and the remaining alternatives in Tradeoff, has a likelihood of 0.037. All other scenario categorizations account for the remaining 0.039 probability mass. These insights confirm the earlier observation that the Tradeoff category predominates, but deepen our understanding that trading between Scrn1 and Scrn2 (which essentially identifies preferences) occurs with probability 0.924, on average.

These results suggest an unexpected use (at least for the authors!) of the GBCatL model as a diagnostic tool to evaluate the success of a DCE setup and design in generating the conditions for reliable inferences about preferences. Specifically, a DCE that has low aggregate probabilities for the Tradeoff state (Table 3a) and/or low probabilities for high Tradeoff scenario categorization (Table 3b) should be viewed skeptically and its preference inferences questioned with respect to validity and generalizability.

Another interesting perspective on the categorization process is shown in Table 4, which presents the CRC screening configurations that maximize probabilities of categorization into each of the three states. An alternative that has Effectiveness=60%, Probability of Not Finding Cancer=0.35, a 3-week Wait for Test Results, a 2-week Wait for Follow-up and a 1-year interval for Frequency of Screening, results in an estimated probability of 0.991 of categorization into the Tradeoff state. The other configurations shown in the table correspond to the Dominance and Rejection states. (In evaluating the categorization performance profiles presented in Table 4, it is important to remember that the roles of attributes can be quite different between the categorization and utility functions.) This type of information can by itself be used to design screening options to avoid particular undesirable target states, such as Rejection, or to increase likelihood that a configuration becomes categorized into the Dominance state. Earlier, Table 3 suggested that across the CRC screening DCE the likelihood of Dominance was very low, on average. But Table 4 suggests that, across the DCE design space, both the Dominance and Rejection states can display quite large probabilities (0.594 and 0.427, respectively). Thus, Table 3 should be used to give a bird's-eye diagnostic of the DCE, but Table 4 can give more precise insights into alternative configurations that might lead to extremal gist categorizations.

We next switch our attention to the representations of scale and preference heterogeneities in the Tradeoff modality. The selected model has $G = 4$, with segment sizes of 28.2%, 37.1%, 8.5% and 26.2%, respectively (see Table 2). Preference heterogeneity is quite pronounced across the four segments, per the utility function coefficients. All attributes display significant differences across segments, but retain intuitively plausible directions in all cases. We highlight that the Opt-Out constant, indicating average preference for the Opt-Out alternative, shows that segments 1, 2 and 4 are aversive to the Opt-Out, while segment 3 is strongly Opt-Out favorable; additionally, note that the interactions of the Opt-Out constant with several socio-demographics indicate that the choice behavior with respect to the Opt-Out alternative is also strongly heterogeneous across stochastic and systematic sources. Changing focus to an attribute that did not play a leading role in the gist categorization stage, note that all segments value increased Effectiveness, but particularly segments 1 and 3. This impact is conditional on being in the Tradeoff modality, and is in addition to the prior impact on categorization.

Finally, we note that there is some evidence of scale heterogeneity, specifically with respect to Age 65+ (decreases scale somewhat), having a positive attitude towards CRC screening (increases scale) and being numerate (also increased scale somewhat). Other socio-demographics we tested were not individually statistically significantly different from zero in the scale function.

Prostate cancer treatment DCE

Our second DCE's goodness-of-fit statistics are shown in Table 5. In this case, the selected GBCatL model has six segments, since the seven-segment model violates our minimum segment size cutoff criterion. As in the CRC DCE, the GBCatL models demonstrate a large improvement with respect to the baseline MNL and Scaled MNL models: the improvement in BIC from the MNL to the GBCatL($G = 6$) is of 7437.8 points in favor of the latter model, and 7378.9 points from the Scaled MNL in favor of the GBCatL. Contrasting these two baseline models with the non-heterogeneous version proposed model GBCatL($G = 1$), we note BIC improvements of that model with the baselines of 1660.9 and 1602.0 in favor of GBCarL($G = 1$); this is quite a bit larger than the equivalent differences in the CRC screening DCE. As a comment on the degree of heterogeneity, Table 5 suggests that individual preferences in the Tradeoff modality are very different across respondents. In the continuation of this section, we exposit the six-segment model only.

Again, we first focus on the Gist functions (model parameter estimates presented in Table 6). All other things equal, we note from the constants that the Tradeoff gist is negatively impacted (relative to Rejection) by surgery, brachytherapy and radiotherapy relative to active surveillance, approximately to the same degree. In contrast, the Dominance state becomes more likely for radiotherapy relative to active surveillance. All attributes (which are essentially risks and effectiveness, all expressed as probabilities) influence both Dominance and Tradeoff gists, relative to Rejection. Additionally, the risks of permanent urological, erection and bowel problems display heterogeneity with respect to DM prior experience with these conditions.

Finally, the impact of socio-demographics and prior experience on the gists for active surveillance is also apparent in the model.

The aggregate state probabilities over the DCE are shown in Table 7a, which clearly show that on average the Dominance state has probability zero. The other extremal state, Rejection, is on average about 0.04 for the two treatments (labelled PCa1 and PCa2), and 0.111 for the opt-out alternative. This latter figure is half the value estimated for the CRC DCE (Table 3a). Again, the statistics in Table 7a lead us to believe that the task and design overwhelmingly favored the categorization of alternatives into the Tradeoff state, which supports the identification of preferences from the comparisons in this data. Table 7b confirms this by showing the most likely

Table 4
Colorectal cancer screening alternative configurations that maximize individual state probabilities.

Attribute	Screening configuration to achieve maximum gist-based state probability Dominance	Tradeoff	Rejection
Effectiveness	20	60	20
Pr(Not finding cancer)	35	35	15
Wait time for test results (weeks)	1	3	1
Wait time for follow-up (weeks)	8	2	2
Frequency of screening (years)	3	1	3
Predicted probability of categorization into state	0.594	0.991	0.427

Table 5
Goodness-of-fit statistics for Prostate Cancer Treatment DCE model selection.

G	Deviance(=−2*LL)	AIC	BIC	Smallest Cluster (%)	Smallest Cluster (#)
MNL	33,678.4	33,706.4	33,778.1	1.0	1239
MNL Scale	33,555.4	33,601.4	33,719.2	1.0	1239
1	31,725.5	31,835.5	32,117.2	1.0	1239
2	28,119.8	28,259.8	28,618.3	0.422	523
3	27,221.7	27,391.7	27,827.1	0.304	377
4	25,861.0	26,061.0	26,573.2	0.191	237
5	25,648.1	25,878.1	26,467.1	0.141	175
6*	**25,414.4**	**25,674.4**	**26,340.3**	**0.127**	**157**
7	25,261.6	25,551.6	26,294.3	0.063	78

* Selected model.

scenario categorization outcomes, using the average probabilities in Table 7a: all alternatives in Tradeoff state accounts for approximately 0.82 probability mass, while (PCa1, PCa2) in Tradeoff state but opt-out in Rejection accounts for another 0.10 probability mass, leading to a total probability of 0.92 that the runs in the DCE support preference identification. Choice scenarios that categorized with at least one of PCa1 or PCa2 in Rejection state, but the other two alternatives in Tradeoff, accounted for 0.071 of the probability mass. These figures are indicative that the DCE was well-designed for the purpose of preference identification.

We investigated the underlying cause for the Dominance state having essentially zero probability, and found that it is due to the attribute Probability of 5-year Survival. For all treatments, this attribute was varied from $x = 70\%$ to 90%, and all treatments had the same range. The large coefficient for this attribute in the Dominance gist ($\gamma = -2.8672$) created an inescapable barrier with a value of $-2.8672 * \mathrm{LN}(x + 1) = [-12.22, -12.93]$. None of the other attributes offset this very negative value, in fact most reinforced it by making it even more negative, leading to the zero probability for the Dominance gist. Again, from the perspective of diagnosing the DCE setup, this result suggests that it might have been advantageous to vary this attribute over different ranges for each treatment option, and perhaps the ranges could have been extended further. This might or might not be a plausible action to take in this domain, but this illustrates the types of insights afforded by the proposed model that might ultimately lead to improved DCE design.

Both preference and scale heterogeneity are strongly evident in this DCE. We noted beforehand how goodness-of-fit had improved tremendously with the more flexible segment representations, and Table 6 makes it plain that preference heterogeneity arises across all attributes. Segment sizes are, in order, 16.1%, 12.7%, 28.1%, 13.1%, 15.3% and 14.7%, implying that segment sizes in the sample are of at least 157 DMs, which gives us confidence that preference differences are robustly determined. Singling out the Probability of 5-year Survival attributes, we note that there seem to be three tiers of sensitivity: segments 1 and 4, with sensitivities of about 3.0; segments 2 and 3, with sensitivities between 5 and 7; and segments 5 and 6, with very large sensitivities of about 13 and 27. This third group is being strongly driven by treatment outcome differences; while risks are significant in their evaluation, they are not nearly as highly valued as survival probability. All segments, ceteris paribus, value some intervention *vis-à-vis* active surveillance, though segment 2 associates a disutility with surgery relative to surveillance. Interestingly, observe that the interventions viewed positively relative to surveillance have very similar constants within a segment; this suggests that the model in Tradeoff modality is capturing a willingness to undergo any of the surgery, brachytherapy, radiotherapy interventions compared to undertaking active surveillance, differentiating among the interventions in terms of risks and benefits. By definition, this is the behavior posited in the normative models underlying most choice models. In the application of the GBCatL model, we are able to "purify" process-based noise to help the data to meet near-axiomatic behavior; in this experiment, we control for the Rejection possibility, which would compromise direct application of models like the MNL.

In this second DCE, scale heterogeneity is detected in more dimensions than in the CRC DCE: we find statistically significant effects for age 65+, higher education, having prostate cancer in the family, having experienced erection problems before, and having a deliberative decision style.

Discussion

As a generalization of the conventional utility maximization (UM) behavioral paradigm built into usual choice models (e.g., MNL, NMNL and MNP), the Gist-based Categorization (GBCat) model has shown itself a useful extension. It nests the UM behavior within the Tradeoff category it postulates to exist along with Dominance and Rejection categories. In the two cancer DCEs we examined with the proposed model, we find strong support for the concept of categorical gists (Reyna et al., 2015); as postulated by Fuzzy Trace Theory (FTT), it seems that the gists are driven by option attribute information, decision-maker characteristics and prior experience or familiarity with the disease domain.

In both DCEs there is limited (CRC) or no support for the Dominance gist (PCa), at the aggregate level. However, in the CRC DCE we find that there are conditions when both the Dominance and Rejection states have relatively high probabilities (see Tables 3a and 4, particularly). At the aggregate level (by which we mean across DMs and scenarios in the DCE), both experiments display a high rate of categorization of all alternatives or all designed alternatives into the Tradeoff state, with estimated average probability higher than 0.90. These figures highlight a useful assessment that can be made of a DCE's quality as a preference measurement tool: a preference elicitation tool that depends upon preference revelation through comparisons across alternatives depends fundamentally upon a large proportion of scenarios having at least two alternatives in the Tradeoff state. If this does not occur, either because some alternatives are

Table 6
Prostate cancer treatment DCE: selected GBCatL model ($G = 6$).

Goodness-of-fit						
Log likelihood	−12,707.20	# Choice Sets		19,824		
Number of parameters	130	N		1239		
Deviance	25,414.41	Tasks/subject		16		
AIC	25,674.41					
BIC	26,340.27					
Rho-squared	0.2836					
Rho-squared bar	0.2763					
Utility Function (Tradeoff Modality)	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Segment 6
Surgery constant	8.4137***	−1.6663***	11.7674***	5.351***	6.6573***	1.9204***
Brachytherapy constant	8.6651***	0.198	11.7168***	5.4327***	6.7572***	2.1864***
Radiotherapy constant	5.9608***	0.0842	11.7081***	4.0079***	6.2591***	1.9454***
Active surveillance constant	−0-	−0-	−0-	−0-	−0-	−0-
R1=LN(Risk permanent urological problems)	−0.0985	−1.0051***	−0.4452***	−0.7851***	−0.3552***	−0.4895***
R2=LN(Risk permanent erection problems)	−0.5073***	−4.2034***	−1.7539***	−1.2131***	−2.0624***	−0.5242***
R3=LN(Risk permanent bowel problems)	−0.3709***	−1.2871***	−0.479***	−1.0365***	−0.3033***	−0.6843***
LN(Prob 5-year survival)	3.0747***	6.9948***	5.0637***	2.9499***	13.8863***	27.3***
R1 x urological problems before	−0.1304	−0.0251	0.0386	0.7896***	−0.0691	0.2452***
R2 x erection problems before	−0.1688***	−1.1274***	0.0022	0.5376***	−0.1999***	−0.2251***
R3 x bowel problems before	−0.1463	−0.4774**	0.1456*	−1.8371***	−0.4211***	0.1649*
Surveillance x Age 65+	0.4494***	1.5502***	−0.9306***	1.7272***	0.2004	−1.2062***
Surveillance x Experience good health	1.9695***	0.0807	0.5487***	0.2862	0.3044***	0.8559***
Surveillance x No PCa before	1.763***	−1.094***	0.6173**	0.9228***	−0.8586***	−1.6211***
Surveillance x PCa in family	0.3672**	−0.8379***	0.1462	−0.4321	0.3088**	−0.5009***
Scale function: $\ln(\mu_T)$						
Age 65+	−0.2382***					
Higher education	0.1132**					
No PCa before	0.0271					
PCa in family	−0.128**					
Erection problems before	0.2333***					
Urological problems before	−0.1					
Bowel problems before	−0.0873					
Surveillance positive attitude	0.0155					
Deliberative decision style	0.1642***					
Gist Functions	Dominance	Tradeoff	Rejection			
Surgery constant	−0.9522***	−1.7959***	−0-			
Brachytherapy constant	−0.0883	−1.8155***	−0-			
Radiotherapy constant	1.9486***	−1.8968***	−0-			
Active surveillance constant	−0-	−0-	−0-			
R1=LN(Risk permanent urological problems+1)	0.991***	0.0607	−0-			
R2=LN(Risk permanent erection problems+1)	−0.2861***	1.2802***	−0-			
R3=LN(Risk permanent bowel problems+1)	−2.5093***	−1.4679***	−0-			
LN(Prob 5-year survival+1)	−2.8672***	0.8402***	−0-			
R1 x urological problems before	−1.9035***	−0.0934	−0-			
R2 x erection problems before	−0.1149	−0.2812***	−0-			
R3 x bowel problems before	0.3748***	0.2566**	−0-			
Surveillance x Age 65+	0.6969***	−0.0309	−0-			
Surveillance x Experience good health	0.313***	0.185*	−0-			
Surveillance x No PCa before	0.069	−0.9863***	−0-			
Surveillance x PCa in family	0.165	−0.02	−0-			
Surveillance x Surv positive attitude	−1.358***	2.4394***	−0-			
Surveillance x Deliberative decision style	0.5064***	−1.5724***	−0-			
Segment Coefficients			Size (%)			
Segment 1	0.0931		16.1%			
Segment 2	−0.1418		12.7%			
Segment 3	0.6503***		28.1%			
Segment 4	−0.1118		13.1%			
Segment 5	0.044		15.3%			
Segment 6	−0-		14.7%			

Parameter significance coding: Blank: $0.10<p$; *: $0.05<p \leq 0.10$; **: $0.01<p \leq 0.05$; ***: $0<p \leq 0.01$.
LN=natural logarithm.

in Rejection (implying some type of choice set formation being in play – see, e.g., Swait and Ben-Akiva 1987, Veldwijk and Swait 2022) or because others are in Dominance (implying that attribute-based utility comparisons are not employed for choice), the reliability and accuracy of the DCE task and experimental design should be questioned.

Fig. 3 illustrates what we think may be the important takeaway from these empirical exercises with the GBCatL model: the gist-based categorization is an important source of re-interpretation of attribute impacts on choice. The simple insight by the FTT

Table 7
Prostate Cancer Treatment DCE: 3-alternative scenario aggregate state probabilities.

a) Aggregate decision state probabilities over DCE runs			
	Decision State Average Probabilities		
Alternative	Dominance	Tradeoff	Rejection
Screening 1	0.000	0.959	0.041
Screening 2	0.000	0.957	0.043
Opt-out	0.000	0.889	0.111
Average	0.000	0.945	0.055
b) Aggregate gist-based categorization combinations, 3 alternatives			
Alternatives			Aggregate State Probability
PCa1	PCa2	Opt-out	
Tradeoff	Tradeoff	Tradeoff	0.817
Tradeoff	Tradeoff	Rejection	0.102
Tradeoff	Rejection	Tradeoff	0.071
Rejection	Tradeoff	Tradeoff	
	All other states		0.010

conceptualization leads to two major improvements in the choice model explanation of observed behaviors: (1) it separates tradeoff behaviors (utility evaluation and utility comparison across alternatives) from DMs' antecedent decisions about how to make a decision (antecedent volition, Swait and Adamowicz 2014) by indicating the basis for that decision through the attributes and personal characteristics of the DM; (2) it isolates (or conditions) preference heterogeneity into the Tradeoff modality, suggesting that preference heterogeneity inferences based on models ignoring gist-based categorization (as well as other mechanisms embodying antecedent volitions) are attempting to capture gist-driven impacts. Given the strong improvement in goodness-of-fit measures in Tables 1 and 5, it would seem that incorporating gist-based categorization is supported as a strong discriminant between data generation process descriptions for DCE data in health.

While we examined two DCEs to seek some degree of generalizability of results, we are limited in number and scope of domains. Further analysis with the GBCatL of DCEs in other health domains is warranted, to map out the scope of application for this extended model of choice. In our applications of the model, we have come to realize that it can be used as a DCE evaluation tool, by revealing the extent and conditions under which tradeoff operations are likely to hold. Applying the GBCatL across DCEs and domains will not only permit an evaluation of the degree to which preference identification is reliably supported by an individual DCE, but will also help us develop a broader perspective on what design practices and parameters lead to successful DCE applications. (We employ the word "design" very broadly here, referring to the entire DCE complex and not just to the experimental design.)

Though we have presented the model development and testing in the context of DCEs, there is nothing about the model that precludes its application to other data sources (e.g., RP, or revealed preference, behaviors). While this type of behavioral analysis has seen less use in the health and health economics arenas, it is eminently feasible and worthwhile to conduct choice modeling efforts on actual behaviors. These are available from administrative data sources and constitute an unexploited source of behavioral data, that can be analyzed separately or in conjunction with DCEs involving the same focal behaviors. This latter practice is called "data fusion" (among others, see Louviere et al. 2000), and is a promising avenue of investigation for health researchers.

Conclusion

We have proposed and tested an update to Swait's (2009) 3-Mix model, justifying it as an expression of Fuzzy Trace Theory, as exposited by Reyna et al. (2015). We further add preference heterogeneity to the basic framework, nesting it within the tradeoff component of the choice model. We argue that the resulting model (which we term the Gist-based Categorization Logit – GBCatL) builds on fundamental categorization skills that decision makers possess and use to shape (and perhaps expedite or otherwise facilitate) their decision making processes. FTT proposes that individuals form gists about alternatives, making use of information they have about the alternatives themselves, but also what they may have arising from culture, context, knowledge and experiences, *inter alia*. We find strong support for the conceptual framework of FTT (i.e. gist-based categorization) in the analysis of data from two DCEs dealing with different aspects of cancer screening and treatment.

As noted in the Introduction, the GBCatL model helps in mitigating the tension between the need to relax the rationality assumption when dealing with choice data in healthcare applications (as well as in other fields of applied economics), and the need to use welfare theoretic concepts and tools (which require a rational decision maker) for policy evaluation. It does so by nesting the utility maximizing assumption into a specific decision modality (the Tradeoff mode), the drivers of which are endogenously determined by the model. This leads to an internal validity evaluation of the scope of rationality as driven by decision context, aspects (or attributes) of options, and decision-maker characteristics. This parsing of the incidence of utility maximizing behavior in a particular application helps the applied economist to evaluate the degree to which the preference measurement instrument (i.e., DCE) and its statistical expression in the form of a choice model successfully links to the welfare theoretic policy evaluation framework.

Funding sources

J.D. Swait was supported by the Erasmus University via the Erasmus Initiative "Smarter Choices for Better Health". E.W. de Bekker-

Grob was supported by The Netherlands Organization for Scientific Research under NWO-Talent-Scheme-Vidi-Grant No. 09,150,171,910,002, and the DCE studies by the same grantor through NWO-Talent-Scheme-Veni-Grant No. 451–15–039. The funders had no role in the study design, in the collection, analysis, and interpretation of data, in the writing of the report, or in the decision to submit the manuscript for publication.

CRediT authorship contribution statement

J. Swait: Conceptualization, Methodology, Formal analysis, Software, Writing – original draft, Writing – review & editing. **E.W. de Bekker-Grob:** Conceptualization, Formal analysis, Writing – review & editing, Data curation.

Declaration of Competing Interest

None to report.

Appendix – Details of case studies

The first of the DCEs used in this study involves decision making around colorectal cancer (CRC) screening, and is fully reported on by de Bekker-Grob et al. (2019) and de Bekker-Grob et al. (2021). These should be consulted for even more detail than we provide here.

This DCE involved unlabeled alternatives presented in the traditional tabular form, under three conditions: (1) (two alternatives (pairs)) two CRC screening alternatives or a CRC screening alternative with an opt-out for no screening; (2) (three alternatives (triples)) two different CRC screening alternatives plus the opt-out for no screening; (3) a mix of two and three alternatives (mixed). Of note, the DCE design with two alternatives was a better reflection of the actual decision for CRC screening, but exhibits lower statistical efficiency than the DCE design with three alternatives. Respondents were randomly assigned to the conditions, and did 16 tasks each. Five attributes characterized CRC screening alternatives: Effectiveness, Probability of not finding cancer, Waiting time for test results, Waiting time for follow-up test, and Frequency of screening.

The second DCE deals with treatment options for prostate cancer, and generally has the same design and delivery structure as employed for the CRC data. This was a labelled DCE involving four treatment types: active surveillance (effectively a default option), surgery, brachytherapy and radiotherapy. Each scenario/task included two or three alternatives (just like the CRC screening DCE study described above), according to the governing experimental design. The treatments were described by seven attributes: Risk of permanent urological problems, Risk of permanent erection problems, Risk of permanent colorectal problems, 5-year survival probability, Frequency of PSA test, Hospital visits for treatment, and Possibility of cure. However, only the first four attributes were varied by the experimental design, while the remainder were held at fixed values representing standard current practice, and all seven were presented to DMs. In addition, for the active surveillance treatment the first three attributes and the sixth do not apply, making the levels of attributes alternative-specific.

To maximize the D-efficiency of each DCE design for CRC screening and prostate cancer treatment while accommodating for the possibility of substantial respondent heterogeneity, Bayesian heterogeneous DCE designs were used (Sándor et al., 2005). That is, for all six DCE studies (i.e., two medical areas times three DCE design conditions - pairs, triples, and mixed), we generated a heterogeneous DCE design consisting of 10 sub-designs (specialty Fortran programs were developed). Each respondent was assigned one randomly selected sub-design containing 16 choice tasks. Together these sub-designs were optimal to estimate an MNL. The prior preference information (attribute weights) as required for the Bayesian efficient optimization approach was obtained from best guess priors using expert judgement and updated for each of these six DCE studies after a pilot run of 100 respondents each.

The choice tasks of the three DCE studies of each medical area were the following: (i) 16 DCE choice tasks with two CRC screening / prostate cancer treatment alternatives or one CRC screening / prostate cancer treatment alternative and an opt-out option; (ii) 16 DCE choice tasks with two CRC screening / prostate cancer treatment alternatives and one opt-out option; and (iii) 8 DCE choice tasks with two CRC screening / prostate cancer treatment alternatives followed by 8 DCE choice tasks with two CRC screening / prostate cancer treatment alternatives and one opt-out option.

Besides the choice elicitation component, the surveys also collected socio-demographic information useful for our current study (the Z vector in Fig. 1). Although we collected more socio-demographic information, we highlight here information about age, education, cancer in family, and positive attitude towards cancer screening and treatment. The respondents were asked validated Likert scale questions related to their decision style (Pachur and Spaar, 2015), health literacy (Ishikawa et al., 2008; van der Vaart et al., 2012), and numeracy as well (Fagerlin et al., 2007; Zikmund-Fisher et al., 2007); these decision-making skills were of interest based on literature, expert opinions and qualitative research, as they all were hypothesized to have an impact on CRC screening and prostate cancer choices. The questionnaire ended with queries about the complexity and length of the questionnaire.

An online sample of $N = 1219$ individuals aged 55−75 years and $N = 1239$ males aged 55−75 years from the Dutch general population, nationally representative in terms of age, (gender), education, and geographic region, was recruited for the CRC screening and prostate cancer treatment condition, respectively, via Survey Sampling International. Calculation of optimal sample sizes for a DCE is complicated as it depends on the true values of the unknown parameters estimated in the choice models (Lancsar and Louviere, 2008). However, based on our DCE design and pilot run, and using the sample size calculation of De Bekker-Grob et al. (2015), a sample size of 1200 respondents per medical area (hence 400 respondents per DCE study) is sufficient to reliably detect preference differences between attribute levels at the 5% significance level.

References

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

Brewer, N.T., Richman, A.R., DeFrank, J.T., Reyna, V.F., Carey, L.A., 2012. Improving communication of breast cancer recurrence risk. Breast Cancer Res. Treat. 133 (2), 553–561. https://doi.org/10.1007/s10549-011-1791-9.

Chorus, C.G., 2010. A new model of random regret minimization. Eur. J. Transp. Infrastruct. Res. 10 (2) https://doi.org/10.18757/ejtir.2010.10.2.2881.

Clark, M.D., Determann, D., Petrou, S., Moro, D., de Bekker-Grob, E.W., 2014. Discrete choice experiments in health economics: a review of the literature. Pharmacoeconomics 32, 883–902. https://doi.org/10.1007/s40273-014-0170-x, 2014.

Clarke, N., Kearney, P.M., Gallagher, P., McNamara, D., O'Morain, C.A., Sharp, L., 2021. Negative emotions and cancer fatalism are independently associated with uptake of Faecal Immunochemical Test-based colorectal cancer screening: results from a population-based study. Prev. Med. 145, 106430.

Collacott, H., Soekhai, V., Thomas, C., Brooks, A., Brookes, E., Lo, R., Mulnick, S., Heidenreich, S., 2021. A systematic review of discrete choice experiments in oncology treatments. Patient. https://doi.org/10.1007/s40271-021-00520-4 online 5 May 2021.

Daganzo, C., 1979. Multinomial Probit: The Theory and Its Application to Demand Forecasting. Academic Press, New York.

Dawson, I.G.J., Johnson, J.E.V., Luke, M.A., 2012. Subjective judgments of synergistic risks: a cognitive reasoning perspective. Br. J. Psychol. 103 (2), 203–223. https://doi.org/10.1111/j.2044-8295.2011.02065.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.

de Bekker-Grob, E.W., Donkers, B., Veldwijk, J., Jonker, M.F., Buis, S., Huisman, J., Bindels, P., 2021. What factors influence non-participation most in colorectal cancer screening? A discrete choice experiment. Patient 14 (2), 269–281. https://doi.org/10.1007/s40271-020-00477-w.

de Bekker-Grob, E.W., Swait, J., Kassahun, H., Bliemer, M., Jonker, M., Veldwijk, J., Cong, K., Rose, J., Donkers, B., 2019. Are healthcare choices predictable? The impact of discrete choice experiment designs and models. Value Health 22 (9), 1050–1062. https://doi.org/10.1016/j.jval.2019.04.1924.

de Bekker-Grob, E.W., Ryan, M., Gerard, K., 2012. Discrete choice experiments in health economics: a review of the literature. Health Econ. 21 (2), 145–172. https://doi.org/10.1002/hec.1697.

Denberg, T.D., Melhado, T.V., Steiner, J.F., 2006. Patient treatment preferences in localized prostate carcinoma: the influence of emotion, misconception, and anecdote. Cancer 107 (3), 620–630.

Fagerlin, A., Zikmund-Fisher, B.J., Ubel, P.A., Jankovic, A., Derry, H.A., Smith, D.M., 2007. Measuring numeracy without a math test: development of the subjective numeracy scale. Med. Decis. Mak. 27 (5), 672–680.

Flynn, P.M., Betancourt, H., Ormseth, S.R., 2011. Culture, emotion, and cancer screening: an integrative framework for investigating health behavior. Ann. Behav. Med. 42 (1), 79–90. https://doi.org/10.1007/s12160-011-9267-z.

Gigerenzer, G., 1996. On narrow norms and vague heuristics: a reply to kahneman and tversky. Psychol. Rev. 103 (3), 592–596.

Gigerenzer, G., Gaissmaier, W., 2011. Heuristic decision making. Annu. Rev. Psychol. 62, 451–482.

Gigerenzer, G., Goldstein, D.G., 1996. Reasoning the fast and frugal way: models of bounded rationality. Psychol. Rev. 103, 650–669. https://doi.org/10.1037/0033-295X.103.4.650.

Hol, L., de Bekker-Grob, E.W., van Dam, L., Donkers, B., Kuipers, E.J., Habbema, J.D.F., Steyerberg, E.W., van Leerdam, M.E., Essink-Bot, M.L., 2010. Br. J. Cancer. https://doi.org/10.1038/sj.bjc.6605566 published online 2 March 2010.

Hutton, D.W., Belkora, J.K., Shacter, R.D., Moore, D.H., 2009. Are patients getting the "gist" in risk communication? Patient understanding of prognosis in breast cancer treatment. J. Cancer Educ. 24 (3), 194–199. https://doi.org/10.1080/08858190902876452.

Ishikawa, H., Takeuchi, T., Yano, E., 2008. Measuring functional, communicative, and critical health literacy among diabetic patients. Diabetes Care. 31 (5), 874–879.

Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47 (2), 263–292. https://doi.org/10.2307/1914185.

Lancaster, K., 1966. A new approach to consumer theory. J. Polit. Econ. 74, 132–157.

Lancsar, E., Louviere, J., 2008. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. Pharmacoeconomics 26, 661–677.

Layard, P.R.G., Walters, A.A., 1978. Microeconomic Theory. McGraw-Hill, New York.

Lie, M.L.S., Lecouturier, J., Harding, C., 2019. Should I stay or should I go? A qualitative study exploring participation in a urology clinical trial. Int. Urogynecol. J. 30, 9–16.

Loken, B., 2006. Consumer psychology: categorization, inferences, affect, and persuasion. Annu. Rev. Psychol. 57, 453–485.

Loken, B., Barsalou, L.W., Joiner, C., 2008. Categorization theory and research in consumer psychology. Handb. Consum. Psychol. 133–165.

Louviere, JJ, Hensher, DA, Swait, JD, 2000. Stated Choice Methods: Analysis and Applications. Cambridge University Press, Cambridge.

Luce, R.D., 1959. Individual Choice Behavior: A Theoretical Analysis. Wiley, New York.

Luce, M.F., Bettman, J., Payne, J., 2001. Emotional Decisions. The University of Chicago Press, Chicago.

Lupi, F., Phaneuf, D., von Haefen, R.H., 2020. Best practices for implementing recreation demand models. Rev. Environ. Econ. Policy 14 (2), 302–323. https://doi.org/10.1093/reep/reaa007.

Marewski, J., Gigerenzer, G., 2012. Heuristic decision making in medicine. Dialog. Clin. Neurosci. 14 (1), 77–89.

McCaughey, D., Bruning, N., 2010. Rationality versus reality: the challenges of evidence-based decision making for health policy makers. Implement. Sci. 5, 39.

McFadden, D., 2006. Free markets and fettered consumers – presidential address delivered at the 117th meeting of the American Economic Association, Boston, January 7, 2006. Am. Econ. Rev. 96 (1), 5–29.

Medin, D.L., Smith, E.E., 1984. Concepts and concept formation. Annu. Rev. Psychol. 35 (1), 113–138.

Pachur, T., Spaar, M., 2015. Domain-specific preferences for intuition and deliberation in decision making. J. Appl. Res. Mem. Cogn. 4 (3), 303–311.

Payne, J., Bettman, J., Johnson, E., 1993. The Adaptive Decision Maker. Cambridge University Press, Cambridge, UK.

Reyna, V.F., 2008. A theory of medical decision making and health: fuzzy trace theory. Med. Decis. Mak. 28, 850–865. https://doi.org/10.1177/0272989x08327066.

Reyna, V.F., 2012. A new intuitionism: meaning, memory, and development in fuzzy-trace theory. Judgm Decis. Mak. 7 (3), 332–359.

Reyna, V., Nelson, W., Han, P., Pignone, M., 2015. Decision making and cancer. Am. Psychol. 70 (2), 105–118.

Reyna, V.F., Lloyd, F.J., Whalen, P., 2001. Genetic testing and medical decision making. Arch. Intern. Med. 161, 2406–2408. https://doi.org/10.1001/archinte.161.20.2406.

Rosch, E., 1978. Principles of categorization. In: Rosch, E., Lloyd, B. (Eds.), Cognition and Categorization. John Wiley & Sons, New York, pp. 27–48.

Rubenstein, A., 1998. Modeling Bounded Rationality. The MIT Press, Cambridge, MA.

Sándor, Z., Wedel, M., 2005. Heterogeneous conjoint choice designs. J. Mark. Res. 42 (2), 210–218.

Simon, H., 1955. A behavioral model of rational choice. Q. J. Econ. 69, 99–118.

Small, K., Rosen, H., 1981. Applied welfare economics with discrete choice models. Econometrica 49, 105–130.

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, 201–226 d o i:0.1007/s40273-018-0734-2.

Stigler, G., Becker, G., 1977. De gustibus non est disputandum. Am. Econ. Rev. 67 (2), 76–90.

Sutkowi-Hemstreet, A., Vu, M., Harris, R., Brewer, N., Dolor, R., Sheridan, S., 2015. Adult patients' perspectives on the benefits and harms of overused screening tests: a qualitative study. J. Gen. Intern. Med. 30 (11), 1618–1626.

Swait, J., 2009. Choice models based on mixed discrete/continuous PDFs. Transp. Res. Part B 43, 766–783.

Swait, J., 2006. Advanced choice models, in valuing environmental amenities using stated choice studies: a common sense approach to theory and practice. Chapter 9, Barbara Kanninen. Springer, Dordrecht, The Netherlands.

Swait, J., Adamowicz, W., 2014. Choosing how best to choose: antecedent Volition and decision process representation in discrete choice models. J. Choice Model. 13, 1–2. https://doi.org/10.1016/j.jocm.2015.01.003.

Swait, J., Ben-Akiva, M., 1987. Incorporating random constraints in discrete models of choice set generation. Transp. Res. B 21B (2), 91–102.

Swait, J., Brigden, N., Johnson, R., 2014. Categories shape preferences: a model of taste heterogeneity arising from categorization of alternatives. special issue on Antecedent Volition, J. Swait W. Adamowicz, Guest Editors J. Choice Model. 13, 3–23. https://doi.org/10.1016/j.jocm.2014.05.003.
Taylor, T.R., 2000. Understanding the choices that people make. J. Am. Board Fam. Pract. 13 (2), 124–133.
Taylor, T., Pack, R., Hilton, G., 2021. No one loves my baby more than me:" A qualitative study of patients' decision-making regarding cannabis use in pregnancy. J. Obstet. Gynaecol. Can. 43 (5), 672 published abstract.
Thaler, R.H., 1985. Mental accounting and consumer choice. Mark. Sci. 4 (3), 199–214.
Tversky, A., 1972. Elimination by aspects: a theory of choice. Psychol. Rev. 79, 281–299.
van Dam, L., Hol, L., de Bekker-Grob, E.W., Steyerberg, E.W., Kuipers, E.J., Habbema, J.D.F., Essink-Bot, M.L., van Leerdam, M.E., 2010. Eur. J. Cancer 46 (1), 150–159. https://doi.org/10.1016/j.ejca.2009.07.014.
van der Vaart, R., Drossaert, C.H.C., Taal, E., et al., 2012. Validation of the Dutch functional, communicative and critical health literacy scales. Patient Educ. Couns. 89 (1), 82–88.
Veldwijk, J., Swait, J., 2022. The role of attribute screening and choice set formation in health DCEs: modelling the impact of benefit and risk attributes. Value Health in press.
Zikmund-Fisher, B.J., 2013. The right tool is what they need, not what we have: a taxonomy of appropriate levels of precision in patient risk communication. Med. Care Res. Rev. 70 (1), 37S–49S. https://doi.org/10.1177/1077558712458541.
Zikmund-Fisher, B.J., Smith, D.M., Ubel, P.A., Fagerlin, A., 2007. Validation of the subjective numeracy scale: effects of low numeracy on comprehension of risk communications and utility elicitations. Med. Decis. Mak. 27 (5), 663–671.