QUANTUM PROBABILITY: A NEW METHOD FOR MODELLING TRAVEL BEHAVIOUR

Thomas O. Hancock (Corresponding Author)
Choice Modelling Centre and Institute for Transport Studies
University of Leeds
tratoh@leeds.ac.uk

Jan Broekaert
Choice Modelling Centre and Institute for Transport Studies
University of Leeds
jan.b.broekaert@gmail.com

Stephane Hess
Choice Modelling Centre and Institute for Transport Studies
University of Leeds
S.Hess@its.leeds.ac.uk

Charisma F. Choudhury
Choice Modelling Centre and Institute for Transport Studies
University of Leeds
C.F.Choudhury@leeds.ac.uk

ABSTRACT

There has been an increasing effort to improve the behavioural realism of mathematical models of choice, resulting in efforts to move away from random utility maximisation (RUM) models. Some new insights have been generated with, for example, models based on random regret minimisation (RRM, μ-RRM). Notwithstanding work using for example Decision Field Theory (DFT), many of the alternatives to RUM tested on real-world data have however only looked at only modest departures from RUM, and differences in results have consequently been small. In the present study, we address this research gap again by investigating the applicability of models based on quantum theory. These models, which are substantially different from the state-of-the-art choice modelling techniques, emphasise the importance of contextual effects, state dependence, interferences and the impact of choice or question order. As a result, quantum probability models have had some success in better explaining several phenomena in cognitive psychology. In this paper, we consider how best to operationalise quantum probability into a choice model. Additionally, we test the quantum model frameworks on a best/worst route choice dataset and demonstrate that they find useful transformations to capture differences between the attributes important in a most favoured alternative compared to that of the least favoured alternative. Similar transformations can also be used to efficiently capture contextual effects in a dataset where the order of the attributes and alternatives are manipulated. Overall, it appears that models incorporating quantum concepts hold significant promise in improving the state-of-the-art travel choice modelling paradigm through their adaptability and efficient modelling of contextual changes.

1. INTRODUCTION

The random utility maximisation (RUM) framework has dominated the travel choice modelling field for many decades. More recently, RUM has been criticised as being inadequate in explaining the full range of behavioural complexity (Chorus et al., 2008; Guevara and Fukushi, 2016). This has resulted in many attempts to better incorporate behavioural concepts into travel behaviour models, including regret (Chorus et al., 2008; Chorus, 2010), contextual relative advantages (Leong and Hensher, 2014) and prospect theory (Avineri and Bovy, 2008). However, none of these developments have yet rivalled RUM as the preferred model in real-world applications. This is due to difficulties that quickly arise once a modeller departs from the firm economic foundations of RUM (Hess et al., 2018). Consequently, caution is required if we are to step away from random utility models. Departures to models with similar underlying structures, i.e. those with the same error structure such as random regret minimisation (Chorus et al., 2008; Chorus, 2010), result in only small differences whilst facing the same key disadvantage of all departures from (linear in attribute) RUM, the loss of the ability to calculate welfare measures (for a further discussion on welfare analysis with non-linear effects, see e.g. Batley and Dekker (2019) and for regret models, see Dekker 2014). Departures to more different models, such as decision field theory (Busemeyer and Townsend, 1992), whilst sometimes finding improvements in model fit, additionally result in models that become computationally infeasible for large-scale datasets (Hancock et al., 2018). Thus, if we are to move away from RUM, we need to investigate alternative approaches that are computationally simpler yet better reflect behavioural realism. This leads us to explore and compare dynamical modelling ideas from other disciplines which are further away from the tried and tested. Given the success of using concepts from quantum physics in cognitive psychology, one possible alternative is to see if quantum physics can make a similar step into travel behaviour modelling. A fundamental aspect of quantum-like models is that they are intrinsically probabilistic. While in, for example, RUM and RRM, a stochastic sampling of the utility function is added to the model to produce probabilistic output, a quantum-like choice model instead implements the stochasticity at the foundation of the decision process *in the mind of the individual decision-maker*.

Quantum physics, first considered in the early 20th century, was originally developed to account for phenomena and results that could not be explained by classical theories of probability and physics. In particular, physicists noticed that the measurement of one variable could impact the measurement of another. One of the archetypal experiments of quantum mechanics is the double-slit scattering experiment (Feynman et al., 1965). In this experiment, either a beam of light or particles is projected on a screen with two fine parallel slits. Further behind the slit-screen, the intensity of the scattered beam is measured across the receptor screen, in the same direction as the slits are separated (Figure 1). When light waves are applied in this experiment, an intensity pattern emerges on the receptor screen which shows both diffraction, caused by light scattering within each slit, and interference, caused by compounding the light waves coming from the two different slits. The major surprise comes from the fact that these same patterns also occur for beams of material particles - e.g. neutrons or heavy molecules. Even when the source is so scarce that one can legitimately assume that only a single particle at once crosses the two slits, a diffraction and interference pattern will still over time build up on the detection screen. Not only does the double-slit experiment prove that matter behaves as waves, but it also shows that a single particle can have non-compatible properties - like being at different locations A and B at the same time. This potential for 'non-locality' of a particle ensues the particle must be at position A and B at the same time in order to produce wave-like interference pattern over the positions C on the detector screen (Figure 1). It has been shown that as soon as the presence of the particle at either of the slits is positively confirmed for sure, the interference pattern on the detection screen disappears.[1]

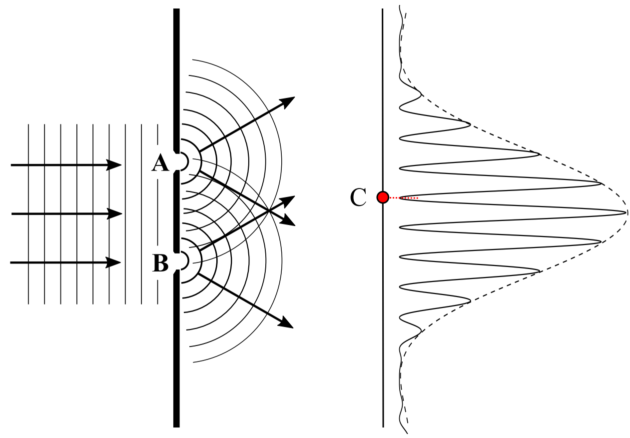

FIGURE 1 : The double-slit experiment (top view). Plain waves are scattered through the double-slit screen (A and B) and produce a diffraction and interference pattern of intensity on the detection screen (C, located at a zero point of the intensity). The dashed line indicates the intensity resulting from single-slit diffraction.

In the double-slit experiment, we can measure for the outcome of the (statistical) propositions $\{a,b,c\}$.

a : 'the particle is present at A'

b : 'the particle is present at B'

c : 'the particle is *never* present at C' (C is at a zero-point of the interference pattern)

Let us suppose single particles repeatedly and individually enter the double-slit device (proposition a or b is true) and we do not observe a particle at C (proposition c is true). Then the outcomes for $(a \vee b)$ and c are both true. Then, by the distributivity law of classical logic, this means that $((c \wedge a) \vee (c \wedge b))$ is also **true**. However, if we evaluate these two expressions according to the procedures of quantum probability, they both are false. First, the proposition $(c \wedge a)$ is false because the procedure shows a particle is never observed at C if we have an affirmative observation of the particle at A, while a non-observation of the particle at location A must set the particle affirmatively at location B and again excludes proposition c. By the same token the $(c \wedge b)$ is false, since never observing a particle at C means the particle cannot have been observed at either of the slits. Thus, for this experiment, $(c \wedge a) \vee (c \wedge b)$ is **false**, an explicit contradiction of the outcome of $c \wedge (a \vee b)$. This particular example clearly exposes that the classical distributivity rule of 'and' and 'or' does not apply for non-compatible features in quantum theory.

These findings resulted in the creation of a new theory of probability, known as quantum logic (Birkhoff and Von Neumann, 1936). Under quantum logic (which is also known as quantum probability), a new set of probability rules were defined, which crucially did not include the axiom of

[1]See also Englert (1996) and Greenberger and Yasin (1988) for the expression of the gradual relation between interference visibility and position predictability.

distributivity. This new theory of probability has subsequently made the transition into cognitive psychology (Bruza et al., 2015) and has also been introduced into transport behaviour modelling. For example, Vitetta (2016) introduced a quantum model based on random utility models with the addition of an interference term for route choice problems. Additionally, Yu and Jayakrishnan (2018) demonstrated that quantum cognition models can be used effectively to capture the difference in state of mind between choices made under stated preference and revealed preference settings. However, as far as the authors are aware, there has not been an actual choice model developed with quantum concepts that incorporates attribute values for individual alternatives and can work for general choices as well as 'changes in perspective'. Thus, the focus of this paper is to explore ways to develop a choice modelling framework based on quantum probability theory that can be used for choices in general, as well as efficiently capturing effects caused by ordering and context, by engendering interference and rotation effects which adequately reflect the changes in the 'state of mind' of the respondents.

In our present study, we will present two quantum models using distinct approaches. The first model, named the *'amplitude model'*, is an innovative approach related to geometrically based quantum-like models. In (all) quantum-like models the belief-action state of a respondent is described by a vector in a Hilbert space. The amplitude components of the vector represent the latent motivation to choose each of the alternatives. In essence, the 'amplitude model' implements the expressions of utility (or regret) immediately in the amplitudes of the belief-action state of a respondent. As such, the amplitude model puts the support for each of the alternatives in a trial directly at the level of a measurable quantity, the probability (amplitude).

The second model, designated as the *'Hamiltonian model'*, is based on a dynamic principle in which the change of the belief state results from attribute comparisons of the alternatives. In this model, therefore, the 'deliberation process' itself is implemented. The dynamic approach to quantum-like modelling uses the 'energy operator', or Hamiltonian, of quantum mechanics to implement the change of the belief state over time. The changes are caused by the information in (and effects from) the input, such as descriptions, questions and choice alternatives or other presented sensory resources (Pothos and Busemeyer, 2009; Atmanspacher and Filk, 2010; Martínez-Martínez, 2014; White et al., 2014). In our present study the expressions of utility or regret are implemented in the phenomenological Hamiltonian. This Hamiltonian then causes the evolution of the initial belief-action state of the respondent towards the informed state in which the decision is made.

The remainder of this paper is organised as follows. First, we introduce quantum probability theory and discuss the relative benefits of using such a system. We then mathematically describe quantum probability theory, discussing how it can be incorporated into a choice model and detailing two different formulations for new models. We next test the performance of our proposed models against typical choice models such as multinomial logit, random regret minimisation and also decision field theory, in the context of travel decisions. Finally, we test the use of 'quantum rotations' on best-worst and contextual choice data, before drawing some conclusions.

2. QUANTUM PROBABILITY THEORY

In this section, we first give a general overview of quantum probability theory. We then give the mathematical definitions for how quantum probability theory works for basic choices. We conclude by describing how it works for a series of related choices. It is in the transformation from one choice task to another that a modelling framework based on quantum probability theory looks very different from traditional choice models.

2.1. Overview of quantum probability theory

A simple example of how quantum probability theory works is given in Figure 2. Initially, a decision-maker might be making a single choice between two alternatives, travelling by car or by train. Each of these alternatives is represented by vectors, $|T\rangle$ and $|C\rangle$ respectively (the axes in Figure 2). Under quantum probability theory, the decision-maker has some belief state, denoted $|z\rangle$, regarding whether they will choose car or train.

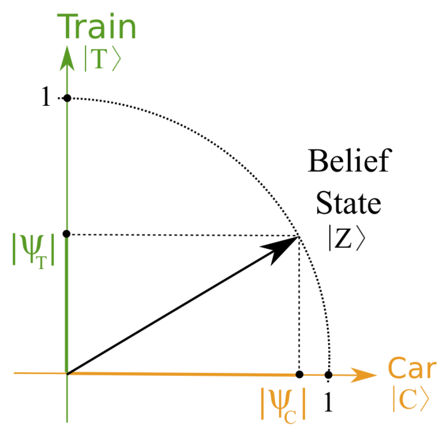

FIGURE 2 : Schematic representation of the belief state in the geometric quantum-like model for a binary choice 'Train' or 'Car'. The belief state $|Z\rangle$ is a superposition of $|T\rangle$ and $|C\rangle$, expressing support for both the choice of Train and Car. The (modulus of the) complex-valued amplitudes of the projections on the respective axes provides the probabilities of each alternative by squaring the projection lengths $|\psi_T|$ and $|\psi_C|$. In this schematic representation, the units on the axes are reals and the normalised belief state is a point on the unit circle.[2] The cosine-similarity of the overall belief state and the 'Train' choice outcome is given by $|\psi_T| = |\langle T|Z\rangle|$.

The action of making a choice (or equivalently coming to some result or making a judgement) results in a reduction of the state. This can be represented graphically by projecting the belief state vector onto the vector corresponding to the chosen alternative. In this example, ψ_T, represents the scalar projection of $|z\rangle$ onto the unambiguous state $|T\rangle$ for choosing the train. The 'length' of this projection is then denoted $|\psi_T|$. In Figure 2, these projections are directly over the corresponding vectors, and on the axes we denoted the norm of the respective amplitudes.

For example, when the belief state vector is at 45 degrees (with respect to the Car and Train axes), the two projections are of equal length and the choice probabilities are thus 50% each. In the example in Figure 2, the car alternative has a higher probability since it shows a larger amplitude than the train component. The full mathematical description for this is given in the following section on a basic choice under quantum probability theory, which also gives a 3-dimensional example. The 'longer' the projection onto the vector for an alternative, the more likely it is for that alternative to be chosen. The crucial difference in using such a system is how an additional question or nudge can impact the decision-maker's choice for the first question (car or train). If,

[2]This schematic representation should not be confused with the Bloch sphere representation for the spin-1/2 particle in quantum mechanics.

for example, the decision-maker was asked 'are you environmentally friendly?' before they had made up their mind between the choice of car or train, they would then be initially answering a different question and making a different choice. As a result of the decision-maker deciding 'I am environmentally friendly', the decision-maker's state moves from the initial starting state and is projected onto the vector representing 'environmentally friendly' and vice versa if they decide 'I am not environmentally friendly' (see Figure 3). This results in making the choice between car and train from a different state. Consequently, the length of the projections ($|\psi_C|$ and $|\psi_T|$) onto the vectors for car and train have changed. This is graphically represented in Figure 3, with the projection length $|\psi_T|$ being longer if the initial state is first projected onto the environmentally friendly vector before being projected onto the train vector, relative to the projection length if train is chosen directly from the initial state. Consequently, the probabilities for choosing car and train are altered.

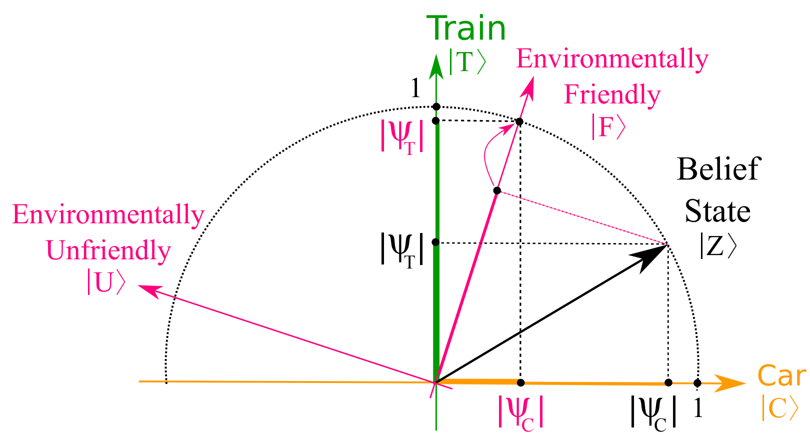

FIGURE 3 : Schematic representation of making two consecutive binary choices under quantum probability theory in the geometric quantum-like model; first 'Environmentally friendly' or 'Environmentally unfriendly', followed by 'Train' or 'Car'. The preparatory ecological question recasts the belief state on the basis $\{|F\rangle, |U\rangle\}$ and will increase the belief support for the choice 'Train' on a *positive* outcome for 'Environmentally Friendly' since the amplitude norm $|\psi_T|$, in pink, is then larger than amplitude norm $|\psi_T|$, in black, and the reverse is true for 'Car'. Notice that while the initial belief state $|Z\rangle$ only had some latent tendency for responding 'Environmentally Friendly', after the positive outcome the updated belief state coincides with the environmentally friendly belief state $|F\rangle$ (pink arrow).

Cognitive psychologists have discussed many key reasons for using quantum probability theory within cognitive modelling (Busemeyer et al., 2011), with many of these reasons also being transferable and relevant within travel behaviour modelling. Firstly, a belief state is most often initially 'indefinite'; it may either have some underlying preference in favour of an alternative or it may express uniform indifference with respect to the alternatives. This may come about due to distorted processing or lack of proper informative input. Furthermore, the final belief state is also in many instances *created* rather than just *recorded* by an effort to measure it. For example, a decision-maker might only start considering how environmentally friendly they are after they have been asked (or reminded) about how environmentally friendly they are (White et al., 2014). For this reason, it is often seen as essential that surveys including both choice tasks and attitudinal questions require the respondent to complete the choice tasks first if the researcher wishes to avoid bias in the choice task (Ben-Akiva et al., 2019). However, conversely, a decision-maker may try to 'justify' their choices with their responses to the attitudinal questions (Cunha-e Sá et al., 2012). Consequently, it is difficult to measure a decision-maker's 'true' attitudes, opinions and preference without some form of bias. It is easy to see how this relates to issues for choice modellers with, for example, analysts often having concerns about the biases or truthfulness within stated preference data (Mahieu et al., 2016).

Secondly, psychologists have put forward the argument that cognition behaves like a rippling wave pattern rather than a classical particle trajectory (Trueblood and Busemeyer, 2012). A decision-maker might consider the advantages of getting the train but then also consider the advantages of driving. Indeed, many models developed in mathematical psychology assume preferences for alternatives that update stochastically (Busemeyer and Townsend, 1993; Krajbich et al., 2012). Under quantum probability theory, preference over time 'behaves like a wave' and consequently exhibits interference patterns and fluctuates over time. It is only when a decision-maker makes up their mind and makes a decision that their preference exists as some measurable definite state. Before an action or choice is made, an observer does not know for sure what the decision-maker will do. There are many preference states within travel behaviour that could similarly be described as 'wave-like', such as anticipating merging onto a new lane when driving, changing travel mode when the weather worsens, or choosing which route to take depending on traffic conditions.

One of the most crucial quantum concepts, however, is the idea of interferences, as e.g. change caused by nudges (such as the previous example of being asked about the environment whilst in the process of making a mode choice). After the development of quantum physics to explain ordering effects of observed variables (Birkhoff and Von Neumann, 1936), a wide range of quantum models, often based on the idea of quantum interference, have been put forward in cognitive psychology (Bruza et al., 2015). These include a quantum model to explain ordering effects (Trueblood and Busemeyer, 2011), a quantum similarity model (Pothos et al., 2013), a quantum judgement model (Busemeyer et al., 2011) and the disjunction effect in the Prisoner's Dilemma (Moreira and Wichert, 2016) and in the two-stage gamble paradigm (Broekaert et al., 2020). These models perform a similar function to choice models that include state dependence, where a number of different models (Seetharaman, 2003) have been applied to capture the temporal correlation of choices over time. Furthermore, should measurement data for both attitude and choice be provided, a higher dimensional representation could be built. Such models have been presented in the literature and have been applied in various contexts; e.g. for choice and confidence level (Kvam et al., 2015), for choice and categorisation (Busemeyer et al., 2009) and choice, confidence and response time (Busemeyer et al., 2006; Kempe, 2003). Given the success of quantum models at explaining ordering effects within cognitive psychology, there is ample scope for quantum logic and quantum ideas within travel behaviour modelling and choice modelling in general.

2.2. Choice making under quantum probability theory

More formally, under quantum probability theory, a measurement (or choice scenario), X, can be related geometrically to a subspace L_x in a multidimensional complex-valued Hilbert[3] space

[3]A Hilbert space is a vector space over the set of real or complex numbers $\mathbb{C}$, (see e.g. Aerts and Gabora 2005). It is the more general form of a Euclidean space, extended to allow for complex numbers and defined over multiple (possibly infinite) dimensions and it is complete; i.e. a space for which convergent sums of vectors are again elements of the vector space. For the work in this paper, our Hilbert space is n-dimensional, where n is the number of choice

(Trueblood et al., 2014b). For each measurement, a number of discrete projection 'events' are possible. These projection events, if mutually exclusive, are related to orthonormal vectors[4] in subspace L_x, which are denoted $|x_1\rangle$, $|x_2\rangle$, ... $|x_J\rangle$ (with J the number of alternatives). For these vectors, we use 'bra-ket' notation in keeping with the standard notation used in quantum mechanics and quantum cognition (c.f. Trueblood and Busemeyer 2011). Under bra-ket notation, a column vector in a Hilbert space is represented by a 'ket' vector, $|\cdot\rangle$, with the corresponding row vector (with each element being complex conjugated) a 'bra' vector, $\langle\cdot|$ (see e.g. Yu and Jayakrishnan 2018). This bra-ket convention simplifies the expression of the inproduct of two states, in particular the squared norm of a complex-valued vector $|Z\rangle$ is then given by the real $\langle Z|Z\rangle$.

These orthonormal vectors, $|x_i\rangle$, then form a basis for the subspace L_x. Consequently, the Hilbert space for a choice task with J alternatives can be represented by a J-dimensional space. This means that for a choice set where there are three alternatives, the Hilbert space is a 3-dimensional space (illustrated in Figure 4).

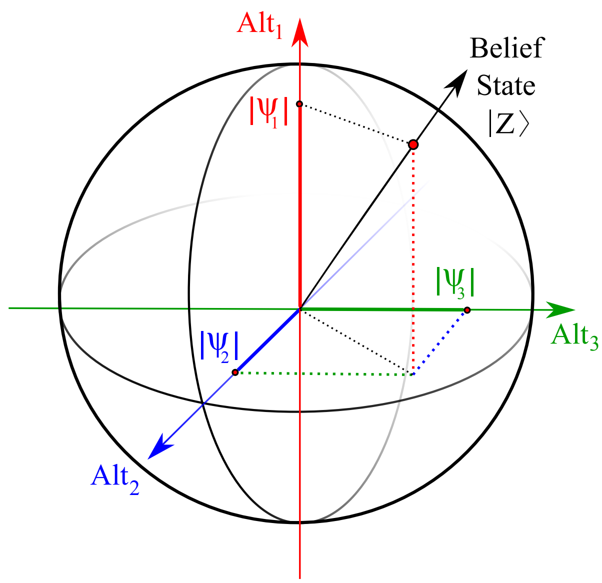

FIGURE 4 : Schematic representation of the belief state in the geometric quantum-like model for a three-choice paradigm {'Alt$_1$', 'Alt$_2$', 'Alt$_3$'} on the unit sphere (see Equation 2). The squared modulus of the amplitude obtained by projection on the axes for each alternative produces the respective probability for that choice $p(\text{Alt}_j) = |\psi_j|^2$.

A basic choice. Under quantum probability theory, a decision-maker has some 'belief state' regarding their preferences over alternatives, which itself is probabilistic (in that a decision-maker inherently has some level of certainty over their preferences and opinions) and is denoted $|z\rangle$, which can be represented by a vector of unit length (see Figure 4). When a decision-maker makes a choice, their state goes from 'indefinite' to 'definite', by projecting onto the vector representing the chosen alternative. This means that for each alternative Alt$_j$, with subspace L_{x_j} there is a corresponding projection operator P_{x_j} - formally $P_{x_j} = |x_j\rangle\langle x_j|$ - to project $|z\rangle$ onto the vector $|x_j\rangle$.

alternatives in a given choice task.

[4]More generally these can have more than one dimension, hence orthogonal subspaces should then be used.

The choice probability, $Pr[j]$, for a specific alternative Alt_j is given by the modulus square of the amplitude for that alternative appearing in the decision-maker's belief state;[5]

$$Pr[j] \;=\; |P_{x_j}\,|z\rangle\,|^2 \;=\; |\left\langle z|x_j\right\rangle|^2 \;=\; |\psi_j|^2. \tag{1}$$

Since we assume the presented alternatives exhaust all possible choices and each alternative is represented by an orthonormal vector (or set of such vectors), the belief state vector must be of unit length:

$$\sum_{j=1}^{J} |\psi_j|^2 = 1, \tag{2}$$

A visual check of this fact appears in Figure 4. The lengths of the three projections can be visualised as the three sides of the cuboid in 3-dimensional space in which Pythagoras' theorem can be applied sequentially.

A sequence of choices. If a decision-maker makes a second choice across a different set of alternatives, this choice may be influenced by the first. Quantum probability theory captures this influence by representing the two measurement events by two separate subspaces within the Hilbert space, L_x and L_y. Each subspace is separately defined by a set of orthonormal vectors representing the alternatives in each measurement event. This means that L_x is spanned by $|x_1\rangle$, $|x_2\rangle$, ... $|x_J\rangle$ and L_y is spanned by $|y_1\rangle$, $|y_2\rangle$, ... $|y_K\rangle$, where there are J alternatives for choice scenario X and K alternatives for scenario Y (while it must be assured that both scenarios span the same Hilbert space).

Revisiting the example presented in Figure 3, a decision-maker might initially be making a choice between commuting by car or train. Under quantum probability theory, the decision-maker has some initial belief state, informed by past experience, regarding whether they will choose car or train. In this measurement event, all possible states are spanned by the basis vectors $|x_{car}\rangle$, $|x_{train}\rangle$. The closer the vector representing the decision-maker's state is to the vector representing an alternative, the more likely it is for that alternative to be chosen. However, the decision-maker could first be asked a different question (Y) about whether they consider themselves to be environmentally friendly or not. In the 'change of perspective' approach of quantum probability theory, the initial belief state does not change under the new perspective under question Y, but the reference frame does. This means that the probabilities for alternatives being chosen in question Y are different from the probabilities for alternatives being chosen in question X because the choice in question Y is represented by a different set of basis vectors, $\left|y_{env-friendly}\right\rangle$, $\left|y_{env-unfriendly}\right\rangle$. Consequently, if the decision-maker makes the choice 'I am environmentally friendly', their belief state moves through the Hilbert space, projected onto the environmentally friendly vector, $\left|y_{env-friendly}\right\rangle$ (see Figure 3). This means that their new belief state is the vector $\left|y_{env-friendly}\right\rangle$ itself. Hence, by making choice in question Y first, the original choice X between car and train is made from a different belief state.

Crucially, by moving their belief state - through what we call a 'quantum rotation' - the size of

[5]Using the bra-ket notation, one can easily see:

$$|P_{x_j}\,|z\rangle\,|^2 = |\,|x_j\rangle\langle x_j|z\rangle|^2 = \langle z|x_j\rangle\langle x_j|x_j\rangle\langle x_j|z\rangle = \langle z|x_j\rangle\langle x_j|z\rangle = |\langle z|x_j\rangle|^2$$

the projected amplitudes onto the vectors for train and for car have changed.[6] As a result, in this example, the decision-maker is more likely to choose to commute by train if they first decide that they are environmentally friendly. This is graphically represented in Figure 3, where the size of the projected amplitude onto the basis vector representing train being chosen has increased, resulting in an increased probability of choosing train.

3. BUILDING A CHOICE MODEL FROM QUANTUM PROBABILITY THEORY

Whilst Lipovetsky (2018) has applied quantum models to consumer recall tasks with multiple alternatives defined on multiple attributes, quantum probability has not ever been applied to multi-alternative, multi-attribute choice scenarios (as far as the authors are aware). In this section, we look at how we can use ideas from quantum probability within a choice model. We do this by first considering what the requirements are for a quantum choice model. Next, we formally define our two alternative quantum-like models, one based on an amplitude approach and the other on a Hamiltonian approach. We then consider how similar or related choice tasks could be mathematically explained by a 'quantum rotation'. Finally, we discuss a number of different value functions that we implement within both standard choice models and our new quantum choice models.

For our choice model to use quantum probability theory, we need to define a method for constructing an indefinite state vector. If this state vector is of unit length and we take projections from it to a set of orthonormal basis vectors (with one vector for each discrete alternative), then the sum of the squared length -more precisely the amplitude- of these projections will equal one. Consequently, for each alternative, we need to find the amplitude of the projection, as the square of this 'length' equals the probability with which the alternative is chosen (see Figure 4). This means that we must first consider how best to represent the state vector, $|z\rangle$.

If, for example, we imagine that we are making a route choice between three alternatives, the development of a state could be represented by Figure 5. When the decision-maker considers factors favouring alternative 1, the state vector extends in the direction of the vector representing alternative 1 (and hence increasing the amplitude of the projection onto alternative 1). Similarly, the decision-maker may consider factors that favour alternatives 2 and 3, resulting in the state vector extending in the direction of the vector for alternative 2 or 3. At some point, the decision-maker reaches some eventual state and makes the actual (probabilistic) choice.[7] To generate this state, we need to know the relative importance of the attributes. This means that one option is to calculate 'value functions' for each alternative. However, if we write the value functions, $V_j = \beta' x_j$, where β is a vector of coefficients and x_j is a vector of observed variables relating to alternative j, then V_j can be positive or negative. As the probability of an alternative is the squared 'length' of the projection from the state vector onto the vector for the alternative, positive and negative values would lead to the same result. This means that care is required when defining how the relative values of the attributes impact the probability amplitudes.

A further requirement for quantum models is some method for capturing underlying preferences towards an alternative. In the representation for the development of an informed state in Figure 5, this simply means having some initial state that is still uninformed by the attributes, but

[6] Two choices that require a different set of basis vectors are known as 'incompatible'. If the choices are in fact compatible and can be represented by the same set of basis vectors, then the order in which the choices are made has no impact on the probabilities of each alternative being chosen. Consequently, quantum probability collapses back into classical probability (Hughes, 1992).

[7] Note that this choice remains probabilistic unless the decision-maker is 100% certain about their choice.

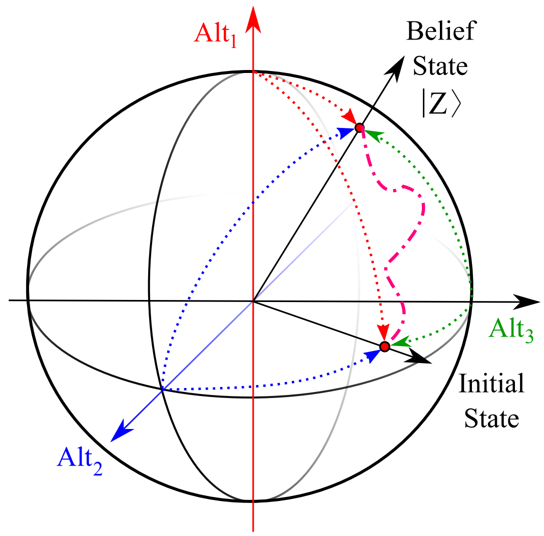

FIGURE 5 : Schematic representation of the development of an informed belief state. The quantum-like dynamical approach lets the 'uninformed' - or possibly biased - initial state evolve to an informed state which leads to the final belief state underlying the probabilistic decision. In the amplitude model this transformation is caused by the subjective utility comparisons immediately in the vector components, while in the Hamiltonian model these utility differences drive the Hamiltonian operator of change over time.

is only based on underlying preferences towards an alternative. Thus, the initial state should be defined on some parameters that act equivalently to attribute specific constants. Then subsequently, from this initial state, the evolution happens when the decision-maker considers the attributes of the alternatives.

3.1. The quantum-amplitude model

Similar to geometric quantum models, the quantum-amplitude model is directed at implementing a specific functionality of the amplitude components of the belief-action state themselves. The innovative approach is to implement value functions for the attributes of the alternatives in the amplitudes. This approach will show an increased optimisation performance since the supporting factors for each of the alternatives are directly expressed in the choice probabilities - through the respective amplitudes.[8]

In geometric approaches, belief states are mostly real-valued vectors of the n-dimensional Euclidean space, e.g. Pothos et al. (2013); van Rijsbergen (2004), or points on the n-dimensional hypersphere.[9]

In the 'quantum-amplitude' approach, we consider the full potential of complex-valued belief

[8]This is in contrast with the Hamiltonian model in which the value functions are implemented in the Hamiltonian components which drive the decision process by progressing the belief state over time and which thus only indirectly produce the choice probabilities.

[9]These belief-state vectors can then be expressed either using generalised spherical coordinates (e.g Lipovetsky (2018); Blumenson (1960)):

$$\psi_1 = \cos\phi_1, \quad \psi_2 = \sin\phi_1 \cos\phi_2, \quad \psi_3 = \sin\phi_1 \sin\phi_2 \cos\phi_3, \quad \ldots$$
$$\psi_{n-1} = \sin\phi_1 \sin\phi_2 \cdots \sin\phi_{n-1} \cos\phi_n, \quad \psi_n = \sin\phi_1 \sin\phi_2 \cdots \sin\phi_{n-1} \sin\phi_n$$

amplitudes to *directly estimate the choice probabilities*. Thus, rather than *evolving* an initial state to a final belief-state vector, in this model we specifically optimise the proper complex amplitudes of the final belief state itself.

For each individual, alternative i in a given choice task has an amplitude, ψ_i, which is estimated as the sum of subjective differences between it and other alternatives j:

$$\psi_i = \left(\delta_i + \sum_{i \neq j} \Delta_{ij}\right) / \sqrt{\mathscr{N}}, \tag{3}$$

where Δ_{ij} is the subjective difference between alternatives i and j (see Section 3.4 for details on the four different value functions that we test to represent this subjective difference) and δ_i is a constant for alternative i which implements the mean impact in the sample of any factors omitted from the specification of the value function for that alternative. This can cover both omitted attributes as well as underlying preferences for specific alternatives. These constants will only take a value of zero if, in the situation of all included explanatory variables taking the same value for the alternatives, the probabilities will be equal. Both Δ_{ij} and δ_i depend on the individual respondent and the task at hand. The normalisation factor (which ensures Equation 2 holds), $\sqrt{\mathscr{N}}$, is obtained from the sum of the squared moduli:

$$\mathscr{N} = \sum_{i}^{J} \left| \delta_i + \sum_{i \neq j}^{J} \Delta_{ij} \right|^2 . \tag{4}$$

Whereas adding the same constant to the utility of every alternative does not have an impact in random utility models, the multiplication of the amplitudes by the same constant does not impact the choice probabilities of alternatives under a quantum system (see Equation 2). Consequently, we can have J parameters to capture the underlying preference towards the J alternatives. The greater the magnitude of these constants, δ_i, relative to the magnitude of the differences, Δ_{ij}, the less deterministic the choices become. Note that from a mathematical point of view we can equivalently estimate the corresponding probabilities, instead of the amplitudes Equation (3), in the model parameter optimisation process (Section 5). Finally, we note that the amplitude model is more general than the cosine similarity model in that it also allows for complex-valued functional expressions, Δ_{ij}, of subjective attribute differences of the alternatives (see section 5.2).

3.2. The quantum-Hamiltonian model

In the search for an adequate dynamical approach to the decision process, quantum *dynamical* elements have proven effective in covering experimental choice paradigms involving ordering and contextuality (Aerts et al., 1999; Busemeyer et al., 2006; Atmanspacher and Filk, 2010; Fuss and Navarro, 2013; Martínez-Martínez, 2014; Asano et al., 2015; Kvam et al., 2015; Broekaert et al., 2017, 2020; Bagarello, 2019). An introductory treatment of this quantum dynamical approach in decision making can be found in Busemeyer and Bruza (2012)'s handbook. At the core of this

---

or by using similarity angles (Pothos et al., 2013) as given by (see Figure 2);

$$\psi_1 = \cos\theta_1,\ \ldots,\ \psi_i = \cos\theta_i,\ \ldots,\ \psi_n = \cos\theta_n$$

in which the similarity cosines must satisfy state normalisation $\sum_{i=1}^{n} \cos^2\theta_i = 1$.

approach is the operator which drives the change of a state vector in quantum theory; the *Hamiltonian*.[10] In contrast to the quantum-amplitude model (see Section 3.1), the 'quantum-Hamiltonian' model thus implements the behavioural decision process as an evolution of the belief state over time. In this dynamical choice model the stochastic process underlying the change of a participant's belief state over time will now be driven through a Hamiltonian which implements the comparison of the attributes of the alternatives. More specifically, the Hamiltonian operator H controls the change of the state vector according to the dynamics of the Schrödinger equation:

$$-i\frac{d}{dt}\Psi \quad = \quad H\Psi, \tag{5}$$

where we have assumed dimensionless expressions for time and 'energy'.[11] The only formal requirement on the Hamiltonian is Hermiticity, $H^\dagger = H$, i.e. the transpose conjugate of matrix H returns H itself. This property assures that the time evolution will conserve the normalisation of the belief state at all times and thus ensure that the choice probabilities across the alternatives add up to 1.

The driving factors of the decision task are formally integrated in the Hamiltonian according to a parametrised Hadamard gate (Busemeyer and Bruza, 2012; Broekaert et al., 2017)

$$H = \begin{pmatrix} h_{11} & \delta_{12} + \Delta_{12} \\ \delta_{12}^\star + \Delta_{12}^\star & -1 \end{pmatrix}, \tag{6}$$

where * indicates the complex conjugate. It should be remarked however that factors in the off-diagonal elements of the Hamiltonian serve a different dynamical function than in the amplitude model (Equation 3). In the Hamiltonian model, the drivers embody pairwise symmetric comparisons of attributes of two alternatives which dynamically compete with each other. For more than two alternatives, we can estimate additional diagonal elements, h_{jj} and each pairwise comparison of alternatives is implemented in a separate parallel process by allocating the drivers to the proper matrix positions;[12]

$$h_{ij} \quad = \quad \delta_{ij} + \Delta_{ij}. \qquad\qquad (i \neq j) \tag{7}$$

In the amplitude model, on the other hand, all pairwise attribute comparisons are immediately summed into the resulting probability amplitude. The driving factors, $\{\delta, \Delta\}$, therefore serve a very different modelling purpose in the two quantum approaches.

The changed belief state at each moment of time is the solution of the Schrödinger equation (see equation 5). This solution can be easily expressed by calculating the propagator:

$$U(t) \quad = \quad e^{-iHt}, \tag{8}$$

[10] A formally very similar dynamic model is provided by continuous-time Markov chain theory in which the operator of change is the transition rate matrix or 'intensity' matrix (see e.g. Busemeyer and Bruza 2012).

[11] In quantum-like modelling in decision making, Planck's constant is set equal to 1 as a standard. This essentially introduces a scale factor to 'time' in the decision process. The Hamiltonian is the generator of change over time but is further devoid of energy connotation.

[12] The Hamiltonian model can be extended to encompass non-symmetric comparison of attributes by doubling the dimension of the Hilbert space. The belief state for each alternative then consists of a two-dimensional subspace.

and applying it to the initial belief state:

$$\Psi(t) = U(t)\Psi(0). \tag{9}$$

In the Hamiltonian model, the unitary time-propagator thus evolves the uninformed - and in general, unbiased - initial belief state $\Psi(0)$ into the evolving informed belief state $\Psi(t)$

$$\Psi(0) = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \quad \longrightarrow \quad \Psi(t) = \begin{pmatrix} \psi_1(t) \\ \psi_2(t) \end{pmatrix}. \tag{10}$$

Like in the general quantum-like approach, to obtain the choice probability for a particular alternative in the experimental paradigm, the corresponding subspace projector M_j should be applied and its outcome norm-squared; $||M_j\Psi(t)||^2$. One more crucial formal element in the Hamiltonian formalism for a decision process is thus the *time of measurement*. Since the datasets we cover in our present study do not include reaction times, we can fix this time to $\pi/2$ in accordance with standard time-scaling procedures (Busemeyer and Bruza, 2012).[13]

3.3. Quantum rotations

In Section 2.1 and Figure 3, we demonstrated how a 'change of perspective' could be accomplished by a projection onto a system with rotated axes representing the new context of the decision. In an equivalent active implementation, this change of perspective can be incurred by applying a rotation operation on the belief state itself (whereas a passive implementation would rotate the basis vectors). For the simplest example with two alternatives, this rotation occurs in a 2-dimensional Hilbert space. The quantum generators of such rotations are the Pauli matrices, e.g. (Feynman et al. 1965, Ch.11);

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{11}$$

The rotation operator, R, itself - about axis $\mathbf{n} = (n_x, n_y, n_z)$ and over angle ϑ - is then given by;[14]

$$R \quad = \quad e^{-i\vartheta \mathbf{n}\cdot\sigma}, \tag{12}$$

where $\mathbf{n}\cdot\sigma$ gives some combination of the Pauli factors, with the restriction that $|\mathbf{n}| = 1$. These rotation operations will be applicable in two of the covered experimental paradigms in our present study (see Section 5).

3.4. Value functions: linear difference, asymmetric decay, soft plus

As well as the use of different 'error structures' provided by the different models, we can also improve our models through the use of non-linear value functions to translate objective differences

[13] Note that a more elaborate quantum model, with intermediate and iterated response/no-response reductions, is required to handle response times (Busemeyer et al., 2006; Kempe, 2003)

[14] Using the equivalence

$$e^{-i\vartheta \mathbf{n}\cdot\sigma} = \mathbf{1}\cos\vartheta - i\,\mathbf{n}\cdot\sigma\sin\vartheta,$$

it is easily verified that for $n_y = 1$, one retrieves the classical expression for a rotation matrix in the real plane, e.g. (Busemeyer and Bruza, 2012; Broekaert et al., 2017).

into subjective ones. In this paper, we test four different value functions across our logit and quantum-like models.

Linear Difference function (LD). The first value function we use simply calculates the relative importance of the linear differences in attributes. Thus, for respondent n in choice task t, we define the subjective difference between alternatives i and j as:

$$\Delta_{ij} = \sum_{k=1}^{K} \beta_k \cdot (x_{ik} - x_{jk}), \qquad \text{(LD)} \qquad (13)$$

where $k = 1,...,K$ is an index across attributes, β_k a coefficient for the relative importance of attribute k and x_{ik} and x_{jk} are the values for alternatives i and j for attribute k.

Asymmetric Decaying Linear Difference function (ADLD). The second value function we test is based on the use of drift rate functions from the multi-attribute linear ballistic accumulator model (Trueblood et al., 2014a). The linear ballistic accumulator (LBA), was originally designed within mathematical psychology, and is a model designed to capture both choices and response times (Brown and Heathcote, 2008). In this approach, a decision-maker starts with a random amount of evidence for each alternative. The evidence for each alternative then grows linearly according to a set of drift rates (with one rate for each alternative). The first to reach some threshold is then the chosen alternative. This model was then adjusted for alternatives with multiple attributes (MLBA) and has been used successfully to explain choices between ratings for eyewitness testimony (Trueblood et al., 2014a), consumer and perceptual choices (Turner et al., 2018) and gambling and accommodation choices (Cohen et al., 2017). In the approach for multiple attributes, the drift rates are generated from a normal distribution where the mean drift rates are a function of the attributes of the alternatives. The non-linearity in the specification for the drift rates allows for the explanation of the similarity, attraction and compromise effects. Notably, work such as Guevara and Fukushi (2016) and Hancock et al. (2018) demonstrate that models that can account for these context effects can be effective for understanding travel behaviour. The corresponding non-linear expression for the drift rate is the second value function we test in this paper:

$$\Delta_{ij} = \sum_{k=1}^{K} w_{x_{ijk}} \cdot \beta_k \cdot (x_{ik} - x_{jk}), \qquad \text{(ADLD)} \qquad (14)$$

where $w_{x_{ijk}}$ is a similarity weighting and β_k, x_{ik} and x_{jk} are defined as before. Whilst similar in appearance to regret functions (see Equation 16), this function, rather than using a logarithm, uses similarity weightings. These are defined such that they are an exponentially decaying function of distance (dropping the indices for individual and task):

$$w_{x_{ijk}} = \exp\Big(- \big(\lambda_1 \cdot \left[x_{ik} \geq x_{jk} \right] + \lambda_2 \left[x_{ik} < x_{jk} \right] \big) \cdot \beta_k \cdot |x_{ik} - x_{jk}| \Big), \qquad (15)$$

where the square brackets convert to 0 or 1 according to the conditional test whether attribute value k is larger in Alt_i than in Alt_j, or vice versa. Under MLBA, two different values, λ_1 and λ_2, are used to capture Tversky (1977)'s findings that the subjective similarity between A and B and the subjective similarity between B and A may not be equal. Given that differences between losses and gains have regularly been shown to be important in a transport context (Hess et al., 2008; Masiero and Hensher, 2010; Stathopoulos and Hess, 2012), this is a useful feature for this quantum model as well. Both λ values should be greater than zero to ensure that attributes that are more similar have a higher similarity value $w_{x_{ijk}}$. This results in weights that are between 0 and 1.[15] Whilst MLBA models typically use just a single pair of λ parameters, another option is to have pairs that are specific to each attribute (i.e. λ_{1k} and λ_{2k}), though this would of course lead to a large increase in the number of estimated parameters if there is a large number of attributes.

Softplus function (S+). The third value function we test is derived from 'softplus' functions, which are used for the activation of a node depending on inputs in a neural network (Hahnloser et al., 2000) and are frequently implemented within machine learning (Dugas et al., 2001; Nair and Hinton, 2010; Zheng et al., 2015). This function is better known in choice modelling for their use within regret functions from random regret minimisation (RRM). The deterministic regret (Chorus, 2010) for the difference between two alternatives i and j is:

$$\Delta_{ij} = \sum_{k=1}^{K} \ln(1 + e^{\beta_k (x_{jk} - x_{ik})}), \qquad \text{(S+)} \tag{16}$$

with β_k, x_{ik} and x_{jk} defined as before.

μ-RRM function (μ-RRM). The final value function we use is based on μ-RRM (van Cranenburgh et al., 2015), which is designed to estimate the 'profundity of regret'. It is defined as:

$$\Delta_{ij} = \mu \cdot \sum_{k=1}^{K} \ln(1 + e^{\frac{\beta_k}{\mu} (x_{jk} - x_{ik})}), \qquad (\mu - \text{RRM}) \tag{17}$$

where μ is a parameter that results in the function collapsing to the LD function (Equation 13) if it is arbitrarily large, and to the S+ function (Equation 16) if it is close to a value of 1.

The use of the four different value functions for Δ_{ij} together with Equations (A1, A2), in the Appendix, of the Logit approach, thus correspond to a multinomial logit (MNL), a contextual utility model and random regret minimisation models (RRM, μ-RRM), respectively. We compare these base models against all of these value functions combined with quantum choice models in Section 5.1.

4. DATA FOR EMPIRICAL EXAMPLES

In this paper, we test our different specifications of quantum models on a number of travel behaviour datasets, which we now describe in turn.

4.1. Swiss value of time dataset

The first dataset we use comes from the Swiss value of time study (Axhausen et al., 2008), where 389 participants each make 9 binary route choice tasks. The two alternatives are described by travel cost (CHF), travel time (minutes), headway (minutes) and the number of interchanges required to complete the trip. This is a basic route choice dataset, without the possibility of testing for interference effects, i.e. an absence of conditions that are specifically suitable for quantum models. We include this 'basic' dataset to test how our quantum models perform under basic settings (i.e. when there is no need for a 'quantum rotation'). This allows us to test whether the underlying structure for the quantum models is effective for modelling travel behaviour.

[15] Note that we adjust the drift rate specification and the weighting functions from the standard specification in Trueblood et al. (2014a) to include weights (β_k) for the relative importance of different attributes.

4.2. UK value of time dataset

The second dataset that we use in this paper comes from the most recent value of travel time study conducted in the UK (Batley et al., 2017). This dataset comprises of 15,045 choices between two balanced alternatives, one of which is cheaper and the other faster (SP1 in Batley et al. 2017). This second dataset allows us to consider quantum rotations to understand the impact of a change in the position of the alternatives or the attributes. In the lay-out of the UK value of time paradigm, two travel alternatives are juxtaposed and are ordered according to two variations;

Time on top: 1) the shorter time but more expensive alternative on the left of the page and thus the longer time but cheaper alternative on the right 't/C-T/c', and 2) the longer time alternative on the left of the page and thus the shorter time on the right 'T/c-t/C'.

Different respondents received the same alternatives and orders, but with inverted ordering of the textual formulation ('phrasing') of the time and cost of the alternative. With these adapted formulations of the options, the two alternatives were again presented in both relative positions.

Cost on top: 1) the configuration with shorter time alternative on the left 'C/t-c/T', and 2) the configuration with longer time alternative on the left 'c/T-C/t'.

The aggregated respondent preferences given in Table 1, show the option order variation[16] to have a significant influence on choice.[17]

TABLE 1 : Observed choice share for alternative 1 in UK-Context paradigm

	Option Order 1	Option Order 2
Textual Order 1	0.495	0.474
Textual Order 2	0.517	0.473

Initial tests suggest that the option order shows a bias effect on the choices made by the respondents, with $\chi^2(1, N = 15045) = 15.884, p = 6.735e-5$. However, we see that the textual order does not, with $\chi^2(1, N = 15045) = 1.628, p = 0.280$. These effects cannot however be disentangled from the impacts of changes in attributes levels in choice tasks, as whilst a balanced design is used to create the choice tasks, the attribute levels are based on a reference trip, meaning that contextual effects can only be disentangled through the estimation of models jointly incorporating the impact of all attributes.

4.3. UK best-worst dataset

The third dataset uses the best-worst format, allowing us to test quantum rotations for their ability to capture both best and worst choices simultaneously. The best-worst dataset we use comes from a survey asking public transport commuters living in the UK to make a set of ten choices between three route alternatives in a stated preference survey. Each choice task involves an invariant reference trip and two hypothetical alternatives. In each instance, the first alternative corresponded to the current respondent-specific conditions. The attributes of the two other alternatives are pivoted around the attributes of the status quo alternative, where the design process ensured that none of

[16] Note that $\text{Prob}(\text{Alt}_2) = 1 - \text{Prob}(\text{Alt}_1)$ in Table 1.

[17] Notice that order effects in quantum-like modelling have been covered previously for consecutive execution of tasks over time and in varied order of execution. In the current paradigm, the order effect relates to variations of visual presentation and phrasing ordering all in the same instance of time.

the three alternatives dominates. Each alternative is described by six attributes: travel time (in minutes), fare (£), rate of crowded trips, rate of delays (both out of 10 trips), the average length of delays (across delayed trips) and the provision of delay information service (which could be unavailable, available at a cost, or available for free). A total of 391 participants completed 10 choice tasks. The participant's task consists of choosing the best option out of the three presented alternatives, and the worst option out of the two remaining alternatives. As participants choose a best and a worst alternative in each choice task, we have a total of 7,820 choices. For full details of the dataset, readers should refer to Stathopoulos and Hess (2012). Crucially, in this best-worst choice data, a bias can be observed in the respondent choice shares, which are given in Table 2, with respondents tending to choose alternatives 2 or 3 as their least favoured alternative more often than their current trip (alternative 1), $\chi^2(2, N = 7820) = 899.9, p < 2.2e-16$.

TABLE 2 : Joint Choice Probabilities and marginals in UK-Best/Worst paradigm

	Alternative 1 worst	Alternative 2 worst	Alternative 3 worst	Sum
Alternative 1 best	•	0.198	0.149	0.347
Alternative 2 best	0.098	•	0.251	0.349
Alternative 3 best	0.058	0.246	•	0.303
Sum	0.156	0.444	0.400	1

5. EMPIRICAL APPLICATIONS

In this section, we describe the various empirical exercises conducted on the data described in Section 4. We start with basic models, before increasing the complexity of the models. Finally, we consider out of sample validation for quantum rotations. For all models, we use R packages maxLik (Henningsen and Toomet, 2011) and Apollo (Hess and Palma, 2019) for estimation of the log-likelihood functions.

5.1. Basic models: logit, DFT, q-Hamilton, q-amplitude.

For the first test of our quantum models, we use all three datasets (Swiss, UK-Context and UK-Best/Worst). At this point, we do not yet consider quantum rotations, simply focussing on comparing the different modelling approaches as well as testing the impact of using different value functions. Whilst these value functions can incorporate real and imaginary parts for the quantum choice models, we test real-only value functions in this section, with comparisons using imaginary parts in Section 5.2. For all three datasets, we compare the quantum models against multinomial logit (MNL), random regret minimisation (RRM), $\mu - RRM$ and a contextual utility model with ADLD value functions (from MLBA theory, as defined in Section 3.4). All of these models have the assumption of no error correlation across choices (thus all choices are treated as being independent from each other, with no correlation assumed between sequential choices), as at this point, we wish to test the impact of simply changing the value function or changing from a classical error structure (which assumes extreme value errors) to quantum choice models.[18] We also test

[18] We report robust standard errors throughout as the computation of the covariance matrix then also accounts for the repeated choice nature of the data, which generally results in an upwards correction of standard errors (cf. Daly and Hess (2010).

our models against Decision Field Theory (DFT), which was demonstrated to outperform standard choice models in our previous research (Hancock et al., 2018). DFT is a dynamic stochastic choice model under which the preferences for different alternatives update over time within the context of a single choice. For a full description of the model, please refer to the Appendix. We first look at specific considerations for the best-worst data, before discussing model specification more generally, and then presenting the results.

5.1.1. Best-worst data modelling methodology

For the best-worst data, we at this point make the basic assumption that best is the opposite of worst. In a utility context, it is common practice to assume symmetry between best and worst,[19] such that:

$$U(\text{Alt}_{i\ worst}) = -U(\text{Alt}_{i\ best}). \tag{18}$$

For quantum models, however, this translation is not as simple for amplitudes. This is a consequence of using the squared amplitudes to calculate the probability of choice of alternatives (see Equation 1), when negative amplitudes for each projection will result in the same probabilities for each alternative as the corresponding positive projections. In the case of only three alternatives (a regular setting for many surveys), there however exists a simple transformation. Given that there are two alternatives left after choosing the most preferred, the probability of picking one alternative as the second best (or second most preferred) equals the probability of picking the other as the worst. Consequently, given alternatives i and j, we can simply define the amplitudes for alternative i being the worst as:

$$|\psi_{worst_i}| = |\psi_{best_j}|, \qquad \text{(basic inversion)} \tag{19}$$

which we define as a 'basic inversion' as it corresponds to the utility model in Equation 18, in that the factors that determine best and worst choice are identical.

For all models, the decision process of the best-worst choice task can be analysed as a progression of a single encompassing process in which valuations of the first stage of choosing the best alternative are carried over into the subsequent process of choosing the worst alternative. On the other hand, these two stages of choice making can be considered to occur independently of each other without carry-over of previous outcomes. Mathematically, this means that in a 'continued deliberation' approach, utilities or amplitudes are generated using the appropriate value function to estimate the probability of each alternative being chosen as the best. Equations 18 and 19 are then used to generate the probability for the worst alternative directly, without a new evaluation of utilities or amplitudes. In an 'independent' evaluation approach, the utilities and amplitudes are re-evaluated for worst choice, where attribute differences between the alternative chosen as the best and the remaining alternatives are not included (thus, for example, under the amplitude model, Equation 3 would no longer have a summation, simply requiring Δ_{ij} where i and j are the only two remaining alternatives).

5.1.2. General points on model specification

For the models tested in this section, we have the following parameters:

[19] This is an oversimplification, with recent work demonstrating that an alternative is to use a scaling parameter, α, for the difference in scale between best and worst Hawkins et al. (2019). Alternatively, one can use a completely separate set of parameters for best choice compared to worst choice (Giergiczny et al., 2017), a point to which we return in Sections 5.3 and 5.4.

- **All models:** A relative importance parameter (β_k) for each attribute (4, 2 and 8 parameters respectively for the Swiss, UK value of time and UK best-worst datasets.[20])

- **Utility models:** $J-1$ alternative specific constants (1, 1 and 2 parameters respectively for the Swiss, UK value of time and UK best-worst datasets).

- **DFT models:** $J-1$ alternative specific constants (1, 1 and 2 parameters respectively for the Swiss, UK value of time and UK best-worst datasets). All DFT models have 1 additional estimated parameter for the number of preference updating steps, with the UK best-worst model additionally having two feedback matrix parameters, ϕ_1 and ϕ_2. These two parameters were found to be insignificant for the Swiss and UK value of time datasets (in line with previous results for datasets comprised of choices tasks with 2 alternatives, Hancock et al. 2020) and were therefore omitted. Note that as we use attribute scaling coefficients in our DFT models (see Appendix), we fix the error $\sigma_\varepsilon = 1$ for normalisation purposes.

- **Hamiltonian models:** 1 alternative pair constant for the Swiss and UK value of time datasets and 3 for the UK best-worst dataset. We also have $J-1$ Hamiltonian matrix diagonals (1, 1 and 2 parameters respectively for the Swiss, UK value of time and UK best-worst datasets). Finally, the Hamiltonian models for UK best-worst models incorporating 'independent deliberations' has different Hamiltonian matrices for best and worst choice. For best choice, a 3x3 matrix is required, thus 2 parameters are required, whilst for worst choice, we require a 2x2 matrix, meaning that 1 additional parameter is required.

- **Amplitude models:** J alternative specific constants (2, 2 and 3 parameters respectively for the Swiss, UK value of time and UK best-worst datasets).

- **ADLD value function:** 2 additional parameters for the utility model, λ_1 and λ_2. For the Hamiltonian model, we only estimate a single λ as we set $\lambda_1 = \lambda_2$ such that $\Delta_{ij} = \Delta_{ji}$ and the Hamiltonian matrix remains Hermitian. DFT similarly requires $\Delta_{ij} = \Delta_{ji}$, thus only has one λ estimated (See explanation of this in the Appendix). For the amplitude models, we fix one λ parameter to a value of 1, as dividing λ by some value x and multiplying the β parameters by x results in amplitudes that are also multiplied by x (hence normalisation results in exactly the same probabilities and an overspecification if we do not fix a λ).

- **ADLDpA value function:** Has a set of lambdas for each attribute ('ADLD per Attr.'). This results in an insubstantial change in log-likelihood for the Swiss and UK best-worst datasets, thus we only show the results for the UK value of time dataset. As this dataset has two attributes, it has a total of 4 λ-parameters in the utility model, and 2 in the Hamiltonian and DFT models (with the same restrictions applied from above).

- **μ-RRM value function:** 1 additional parameter, μ, which measures the profundity of regret.

[20] Note that the UK best-worst dataset has choice alternatives with 6 attributes. The provision and cost of delay information service are treated separately, and we also have a 'reliability' index, which is the expected delay, defined as the interaction between the average time delay and the rate of delays, which was found to be significant in previous research (Stathopoulos and Hess, 2012).

5.1.3. Estimation results

The results for all of the basic models[21] are given in Table 3, with these results complemented by Figure 6.

For the Swiss dataset, the best model fit is obtained by a quantum amplitude model with an ADLD value function. Notably, there is a high degree of non-linearity, with all ADLD functions resulting in a significant improvement over models with a linear difference value function. If only linear differences are considered, DFT results in the best model fit. Gains over the LD value function are also obtained through the use of S+ and μ-RRM value functions for the quantum amplitude model. Whilst the model results obtained from the quantum Hamiltonian models are similar to those of the utility models, DFT and quantum amplitude models both offer substantial improvements in model fit, for all value functions other than a quantum amplitude model with linear differences.

Similar results are also obtained for the UK value of time data, with DFT again substantially outperforming all other models when linear differences are used. This suggests that a standard DFT model can account for non-linearity, as the difference disappears upon moving to ADLD value functions, for which very similar log-likelihoods are obtained across all four models. The quantum amplitude model again gives us the best model fit across the dataset, through the use of attribute-specific λ parameters in the ADLDpA value function.[22] This addition results in the amplitude model substantially outperforming the other models. The μ-RRM model obtains an estimate for μ that is insignificantly different from 1 for the quantum amplitude model, resulting in an equivalent model fit to that of the S+ value function. We again observe no difference in model fit between RRM, μ-RRM and MNL as there are only two choice alternatives in all choice tasks.

For the UK best-worst dataset, our quantum models do not perform as well. This is particularly the case for continued deliberation (when a single value is used to generate probabilities for best and worst choice), for which the best performing model is the ADLD utility model. Whilst DFT achieves similar model fit, neither Hamiltonian nor amplitude models perform as well as the utility models. Notably, very similar log-likelihoods are obtained for MNL, RRM and $\mu-$RRM. We again observe that DFT models perform best for linear difference value functions, with a substantial difference observed for the UK best-worst independent deliberation models. This advantage is reduced through the use of ADLD value functions, though DFT models still give the best model fit.

Crucially, across all three datasets, the best performing quantum amplitude model achieves a better model fit than the best utility model. In comparison to the quantum models, DFT performs similarly for the Swiss dataset, worse for the UK value of time dataset and better for the UK best-worst dataset. The only substantial difference in model fit in favour of the Hamiltonian model over the amplitude model occurs when linear differences are used for the independent deliberation models. This difference is reversed, however, through the use of ADLD value functions. In Table 4, we also give some parameter estimates for the ADLD models run on the Swiss data (model outputs for more complex specifications of these models are given in Tables 7 and 10 for the UK value of time and UK best worst datasets, respectively). Whilst the outputs from quantum and DFT models

[21] Note that we do not report BIC values here, as for the Swiss dataset, the best-fitting version of each model has the same number of parameters, and for the UK datasets, we provide more complex versions of the models in later subsections.

[22] Note that ADLDpA value functions were tested on both the Swiss and UK best-worst datasets but did not result in a significant improvement in model fit for any of the model types.

TABLE 3 : Log-likelihoods for basic versions of logit, DFT, quantum Hamiltonian and quantum amplitude models across all three datasets

		Model Type							
		Utility		Decision Field Theory		Quantum Hamiltonian		Quantum Amplitude	
Dataset	Value function	pars.	LL	pars.	LL	pars.	LL	pars.	LL
Swiss	LD	5	-1,667.97	6	-1,575.40	6	-1,666.92	6	-1,682.83
	ADLD	7	-1,631.46	7	-1,570.56	7	-1,638.96	7	-1,569.05
	S+	5	-1,667.97					6	-1,587.00
	μ-RRM	6	-1,667.97					7	-1,576.56
UK value of time	LD	3	-9,603.17	4	-9,390.42	4	-9,412.23	4	-9,524.21
	ADLD	5	-9,306.86	5	-9,309.98	5	-9,310.32	5	-9,313.73
	ADLDpA	7	-8,936.61	6	-8,982.05	6	-9,026.91	7	-8,790.80
	S+	3	-9,603.17					4	-9,369.61
	μ-RRM	4	-9,603.17					5	-9,369.61
UK best-worst [continued deliberation]	LD	10	-5,802.67	13	-5,788.36	13	-5,831.05	11	-5,850.54
	ADLD	12	-5,777.57	14	-5,780.56	14	-5,818.60	12	-5,802.22
	S+	10	-5,803.97					11	-5,812.90
	μ-RRM	11	-5,802.25					12	-5,812.90
UK best-worst [independent deliberation]	LD	10	-5,780.34	13	-5,656.16	14	-5,818.74	11	-5,868.75
	ADLD	12	-5,668.36	14	-5,648.68	15	-5,818.74	12	-5,660.84
	S+	10	-5,815.69					11	-5,683.49
	μ-RRM	11	-5,724.44					12	-5,683.49

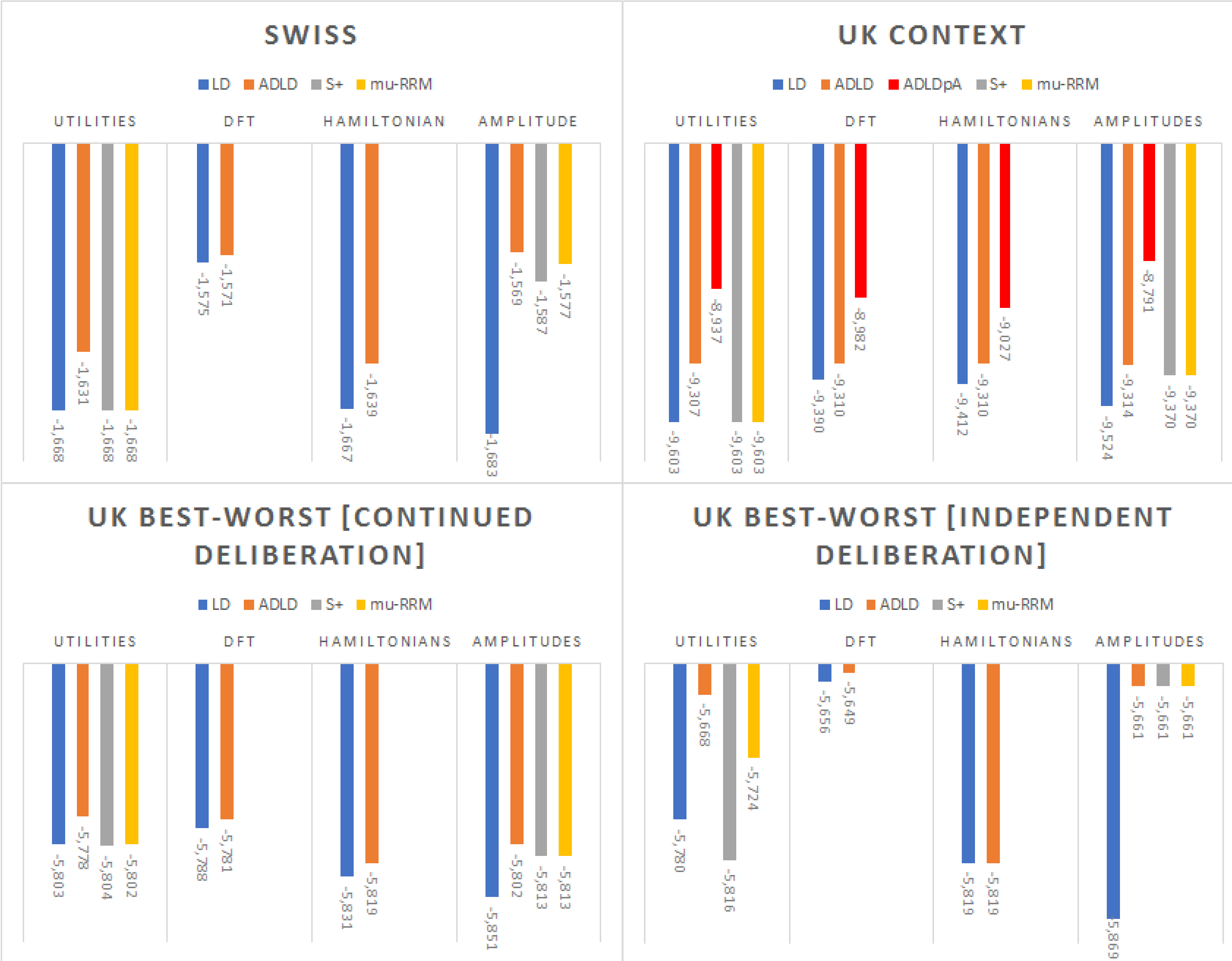

FIGURE 6 : The log-likelihoods of the basic models across the three datasets

cannot be translated into measures such as the value of travel time, we can obtain an indication for the relative importance (RI) of the different attributes by dividing the parameter estimates by the sum of the absolute value of all attribute coefficients. Thus, the relative importance for attribute l is defined:

$$RI_l = \frac{|\beta_l|}{\sum_{k=1}^{K} |\beta_k|}, \tag{20}$$

where $k = 1,...,K$ is an index across attributes. All models find significant estimates with the expected sign for all four attributes, and no significant bias between alternatives 1 and 2. The quantum amplitude model gives similar relative importance weights to the utility model, whereas the DFT and Hamiltonian models give less importance to cost, instead giving a higher weight to the number of changes (β_{CH}). The quantum amplitude model has a better fit than the Hamiltonian model, however, suggesting that differences in attribute importance across the models may not be the driving force behind the differences in model fit.

Additionally, all models find significant estimates for λ_1, which is unsurprising given that all models with ADLD value functions offer a significant improvement in model fit over the corresponding LD value function models. The non-linearity captured by the models utilising ADLD functions is demonstrated for differences in travel time between alternatives in Figure 7. In this figure, the y-axis shows the ‘relative contribution’ to Δ_{ij}, which is defined as Equation 14 but without the multiplication by β_k (thus it is equivalent to $w_{x_{ijk}} \cdot (x_{ik} - x_{jk})$), which allows us to compare the impact of the non-linearity across the models. These results suggest that the quantum models find a stronger damping effect, resulting in greater differences having less of an impact in these models relative to their impact in the DFT and logit models, which have nearly identical satiation rates.

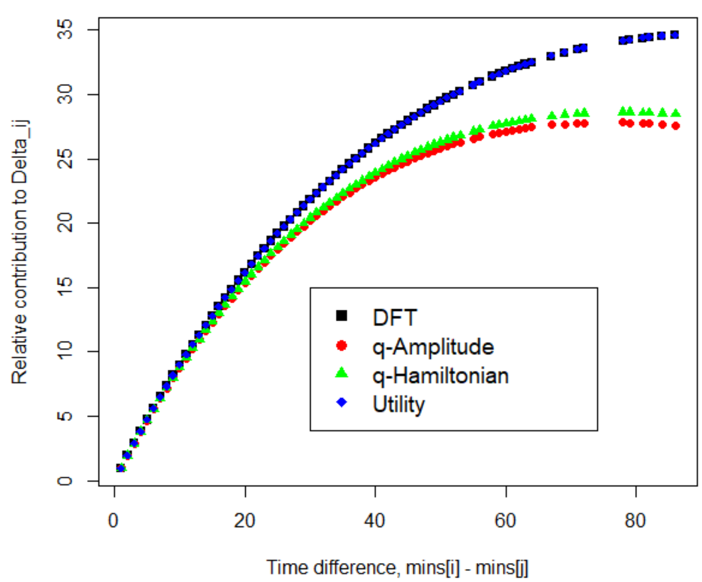

FIGURE 7 : The non-linearity for time differences in the Swiss models

TABLE 4 : Parameter estimates from the models for the Swiss dataset with ADLD value functions, with rel. weight giving the relative importance of the different attributes

		Utility	DFT	q-Hamiltonian	q-Amplitude
Parameters		7	7	7	7
Log-likelihood		-1,631.46	-1,570.56	-1,638.96	-1,569.05
β_{TT}	est.	-0.1231	-3.8682	-0.0300	-0.4071
	rob. t-rat.	-5.00	-9.35	-9.55	-3.82
	rel. weight	8.3%	5.8%	6.3%	6.8%
β_{COST}	est.	-0.3575	-12.7524	-0.0883	-1.3585
	rob. t-rat.	-4.17	-6.55	-5.82	-3.23
	rel. weight	24.1%	19.1%	18.5%	22.6%
β_{HW}	est.	-0.0257	-1.2633	-0.0100	-0.1051
	rob. t-rat.	-7.68	-5.77	-14.54	-4.23
	rel. weight	1.7%	1.9%	2.1%	1.7%
β_{CH}	est.	-0.9773	-48.9351	-0.3496	-4.1473
	rob. t-rat.	-5.94	-6.49	-15.04	-4.42
	rel. weight	65.9%	73.2%	73.2%	68.9%
λ_1	est.	0.7560	0.0027	0.4300	0.0325
	rob. t-rat.	2.29	3.07	7.81	4.35
λ_2	est.	0.0859			1.0000
	rob. t-rat.	0.79			**fixed**
δ_1	est.	-0.0150	-1.1200		0.6400
	rob. t-rat.	-0.37	-0.37		2.06
δ_2	est.	0.0000	0.0000		0.6500
	rob. t-rat.	**fixed**	**fixed**		2.12
δ_{12}	est.			-0.0025	
	rob. t-rat.			-0.22	
σ_ε	est.		1.0000		
	rob. t-rat.		**fixed**		
t	est.		5.9293		
	rob. t-rat. (vs. 1)		5.96		
h_{11}	est.			0.9051	
	rob. t-rat. (vs. 1)			2.37	

5.2. Extension to complex value-functions

Proper to a quantum-like approaches, the value functions (see Section 3.4), which build up the amplitudes, Equation (3) and drive the Hamiltonians, Equation (7), need not to be restricted to real-valued expressions. This increased flexibility allows for interactions between different components within an evaluation of an alternative.

5.2.1. Specification

Whilst there are many different possibilities for how to construct real and imaginary parts within a value function, in this section we consider only simple specifications where the alternative specific constants (δ) and attribute comparisons (Δ) are either real or imaginary. For imaginary parts, we simply multiply the corresponding component by i. Thus, for example, in the Hamiltonian model, Equation 7 becomes $h_{ij} = \delta_{ij} + i \cdot \Delta_{ij}$, for a model with real alternative specific constants and imaginary valued attribute differences. This gives us four alternative specifications for the Hamiltonian models. For the amplitude model, the equivalent model with real alternative specific constants and imaginary valued attribute differences is (for the UK value of time dataset):

$$\psi_1 = (\delta_1 + i \cdot wtc_{12} \cdot tc_{12} + i \cdot wtt_{12} \cdot tt_{12}) / \sqrt{\mathcal{N}} \quad (21)$$

$$\psi_2 = (\delta_2 + i \cdot wtc_{21} \cdot (-tc_{12}) + i \cdot wtt_{21} \cdot (-tt_{12})) / \sqrt{\mathcal{N}} \quad (22)$$

with tc_{12} the difference in travel cost between alternatives 1 and 2, tt_{12} the difference in travel times between alternatives 1 and 2, multiplied by their similarity weightings (which are defined in Equation 15). Finally, we have:

$$\mathcal{N} = |\delta_1 + i \cdot wtc_{12} \cdot tc_{12} + i \cdot wtt_{12} \cdot tt_{12}|^2 + |\delta_2 + i \cdot wtc_{21} \cdot (-tc_{12}) + i \cdot wtt_{21} \cdot (-tt_{12})|^2, \quad (23)$$

which ensures that the sum over the probabilities of each alternative equals 1. We only have two different real-imaginary combinations for the amplitude models, as a result of the use of the norm in Equations (4) and (23), implying $\left| i \cdot \delta_i + \sum_{i \neq j}^J \Delta_{ij} \right|^2 = \left| \delta_i + i \cdot \sum_{i \neq j}^J \Delta_{ij} \right|^2$ and also $\left| i \cdot \delta_i + i \cdot \sum_{i \neq j}^J \Delta_{ij} \right|^2 = \left| \delta_i + \sum_{i \neq j}^J \Delta_{ij} \right|^2$. For the Hamiltonian models, each of the four real-imaginary combinations leads to a different dynamical evolution as a result of Equation (8).

5.2.2. Results

The results of each of these specifications is given in Table 5 for the ADLD value function model for each dataset.

For the Hamiltonian models, the addition of imaginary differences (Im-Δ) has a negative impact for the Swiss and UK value of time datasets, resulting in a far inferior model fit. The alternative specific constants (which are not significant for the Swiss dataset, see Table 4) have little effect on the model, with the consequence that there is little impact on model fit by changing from real to imaginary alternative specific constants. For the UK best-worst dataset, we observe far superior model fits through the use of imaginary attribute differences, with the improvement for the independent deliberation model resulting in the Hamiltonian model becoming more similar in model fit in comparison to the quantum amplitude, utility and DFT model results for the same data. For the amplitude models, we observe a better model fit in all cases for real-real/imaginary-imaginary combinations. Overall, these results suggest that there is ample scope for future exploration of alternative specifications of real and imaginary parts within quantum choice models.

TABLE 5 : Results from models incorporating real and imaginary parts for both q-Hamiltonian and q-Amplitude models

	Re-δ Re-Δ		Re-δ Im-Δ		Im-δ Re-Δ		Im-δ Im-Δ	
	q-Hamiltonian		q-Hamiltonian		q-Hamiltonian		q-Hamiltonian	
Swiss	7	-1,638.96	7	-1,785.75	7	-1,638.88	7	-1,785.89
UK value of time	6	-9,026.91	6	-9,395.86	6	-9,028.27	6	-9,394.86
UK best-worst [continued]	14	-5,818.60	14	-5,792.70	14	-5,814.29	14	-5,811.96
UK best-worst [independent]	15	-5,818.74	15	-5,684.40	15	-5,762.21	15	-5,810.28
	q-Amplitude		q-Amplitude		q-Amplitude		q-Amplitude	
Swiss	7	-1,569.05	7	-1,569.56	7	-1,569.56	7	-1,569.05
UK value of time	7	-8,790.80	7	-8,967.92	7	-8,967.92	7	-8,790.80
UK best-worst [continued]	12	-5,802.22	12	-5,834.97	12	-5,834.97	12	-5,802.22
UK best-worst [independent]	12	-5,660.84	12	-5,673.93	12	-5,673.93	12	-5,660.84

5.3. Models with quantum rotation: contextual and ordering effects in the UK value of time dataset

Thus far, we have only implemented quantum choice models without the use of quantum rotations. In Section 3.3, we demonstrate how 'a change of perspective' in a decision task can be represented within a quantum choice model by performing a quantum rotation on the belief state. In this section, we give the results of models incorporating rotations for contextual and ordering effects in the UK value of time dataset.

This theoretically works well for the UK value of time dataset, which has some choice sets with the cheaper alternative shown first, and some with the faster alternative shown first, as well as having cost sometimes on top and sometimes at the bottom. Whilst we could again use a full set of different parameters for the four different scenarios, these ordering effects have previously been found to be significant (Hess et al., 2017), making this an appropriate dataset to test quantum rotations.

5.3.1. Model specification

Contextual effects such as described above can be captured in both our amplitude and Hamiltonian quantum models through a supplementary rotation for a switched option order and/or switched textual representation. The sizes of the two rotation angles give an immediate process-based assessment of the location order bias and the textual phrasing bias.

In the Hamiltonian model, there is a Hamiltonian operator for the basic change in belief state due to the different values of the attributes of the two alternatives, and additionally an effect from a supplementary rotation for switched option order and for switched textual representation. The order effect is implemented by a rotation of the belief state $\Psi = (\psi_1, \psi_2)$ in the Hilbert space where the first component ψ_1 represents the belief amplitude for alternative 1 and the second amplitude ψ_2 sustains the choice for alternative 2. Thus, for the reference configuration, 'option order' (OptOrd=1) and 'time-cost order' (TCOrd=1), we have a basic Hamiltonian (H_{11}) set up as before, based on Equation 6. This is again a basic parametrised Hadamard gate, that is commonly used to implement the dynamics in a binary choice (Busemeyer et al., 2011). When the time-cost order remains as in the reference configuration (TCOrd=1) while the option order is switched (OptOrd=2), the basic Hamiltonian is complemented with a small rotation over an angle ϑ_{LR} to implement the bias for option ordering. When on the other hand, only the time-cost order is changed with respect to the reference configuration and the option order remains unchanged (TCOrd=2, OptOrd=1), the basic Hamiltonian is again complemented with a rotation but with different angle size ϑ_{Phr}. When both time-cost order and the option order are changed with respect to the reference configuration, (OptOrd=2, TCOrd=2), the basic Hamiltonian is now complemented with the effect of both rotations with the combined angle $\vartheta_{LR} + \vartheta_{Phr}$. Thus, the respective Hamiltonians are given by:

$$H_{11} = \begin{pmatrix} h_{11} & \delta_{12} + \Delta_{12} \\ \delta_{12}^{\star} + \Delta_{12}^{\star} & -1 \end{pmatrix} \tag{24}$$

$$H_{12} = H_{11} + \vartheta_{LR}\ \sigma_y \tag{25}$$

$$H_{21} = H_{11} + \vartheta_{Phr}\ \sigma_y \tag{26}$$

$$H_{22} = H_{11} + (\vartheta_{LR} + \vartheta_{Phr})\, \sigma_y. \tag{27}$$

Note that this is equivalent to adjusting the off-diagonals of H_{11}, with, for example, the upper and lower off-diagonals of H_{12} being $\delta_{12} + \Delta_{12} - i\ \vartheta_{LR}$ and $\delta_{12}^{\star} + \Delta_{12}^{\star} - i\ \vartheta_{LR}$, respectively. Consequently, for this implementation of the Hamiltonian model, the process of attribute comparison and ordering bias occurs simultaneously (addition on Hamiltonian level). The choice process could also be modelled sequentially by separating the Hamiltonian evolution operator, Equation (8), from the rotation operator, Equation (12), and applying them consecutively to the initial belief state. This method can also be used for the amplitude-approach, under which the attribute values of the two alternatives initially determine the basic reference probability amplitude, while dedicated rotations implement the bias process for switched option order and for switched textual representation. Thus, in principle, we implement the same respective rotation operators (based on Equation 12), for time-cost ordering and option ordering;

$$R_{Phr} = e^{-i\vartheta_{Phr}\sigma_y}, \tag{28}$$

$$R_{LR} = e^{-i\vartheta_{LR}\sigma_y} \tag{29}$$

in both the amplitude model and Hamiltonian model.

We compare different levels of complexity for each of the different model structures. We test three variations for contextual ‘changes in perspective’ for both the Hamiltonian and amplitude models as well as trying ‘separate’ parameter models where a set of attributes (based on the ADLDpA basic models) are estimated for each of the four contextual framings. Thus, we have the following five specifications:

1. A basic model. These models correspond to those given by the ADLDpA models in Table 3.

2. A model where a shift is made to Δ_{12}, such that the constant δ_{LR} is added if (OptOrd=2) and the constant δ_{Phr} is added if (TCOrd=2). For the Hamiltonian model, this corresponds to Equations (24-27) with the use of σ_x in place of σ_y.

3. A model where an *imaginary* valued shift is made to Δ_{12}, such that $i \cdot \delta_{LR}$ is added if (OptOrd=2) and $i \cdot \delta_{Phr}$ is added if (TCOrd=2). For the Hamiltonian model, this corresponds to Equations (24-27). The use of imaginary numbers here means that this version cannot be implemented in the utility and DFT models.

4. A model with the application of a quantum rotation of R_{Phr} as defined by Equation 28 if (TCOrd=2), and a rotation of R_{LR} (see Equation 29) if (OptOrd=2). This results in the estimation of two additional parameters, ϑ_{Phr} and ϑ_{LR}.

5. A 'separate' model, in which the basic models are applied separately to subsets of the data corresponding to each contextual scenario. As we have four different scenarios, this results in a fourfold increase in the number of parameters for each model.

5.3.2. Results

The results of all possible specifications are given for each of the four model frameworks in Table 6.

TABLE 6 : Log-likelihood and BIC performance of Logit, DFT, Hamiltonian and Amplitude models for the UK-context paradigm, with models 2-5 also giving the improvement in log-likelihood over model 1.

		Logit	DFT	q-Hamiltonian	q-Amplitude
[1] Basic model	pars.	7	6	6	7
	LL	-8,936.61	-8,982.05	-9,026.91	-8,790.80
	BIC	17,941	18,022	18,112	17,649
[2] Context shifts added to Δ_{12}	pars.	9	8	8	9
	LL	-8,927.59	-8,976.53	-9,018.09	-8,783.64
	LL improvement	*9.01*	*5.52*	*8.82*	*7.16*
	BIC	17,942	18,030	18,113	17,654
[3] (Im) Context shifts added to Δ_{12}	pars.			8	9
	LL			-9,016.93	-8,789.96
	LL improvement			*9.97*	*0.84*
	BIC			18,111	17,666
[4] Rotation operators, R_{Phr} and R_{LR}	pars.			8	9
	LL			-9,017.76	-8,779.71
	LL improvement			*9.14*	*11.09*
	BIC			18,112	17,646
[5] Separate pars.	pars.	28	24	24	28
	LL	-8,909.48	-8,959.43	-8,990.00	-8,769.81
	LL improvement	*27.12*	*22.62*	*36.91*	*20.99*
	BIC	18,088	18,150	18,211	17,809

For all specifications, the amplitude model outperforms the utility model, which in turn has a better model fit than DFT and the Hamiltonian models. For all model frameworks, it appears that using separate sets of parameters instead of a basic model results in an improvement in model fit but a worse BIC. Whilst the quantum rotation models are not as successful as capturing the difference between the contextual situations as models with separate parameters, these models return favourable BICs as they only have two additional parameters. Notably, the best performing Hamiltonian model (in terms of BIC) implements an imaginary shift, as defined by Equations (24-27), and the model with the overall best BIC value is obtained with an amplitude model with rotation operators. The key parameter outputs for these models are given in Table 7.

TABLE 7 : Key model outputs from the context models with shifts or rotations for a change in context

Model	Version	Gain in Log-likelihood	Rel. importance of time (£/hour)	$\delta_{LR}/\vartheta_{LR}$ Estimate	$\delta_{LR}/\vartheta_{LR}$ Rob. t-rat.	$\delta_{Phr}/\vartheta_{Phr}$ Estimate	$\delta_{Phr}/\vartheta_{Phr}$ Rob. t-rat.
Utility	2	9.01	6.49	-0.1467	**-3.81**	0.0464	1.45
DFT	2	5.52	3.58	-0.3133	**-2.99**	0.0957	1.13
q-Hamiltonian	2	8.82	3.63	-0.0479	**-3.88**	0.0042	0.41
	3	9.97	3.65	-0.0995	**-3.61**	0.1022	**3.02**
	4	9.14	3.63	-0.0419	**-3.85**	0.0083	0.92
q-Amplitude	2	7.16	5.26	-10.3154	**-3.52**	2.3966	0.97
	3	0.84	5.26	-38.6215	-0.69	59.7161	0.71
	4	11.09	5.25	-0.0342	**-4.29**	0.0147	**2.09**

Crucially, all but one model show a negative estimate for $\delta_{LR}/\vartheta_{LR}$, which means that the probability of alternative 2 increases when we move from option order 1 to option order 2. This result is in line with the original test and confirms the presence of a shift in left-right bias as a result of changing whether the cheaper and slow alternative appears on the left or the right. For most of the models, we also confirm that there is no effect of changing the order of appearance for the attributes. Table 6 also gives the 'relative importance of travel time with respect to travel cost', which is defined as the ratio of the time parameter estimate divided by the cost parameter estimate, multiplied by 60 (see further detail in Hancock et al. (2018) on how this measure can be interpreted). Whilst this does not correspond to welfare measures (as all models use asymmetric decay functions), these values give us an indication as to whether a decision-maker will more likely choose a fast or cheap alternative. In comparison to the utility model, both DFT and quantum models assign a lower importance to travel time.

By considering the parameter outputs for version 4 (quantum rotation) models, we can also calculate the estimated rotation matrices R_{Phr} and R_{LR}. For the Hamiltonian (Ham) and amplitude (Amp) models, the rotation matrices for changing from having the cheaper alternative on the left (first) to on the right (second) are:

$$R_{LR_{Ham}} = \begin{bmatrix} 0.999 & -0.042 \\ 0.042 & 0.999 \end{bmatrix}, R_{LR_{Amp}} = \begin{bmatrix} 0.999 & -0.034 \\ 0.034 & 0.999 \end{bmatrix}, \tag{30}$$

and the estimates for the quantum rotation matrices for changing from having the travel time first to having the travel cost first are:

$$R_{Phr_{Ham}} = \begin{bmatrix} 0.999 & 0.008 \\ -0.008 & 0.999 \end{bmatrix}, R_{Phr_{Amp}} = \begin{bmatrix} 0.999 & 0.015 \\ -0.015 & 0.999 \end{bmatrix}. \tag{31}$$

This results in a shift towards alternative two through the use of R_{LR} and a small shift towards alternative one through the use of R_{Phr}.

5.4. Contextual and ordering effects in the UK best-worst data

The best-worst dataset also presents a suitable paradigm for the implementation of a quantum rotation, as it is possible that the influence of individual attributes may differ between the case of choosing the best alternative and the case of choosing the worst alternative (Giergiczny et al., 2017).

5.4.1. Model specification

In the quantum-like approach, this changed perspective can be obtained by modifying the angle over which the basis vectors representing the choice of the worst alternative, $|\text{Alt}_{i\ worst}\rangle$, $|\text{Alt}_{j\ worst}\rangle$, are rotated with respect to $|\text{Alt}_{i\ best}\rangle$, $|\text{Alt}_{j\ best}\rangle$. This rotation changes the norm of the projected amplitudes (see Figure 3) and thus modifies the probability for choosing the worst alternative, see Equation (1). Formally, the belief state for choosing the worst alternative, Ψ_{worst}, is obtained by applying the rotation matrix R, Equation (12), to the residual belief state after having chosen the best alternative:

$$\Psi_{worst} = R\Psi_{Resid.\ best}, \qquad (\text{Best} - \text{Worst rotation}) \tag{32}$$

where $\Psi_{Resid.\ best}$ is a renormalised vector of the belief state Ψ_{best} over the remaining choice alternatives. One can easily verify that a rotation over an angle $\pi/2$ according to the axis of rotation $n_y = 1$ results in the 'basic inversion' condition, Equation (19). Mathematically, this rotation simultaneously causes a change in both underlying preferences towards alternatives and a change in how deterministic the choice is. In our empirical application, we test the basic inversion, Equation (19), and the more general quantum rotation, Equation (32). Naturally, if there is a mere classical inversion of amplitudes, the parameters will tend towards those that generate Equation (19).

Under both the Hamiltonian and amplitude models, we assume a three dimensional complex Hilbert space for the belief states, $\Psi = (\psi_1, \psi_2, \psi_3)$, in which the respective components constitute the support for the respective alternatives. In the Hamiltonian approach, the decision process for the choice of the best alternative starts from an initial state Ψ_0:

$$\Psi_0 = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}, \tag{33}$$

which can be configured as unbiased $|\alpha| = |\beta| = |\gamma| = 1/\sqrt{3}$. A relative phase can be implemented on the amplitudes to differentiate their engagement with the unitary evolution operator, Equation (8), or a non-process bias with respect to specific alternatives by setting $|\alpha| \neq |\beta| \neq |\gamma|$. The initial belief state is subject to change according to the Hamiltonian for choice of the best alternative:

$$H = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{12}^{\star} & 0 & h_{23} \\ h_{13}^{\star} & h_{23}^{\star} & h_{33} \end{pmatrix}. \tag{34}$$

The expression of the Hamiltonian can be considered as the superposition of three parametrised Hadamard gates which respectively implement the pairwise comparison process of attributes of the alternatives, Equation (7).

In the amplitude model, the summed attribute differences are implemented directly into the probability amplitudes for each alternative:

$$\psi_1 = \left(\delta_1 + e^{i\phi_{bw}}(\Delta_{12} + \Delta_{13})\right)/\sqrt{\mathcal{N}}, \tag{35}$$

$$\psi_2 = \left(\delta_2 + e^{i\phi_{bw}}(\Delta_{21} + \Delta_{23})\right)/\sqrt{\mathcal{N}}, \tag{36}$$

$$\psi_3 = \left(\delta_3 + e^{i\phi_{bw}}(\Delta_{31} + \Delta_{32})\right)/\sqrt{\mathcal{N}}, \tag{37}$$

where $\mathscr{N}$ renders the belief state normalized to 1, similarly to Equation (23). The relative phase ϕ_{bw} between the bias component δ_j and the attribute differences $(\Delta_{ji}+\Delta_{jk})$ implements a difference in processing for both factors.

In both of the quantum models that we implement, we also explore the inclusion of a 'proportion' parameter for the possibility that a respondent would reverse the processing order of choosing the best and worst alternatives. The UK-best/worst survey allows the respondent to either first choose the best alternative and then follow this by selecting the worst between the two remaining alternatives, or vice versa, starting with choosing the worst alternative before then choosing the best out of the remaining two alternatives. The proportion parameter expresses the proportion of choice processes that are taken in reverse order, by weighting the theoretical choice probabilities from models for both orders ('best then worst' and 'worst then best'). Whilst there are many possibilities for specifications for models that implement a 'proportion' parameter, we only test the most basic specification in our empirical application, under which there are no other additional parameters. Thus if a decision-maker considers worst and then best, Ψ_{worst} is generated using $-\delta_{ij}$ and $-\Delta_{ij}$ in place of δ_{ij} and Δ_{ij} in the Hamiltonian model. For the amplitude model, $\Delta_{ij} \neq \Delta_{ji}$, thus we use $-\delta_i$ and Δ_{ji} in place of δ_i and Δ_{ij}. Then, we generate a rotation matrix R that translates best to worst, and use R^{-1} for the translation from worst to best.

In this application, we explore basic models for both 'independent' and 'continued' deliberation assumptions. Given the various extensions to these models discussed above, we consider four further possibilities. This gives us a total of six different specifications for the quantum choice models (and three for the utility and DFT models, which do not implement quantum rotations). These six possibilities are:

1. A basic structure for each model based on independent deliberations, meaning that best and worst choice probabilities are calculated independently with worst choice using only the two unchosen alternatives within the value functions (thus not using the attribute values from the alternative chosen as best). In the first implementation of independent deliberations, we use the same set of estimated parameters for best and worst choice. To estimate the probability for worst choice, we simply calculate the probability of the other alternative being chosen as (second) best. The utility, DFT and amplitude models here correspond to the independent deliberation ADLD models given in Table 3. For the Hamiltonian model, we instead implement imaginary attribute differences,[23] which corresponds to the best performing (independent deliberation) Hamiltonian model from Table 5.

2. The second method again uses the assumption of independent deliberations, but now allows for a completely 'separate' set of parameters for the best choices compared to the worst. This is equivalent to running two separate models where the dataset is split into two subsets: one with only the best alternative choice tasks and one with only the worst alternative choice tasks. Note that whilst this effectively doubles the number of estimated parameters, two DFT parameters are dropped as there is no significant impact of including feedback matrix parameters for the worst choice (as is often the case when choosing between two alternatives, see Appendix). Additionally, the Hamiltonian model estimates 2 diagonal elements for the Hamiltonian for best choice and 1 for the Hamiltonian for worst choice. Consequently, we do not need these 3 parameters twice for separate parameter models of best and worst choice.

[23] Note that all Hamiltonian models in this section are implemented in this way.

3. The third specification is based on the assumption of continued deliberation. This means that the probabilities for best and worst choice are generated simultaneously through the use of a single value function. We assume basic inversions for all models meaning that the same attributes are important for best and worst choice. These models correspond to the continued deliberation ADLD models given in Table 3, with the Hamiltonian model again using imaginary attribute differences, corresponding to the result in Table 5.

4. The fourth model uses continued deliberations, but also allows for the application of a quantum rotation, R (see Equation 32, that is defined by Equation 12). To estimate R, we need to find the relative importance of the three Pauli matrices. In all cases, we find that the impact of including a third Pauli matrix is insignificant, and thus two additional parameters are estimated, ϑ and ω,[24] where $R = e^{-i\vartheta(n_1*\sigma_1+n_2*\sigma_2)}$, $n_1 = sin(\omega)$, $n_2 = cos(\omega)$ and $n_3 = 0$. For the Hamiltonian models, $n_x = 0$, and for the amplitude models, $n_z = 0$. Note that just a single rotation matrix is used here, meaning that the model uses the same rotation to adjust the amplitudes for the two remaining alternatives, regardless of which pair is left. This results in the assumption that the same adjustment happens for ($\mathrm{Alt}_1 \rightarrow \mathrm{Alt}_2$), ($\mathrm{Alt}_1 \rightarrow \mathrm{Alt}_3$) and ($\mathrm{Alt}_2 \rightarrow \mathrm{Alt}_3$).

5. The fifth model is equivalent to the fourth model, except that it also estimates a 'proportion' parameter. This parameter estimates the percentage of decision-makers who 'choose best then worst' or 'worst then best'.

6. The final model is equivalent to the fifth model, with the exception that two different rotation matrices are estimated through the use of two different axes specified by the parameter ω, one for the rotation from best to worst (ω_{bw}), and the other for the rotation from worst to best (ω_{wb}). Additionally, the Hamiltonian model no longer assumes an indifferent initial belief state. Instead, we set $\psi_0 = (1/\sqrt{3}, e^{i\cdot s}/\sqrt{3}, e^{-i\cdot s}/\sqrt{3})$, where s is an estimated parameter. This results in two additional parameters for the Hamiltonian model and one for the amplitude model. The 'worst to best' rotation matrices are then set as $R_{wb} = e^{i\vartheta(n_x*\sigma_x+n_y*\sigma_y)}$, where the weights n_x and n_y are estimated with ω_{wb}. Consequently, $R_{wb} = R_{bw}^{-1}$ if $\omega_{wb} = \omega_{bw}$.

5.4.2. Results

The results of these models are given in Table 8. Given the complex likelihood structure, we use an initial parameter search algorithm based on the heuristic for non-linear global optimisation developed by Bierlaire et al. (2010) in an attempt to reduce the risk of convergence to poor local optima. For all of the quantum models, we try the four different specifications using real and imaginary numbers, as tested in Table 5. For brevity, we show just the best-fitting model in each case, which is a model with Re-δ_{ij} and Im-Δ_{ij} for all of the models that incorporate quantum rotations.

Unsurprisingly, every model finds a significant improvement in model fit by having a separate set of parameters for the best alternatives compared to the worst alternatives (in line with the results of Giergiczny et al. 2017). This suggests that the relative sensitivities to the different attributes for a best alternative are not necessarily the same as the relative sensitivities to the different attributes for a worst alternative. The overall best-fitting model in terms of log-likelihood is the DFT model with separate parameters. However, the quantum rotation models are efficient in parameter use and

[24] The use of ω here within a sine and cosine function ensures that $|\mathbf{n}| = 1$.

TABLE 8 : Results from models for the best-worst dataset

Standard choice models	Deliberation	Utility			DFT		
		pars.	LL	BIC	pars.	LL	BIC
[1] single set pars.	Independent	12	-5,668.36	11,444	14	-5,651.62	11,429
[2] separate pars.	Independent	24	-5,607.63	11,430	26	-5,569.04	11,371
[3] basic model	Continued	12	-5,777.57	11,663	14	-5,780.56	11,687
Quantum choice models	**Deliberation**	**q-Hamiltonian**			**q-Amplitude**		
		pars.	LL	BIC	pars.	LL	BIC
[1] single set pars.	Independent	15	-5,684.40	11,503	12	-5,660.84	11,429
[2] separate pars.	Independent	27	-5,657.22	11,556	24	-5,598.20	11,412
[3] basic model	Continued	14	-5,792.70	11,711	12	-5,802.22	11,712
[4] quantum rotation 1	Continued	16	-5,656.82	11,457	14	-5,742.78	11,611
[5] quantum rotation 2	Continued	17	-5,651.74	11,456	15	-5,612.08	11,359
[6] quantum rotation 3	Continued	19	-5,624.54	11,419	16	-5,611.09	11,366

consequently find good BIC values, with the result that the best BIC is obtained by an amplitude model that uses a quantum rotation. By stepping away from 'best = opposite of worst', the rotation models bring the performance of the continued deliberation models in line with those of the separate parameter independent deliberation models. Consequently, we find that best-worst choices in this dataset are 'incompatible': a quantum rotation is required to move from a set of basis vectors for best choices to a different set of basis vectors for worst choices. Overall, the results from both classical and quantum models suggest that best is not the opposite of worst, which is in line with the biases present in the overall choice shares (in Table 2).

Test of the quantum rotations. Of key importance for the quantum models is to test the impact of the quantum rotation matrices themselves, as the inclusion of these matrices substantially improves both the Hamiltonian and amplitude models. We consider the impact of these matrices by looking at the resulting probabilities generated from the application of the matrix to an initial belief state. As a contrast to the rotation matrices generated by models for the UK value of time dataset, it is not intuitively clear what the impact of these rotation matrices are on complex-valued residual belief states, with the respective rotation matrices for the Hamiltonian (rotation 3 model) and Amplitude models (rotation 2 model) being:

$$R_{Ham_{BW}} = \begin{bmatrix} 0.20+0.13i & 0.97+0.00i \\ -0.97+0.00i & 0.20-0.13i \end{bmatrix}, R_{Ham_{WB}} = \begin{bmatrix} -0.20-0.27i & -0.94+0.00i \\ 0.94+0.00i & 0.20+0.27i \end{bmatrix}. \tag{38}$$

and

$$R_{Amp} = \begin{bmatrix} -0.62+0.00i & 0.24+0.75i \\ -0.24+0.75i & -0.62+0.00i \end{bmatrix}. \tag{39}$$

We thus test these matrices by calculating their impact on the choice probabilities of each respondent of choosing an alternative as worst, given a belief state, $\psi_{Resid.best}$, which corresponds to the renormalised probability amplitudes for the remaining two alternatives following the choice of the best alternative. Vice versa, we check the choice probabilities for best alternative resulting from the application of the inverse rotation on $\psi_{Resid.worst}$, the renormalised state for the remaining two alternatives following the choice of the worst alternative.

This results in Figure 8, in which the amplitudes of the two remaining alternatives - after having chosen best (left panel) and worst (right panel) alternative - are rotated to generate the choice probabilities of the alternative being chosen as worst (left panel) or best (right panel) for each individual respondent. In this figure, the black dots show how the probability changes under a basic inversion (as described by Equation 19). Under a basic inversion, best is the opposite of worst, and as a direct consequence, a belief state of, for example $\psi_{Resid.best} = (1,0)$ generates probabilities for worst choice equal to 0 for the upper positioned alternative, and equal to 1 for the lower positioned alternative (hence swapping the amplitude entries).

In both quantum models, we observe convex and concave transformations of second best (or second worst) choice probabilities into worst (or best) choice probabilities due to the rotation transformation. The impact of this rotation transformation is more easily assessed by checking the image for the value of the initial probability at 0.5, which corresponds to expressing indifference between the two remaining alternatives. In the case of a concave relation, the initial indifference results in the chosen second best (worst) alternative becoming the chosen worst (best) alternative with a higher probability. In contrast, in a convex relation, the probability of chosen second best (worst) alternative will render a lower probability for the chosen worst (best) alternative in comparison to the basic inversion.[25] We can now assess the impact of the quantum rotation in the observed bias effects when choosing the status quo, i.e. Alt_1, as the worst alternative (see Table 2). Choosing the status quo as worst alternative occurs when either Alt_2 or Alt_3 are chosen as best alternatives, hence we must compare p(Alt_3)(vs 1) for worst (blue) with p(Alt_1)(vs 3) for worst (red) and, separately compare p(Alt_2)(vs 1) for worst (purple) with p(Alt_1)(vs 2) for worst (orange). When Alt_2 is chosen as the best alternative, the amplitude model shows a substantial bias effect against the status quo as worst choice in the 'best then worst' order (blue concave, red convex). When Alt_3 is chosen as the best alternative, the rotation produces the bias effect in both processing orders (purple concave and orange convex). This suggests that it is the introduction of the quantum rotation that drives the accurate recovery of the underlying observed choice shares given in Table 2.

Under the Hamiltonian model, the bias effect is not reproduced in exactly the same manner, in particular in the 'best then worst' processing order. When Alt_2 is chosen as the best alternative, the Hamiltonian model renders an increased probability against the status quo as worst by shifting density for p(Alt_3)(vs 1) for worst (blue) towards lower residual probability for Alt_3 being second best, whilst shifting density for p(Alt_1)(vs 3) for worst (red) towards higher residual probability for Alt_1 being second best. In the 'worst then best' processing order, both models appear to use the same relative rotation transformation to produce the bias effect.

We also note that both models show choice order effects, although not necessarily for the same

[25] We notice a different cause of the choice probability p(Alt_1)(vs 3) for worst alternative (red dots) in the Hamiltonian and amplitude model. In the amplitude model, the lowered probability (w.r.t. basic inversion) results from convexity while in the Hamiltonian model it results from density (concentration towards higher probabilities in the residual vector). These density shifts are related to the bias parameter in the initial state in the Hamiltonian model. The same effect is present in the transformation for the choice probability p(Alt_3)(vs 1) for worst alternative (blue dots).

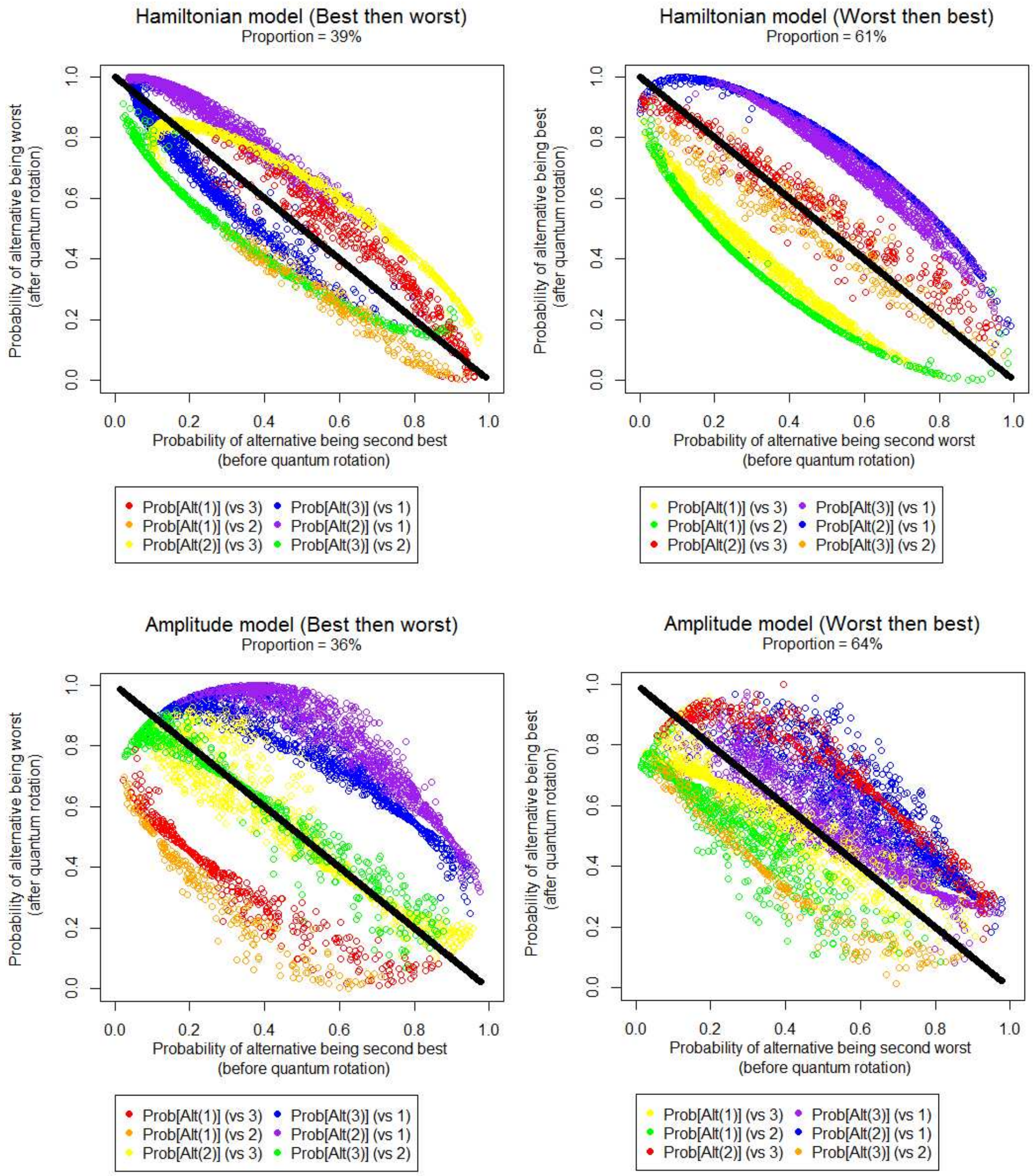

FIGURE 8 : Choice probabilities of individual decision-makers generated by the quantum rotation in the Hamiltonian model rotation [3] (top) and the amplitude model - rotation [2] (bottom) and, choice of best alternative then worst (left) and choice of worst alternative then best (right). The rotation transforms the probability amplitude of the second best choice into the worst choice (left panels), and of the second worst choice into the best choice (right panels). The combinations of chosen alternatives have been consistently colour coded across choice order and models (i.e. red corresponds to Alt_1 as worst, and Alt_2 as best in all four graphs). The 'basic inversion' relation, Equation (19), which switches the *residual* probabilities for best (worst) alternative into worst (best) alternative is marked with black dots and serves to gauge the effect of the optimal quantum rotations.

combinations of best and worst choice. For instance, the combination Alt_2 worst and Alt_3 best does not show significant order effects in the amplitude model, nor in the Hamiltonian model. On the other hand, Alt_1 best and Alt_2 worst is oppositely transformed in both choice orders in the Hamiltonian model, while on the other hand in the amplitude model Alt_2 best and Alt_1 worst is oppositely transformed.

The optimised proportion parameter has shifted the weight moderately towards the 'worst then best' choice order in both models (order ratio 0.39/0.61 for the Hamiltonian model and 0.36/0.64 for amplitude model). This feature is likely a result of the strong bias in the choice for worst alternative which would formally be captured more easily by an immediate implementation in the belief state (for the Amplitude model) and by a dynamical process (in the Hamilton model), than by a rotation. In the 'worst then best' order, both models do not invoke the rotation to render the probabilities for the worst choice, while in the reverse order 'best then worst' the rotation is involved in the final stage of producing the worst choice probabilities.[26]

To analyse the impact of these rotations in relation to the effects they have on model outputs, we consider the overall predicted choice shares across the different models (see Table 9).

TABLE 9 : Observed and predicted choice shares from different best-worst models

Best choice / Worst choice		1 / 2	1 / 3	2 / 1	2 / 3	3 / 1	3 / 2	Average deviation from observed
Utility	[2]	20.4%	14.3%	9.1%	25.8%	6.0%	24.3%	0.5%
DFT	[2]	20.0%	14.3%	9.4%	25.1%	6.3%	24.8%	0.3%
q-Hamiltonian	[3]	20.6%	19.1%	9.9%	19.9%	9.6%	20.8%	3.0%
q-Hamiltonian	[6]	20.1%	14.9%	9.7%	25.0%	6.1%	24.2%	**0.2%**
q-Amplitude	[3]	20.6%	19.1%	9.8%	20.0%	9.6%	20.9%	2.9%
q-Amplitude	[5]	18.8%	14.5%	11.0%	24.9%	6.1%	24.7%	**0.5%**
Observed share		19.8%	14.9%	9.8%	25.1%	5.8%	24.6%	

The average deviation from the observed choice share of best and worst choice demonstrates the impact of the quantum rotation on the quantum models with a quantum rotation. Hamiltonian and amplitude models without a quantum rotation have deviations of 3.0% and 2.9% respectively. These deviations are substantially reduced by moving to versions of the models with quantum rotations, which brings the results in line with those of the utility and DFT models, with the Hamiltonian model in particular almost perfectly capturing the observed choice shares.

Finally, we consider the parameter outputs for the best version of each model in Table 10.

All four models give the expected sign for all of the attributes. Whilst the relative importance of the different attributes is similar across the models, there are some exceptions. In particular, the estimates for travel fare (LF) for the worst choices in the utility and DFT models are very different, with DFT giving the lowest relative importance to travel fare in both best and worst choice. DFT gives a higher importance to the rate of delays (RA) and the provision of a free information service (IFR) than other models. The quantum models provide very similar relative

[26]Further examination of this order effect could be done in a dataset with explicit choice order specifications in the survey.

TABLE 10 : Parameter estimates from the models for the UK best worst dataset, with rel. weight giving the relative importance of the different attributes.

Model		Utility		DFT		q-Hamiltonian	q-Amplitude
Parameters		24		26		19	15
Log-likelihood		-5,607.63		-5,569.04		-5,633.44	-5,612.08
BIC		11,430		11,371		11,437	11,365
choice		best	worst	best	worst	all	all
LL contribution		-3561.75	-2045.88	-3536.743	-2032.30		
β_{TT}	est.	-0.0259	-0.0221	-0.1942	-0.0414	-0.0096	-0.2716
	rob. t-rat.	-5.96	-3.77	-6.55	-3.56	-9.75	-5.85
	rel. weight	**0.5%**	**0.3%**	**0.7%**	**0.5%**	**0.5%**	**0.5%**
β_{LF}	est.	-5.3428	-7.2843	-25.0410	-7.1054	-1.7815	-46.6046
	rob. t-rat.	-4.78	-6.14	-7.84	-4.31	-9.42	-5.69
	rel. weight	**93.0%**	**93.4%**	**89.5%**	**81.6%**	**91.0%**	**92.2%**
β_{CR}	est.	-0.1235	-0.2333	-0.8419	-0.3671	-0.0506	-1.2673
	rob. t-rat.	-4.43	-4.29	-5.57	-3.77	-8.71	-5.58
	rel. weight	**2.2%**	**3.0%**	**3.0%**	**4.2%**	**2.6%**	**2.5%**
β_{RA}	est.	-0.0778	-0.0019	-0.6702	-0.3897	-0.0506	-0.9285
	rob. t-rat.	-3.19	-2.88	-4.04	-3.22	-4.92	-3.69
	rel. weight	**1.4%**	**0.0%**	**2.4%**	**4.5%**	**2.6%**	**1.8%**
β_{RE}	est.	-0.0042	-0.0077	-0.0290	-0.0098	-0.0016	-0.0435
	rob. t-rat.	-1.47	-1.3%	-2.18	-1.47	-2.56	-2.75
	rel. weight	**0.1%**	**0.1%**	**0.1%**	**0.1%**	**0.1%**	**0.1%**
β_{RB}	est.	-0.0101	-0.0025	-0.0760	-0.0239	-0.0037	-0.0792
	rob. t-rat.	-1.75	-0.18	-2.06	-1.24	-2.08	-2.42
	rel. weight	**0.2%**	**0.0%**	**0.3%**	**0.3%**	**0.2%**	**0.2%**
β_{ICH}	est.	-0.0333	-0.0006	-0.2345	-0.1779	-0.0073	-0.5012
	rob. t-rat.	-1.19	-0.90	-1.06	-0.77	-0.71	-1.98
	rel. weight	**0.6%**	**0.0%**	**0.8%**	**2.0%**	**0.4%**	**1.0%**
β_{IFR}	est.	0.1254	0.2494	0.8986	0.5957	0.0529	0.8245
	rob. t-rat.	4.33	2.7%	4.49	3.49	5.2	3.07
	rel. weight	**2.2%**	**3.2%**	**3.2%**	**6.8%**	**2.7%**	**1.6%**
λ_1	est.	0.3272	5.1126	0.0438	0.4302	1.3547	0.0391
	rob. t-rat.	4.38	3.16	3.31	2.74	11.06	4.56
λ_2	est.	1.8508	0.3102				1.0000
	rob. t-rat.	1.69	7.73				**fixed**
δ_1	est.	0.0000	0.0000	1.9037	1.2919		2.8439
	rob. t-rat.	**fixed**	**fixed**	6.55	5.94		5.79
δ_2	est.	-0.3029	-1.0826	1.1384	-0.9907		8.2458
	rob. t-rat.	-2.69	-14.72	4.09	-5.1		7.05
δ_3	est.	-0.5260	-0.7559	0.0000	0.0000		5.7360
	rob. t-rat.	-4.94	-9.63	**fixed**	**fixed**		7.73
δ_{12}	est.					0.0277	
	rob. t-rat.					0.66	
δ_{13}	est.					0.2025	
	rob. t-rat.					9.65	
δ_{23}	est.					-0.1435	
	rob. t-rat.					-5.59	
σ_ε	est.			1.0000	1.0000		
	rob. t-rat.			**fixed**	**fixed**		
t	est.			7.1019	5.4830		
	rob. t-rat. (vs. 1)			8.52	4.53		
ϕ_1	est.			0.0040	0.0000		
	rob. t-rat.			2.67	**fixed**		
ϕ_2	est.			0.3469	0.0000		
	rob. t-rat. (vs. 1)			14.54	**fixed**		
h_{11}	est.					1.6685	
	rob. t-rat.					34.23	
h_{33}	est.					2.1300	
	rob. t-rat.					20.26	
proportion	est.					0.3943	0.3568
	rob. t-rat.					6.71	7.96
ϑ	est.					1.3718	2.2353
	rob. t-rat. (vs. $\pi/2$)					-5.25	13.82
ω_{BW}	est.					0.1346	1.2641
	rob. t-rat.					2.69	25.85
ω_{WB}	est.					0.2788	
	rob. t-rat.					6.21	
s	est.					1.1709	
	rob. t-rat.					18.43	

importances to the utility model. All models suggest that there are differences between best and worst choice, with both the utility model and DFT in particular finding different sensitivities to cost when comparing best to worst. Additionally, both find substantially lower estimates for δ_2 in worst choice compared to best, which is in line with observed choice shares for the 2nd alternative (best-35%, worst-44%). For the quantum models, we observe angles ϑ, and ω estimates, that are significantly different from $\pi/2$ and 0 respectively, which would correspond to best being the opposite of worst for the Hamiltonian model (with ω also significantly different from $\pi/2$ for the amplitude model, equivalently demonstrating that it too suggests that best is not the opposite of worst).

5.5. Validation results: Holdout method

We also test for overfitting, by testing the best performing model (in terms of BIC) for each of the four different types of model in our best-worst data. This corresponds to a separate parameters model for the utility and DFT models, and to models with quantum rotations and a proportion parameter for the Hamiltonian and amplitude models (2 rotations for the Hamiltonian, but just one for the amplitude model). We fit the data to 5 estimation subsets and then estimate the out-of-sample log-likelihood for the remaining validation subset. In each case, 80% of the (participants in the) dataset are assigned to the subset that is used for model estimation, with the remaining 20% used for validation. The log-likelihoods of these models are given in Table 11.

TABLE 11 : The log-likelihood results for the estimation and holdout samples for the different models for the UK best-worst dataset

	Utility		DFT		q-Hamiltonian		q-Amplitude	
	pars.	LL	pars.	LL	pars.	LL	pars.	LL
estimation 1	24	-4,508.98	26	-4,471.93	19	-4,522.45	15	-4,519.58
estimation 2	24	-4,420.34	26	-4,394.00	19	-4,439.09	15	-4,412.87
estimation 3	24	-4,526.41	26	-4,498.69	19	-4,540.49	15	-4,529.31
estimation 4	24	-4,486.07	26	-4,462.50	19	-4,500.03	15	-4,498.81
estimation 5	24	-4,468.13	26	-4,426.69	19	-4,480.68	15	-4,472.40
holdout 1	24	-1,106.20	26	-1,104.27	19	-1,106.51	15	-1,097.21
holdout 2	24	-1,193.77	26	-1,183.34	19	-1,190.84	15	-1,205.68
holdout 3	24	-1,084.95	26	-1,073.84	19	-1,086.96	15	-1,084.99
holdout 4	24	-1,125.21	26	-1,110.85	19	-1,127.82	15	-1,117.01
holdout 5	24	-1,143.46	26	-1,148.45	19	-1,146.99	15	-1,141.51

The results suggest that neither the quantum models nor DFT overfit the data, with DFT giving the best fit in all 5 estimation and 3 of the validation subsets, and the amplitude model having the best fit in the other two validation subsets. The BIC values for these models are given in Figure 9, which penalises the utility and DFT models. This consequently results in the amplitude model obtaining the best BIC value across all 5 estimation and validation subsets.

FIGURE 9 : The BIC results for the estimation and holdout samples for the different models for the UK best-worst dataset

5.6. Elasticities from value of time datasets

In this section, we look at elasticities from the best version of each model (in terms of BIC) for the Swiss value of time and UK value of time datasets. For all models, we estimate an arc elasticity (E) for alternative i with:

$$E_i = \log\left(\frac{\text{Forecasted Trips}_{\text{ i}}}{\text{Base Trips}_{\text{ i}}}\right) / \log(1.1), \tag{40}$$

where 'Base Trips$_i$' is calculated as the sum over the probabilities of choosing alternative i across all choice tasks in the dataset, with 'Forecasted Trips$_i$' calculated equivalently but with adjusted attributes. The corresponding cross elasticities, CE_j, estimate the impact on the probability of choosing alternative j given a change to an attribute of alternative i. We estimate elasticities by using a 10% increase of the controlling factor (see Equation 40), either travel time or travel cost for alternative 1, with the results given in Table 12. We estimate standard errors for the elasticities by taking 30 draws for the parameter values from the corresponding model estimates and robust covariance matrices.

TABLE 12 : Arc elasticities for an increase in travel time or travel cost for the first alternative in the value of time datasets

UK Value of Time dataset						
	arc elasticities for cost (alt1)			arc elasticities for time (alt1)		
	est	s.d.	t-value (vs utility)	est	s.d.	t-value (vs utility)
Utility	-1.5600	0.0617		-2.0476	0.0393	
DFT	-1.7962	0.0488	-3.00	-2.2205	0.0467	-2.83
q-Hamiltonian	-0.9335	0.0564	7.49	-1.6876	0.0766	4.18
q-Amplitude	-1.8411	0.0473	-3.62	-2.4337	0.0647	-5.10
Swiss Value of Time dataset						
	arc elasticities for cost (alt1)			arc elasticities for time (alt1)		
	est	s.d.	t-value (vs utility)	est	s.d.	t-value (vs utility)
Utility	-1.6641	0.2380		-1.4043	0.1241	
DFT	-1.6304	0.1406	0.12	-1.3705	0.0773	0.23
q-Hamiltonian	-1.3034	0.1700	1.23	-1.1648	0.0890	1.57
q-Amplitude	-1.7363	0.1454	-0.26	-1.4199	0.0948	-0.10

For the UK dataset, we observe significantly lower elasticities for the Hamiltonian model, and significantly higher elasticities for the amplitude model (relative to the utility model). There are equivalent results for both cost and time elasticities, with the amplitude and Hamiltonian models predicting greater and smaller shifts, respectively, away from choosing alternative 1 if the cost or time increases, relative to the utility model. For the Swiss dataset, we observe similar patterns to the UK dataset for the quantum models, but these differences are never significant. The elasticities from all models are higher than would be expected, though this is of course typical for SP datasets. Overall, the elasticities from the best performing quantum model (the amplitude model) appear reasonable in comparison to the elasticities given by the utility model.

6. CONCLUSIONS

In this paper, we move away from the tried and tested alternatives to random utility maximisation by considering ideas first developed in quantum physics. With the probability framework developed in quantum physics having made a successful transition to cognitive psychology, we look at whether it can be operationalised into a choice model framework for transportation studies. Under quantum probability theory, a decision-maker has some 'belief state' regarding their preferences over alternatives, from which the probabilities of each alternative can be inferred. Thus a key component of this paper is the development of specifications for the belief state and how these beliefs change through the process of decision making.

We discuss two very different formulations for models generating belief states which incorporate quantum probability theory within a choice model. The first uses a Hamiltonian operator that dynamically evolves the belief state over time. The second is based on directly estimating the probability amplitudes of the belief state for each of the alternatives. We find that our quantum models provide good model fit and outperform standard utility-based models across three route choice datasets as well as providing good out-of-sample fit for the most complex of these. In comparison to Decision Field Theory (DFT), which has also been shown to outperform standard choice models (Hancock, 2019), our quantum models also perform favourably, with the amplitude models recording the best BIC values across all datasets. Additionally, we find good model performance from our quantum Hamiltonian model, although it appears that in the particular choice contexts we test it, the quantum amplitude model performs better. These positive results from our initial tests on quantum choice models suggest that there is ample scope for models with a quantum framework to be used within travel behaviour modelling.

In order to perform fair tests of our quantum choice models against utility-based models and Decision Field Theory, we discuss four different value functions that are used to implement the attribute differences into the choice models. Overall, it often appears that the value functions themselves have a larger impact on model results than the model structure, with the ADLD models for the UK value of time data in particular giving very similar log-likelihoods across models with vastly different paradigms. However, with the exception of linear difference models, it appears that our quantum amplitude model tends to outperform the utility-based models. This exploration of different value functions also leads to the development of a first DFT model which specifically implements non-linear attribute differences. This results in a significant improvement in our DFT model across all datasets, with a substantial improvement recorded for the UK value of time dataset in particular.

A key benefit of the quantum amplitude model over DFT is that it is simple to run and estimate, meaning that it could be applied to a wide range of choice scenarios. However, for these models to make a transition into large-scale modelling, an alternative specification would need to be defined to avoid the same pitfall of random regret minimisation for large numbers of alternatives: using a comparison between every pair of alternatives quickly becomes computationally infeasible and quantum amplitude models with linear attribute differences perform worse than standard multinomial logit models. Another issue with the current specifications of the quantum models is that it could be argued that it is unclear how specifically the use of real and imaginary numbers improves the quantum models, with further work being required to understand the mechanisms at work here. Additionally, by restricting the belief state to be defined by the value functions tested in this paper, we deny the possibility of having a probabilistic belief state, as conceptualised by quantum theory. This limitation can easily be addressed through the incorporation of random parameters for the relative importance of the different attributes, which would naturally allow for a probabilistic belief state.

That being noted, the results from our quantum rotation models in this paper suggest that there is potentially a wide range of benefits of bringing quantum probability theory into choice models. Crucially, our best performing models for the best-worst dataset and the contextual choice dataset, after allowing for model complexity to be taken into account, are the amplitude models including a quantum rotation. This suggests that there is some merit in the concept of quantum rotation, which indicates a different set of basis vectors for choices are required for different choice tasks. For example, the belief state rotation works well for capturing the difference between best and worst choice. Despite the fact that the best-worst choices are related, the quantum rotation suggests that these choices are in fact incompatible: the choices cannot be made at the same time and consequently they may not follow the classical probability law of distributivity. This means that different choices may be observed depending on whether the decision-maker chooses the best or worst alternative first. The use of such information may also provide a number of insights for better specifications of quantum models that incorporate ‘best then worst’ or ‘worst then best’ deliberation processes. An enhanced implementation of our quantum approach demonstrated and improved model performance by integrating both orderings of the choice process, ‘best then worst’ and ‘worst then best’.

While the quantum rotation findings here are just illustrative examples, these results demonstrate that there is major scope for future work within travel behaviour modelling. For example, large-scale models frequently aim to understand a series of related, sequential choices. Given the ability of quantum rotations to capture the translation between best and worst choices, they theoretically should also work for a larger sequence of related choices where continuously adding on separate sets of parameters may not be possible. Ordering effects and state dependence may thus be well captured by models within a quantum framework. Furthermore, it may be possible to mitigate the impacts of contextual effects by applying the appropriate quantum rotation derived from other quantum models that account for the same effect. Future efforts could also compare quantum frameworks against other models that are specifically designed to deal with contextual effects, such as prospect theory or MLBA. Additionally, quantum choice models could be applied to experimental paradigms in which a nudge is involved in some of the choice tasks (for example, a scenario such as the environmentalism example discussed in the introduction of this paper). These future possibilities combined with the positive results in our empirical work mean that this paper serves as a proof-of-concept that quantum ideas can be incorporated into choice models aiming to understand travel behaviour.

ACKNOWLEDGEMENTS

The authors are grateful to four anonymous referees as well as Tony Marley for helpful suggestions on an earlier version of this paper. The authors would also like to acknowledge the financial support by the European Research Council through the consolidator grant 615596-DECISIONS.

REFERENCES

Aerts, D., Broekaert, J., and Smets, S. (1999). A quantum structure description of the liar paradox. International Journal of Theoretical Physics, 38(12):3231–3239.

Aerts, D. and Gabora, L. (2005). A theory of concepts and their combinations ii: A hilbert space representation. Kybernetes.

Asano, M., Khrennikov, A., Ohya, M., Tanaka, Y., and Yamato, I. (2015). Quantum Adaptivity in Biology: From Genetics to Cognition. Springer Netherlands.

Atmanspacher, H. and Filk, T. (2010). A proposed test of temporal nonlocality in bistable perception. Journal of Mathematical Psychology, 54(3):314–321.

Avineri, E. and Bovy, P. H. (2008). Identification of parameters for a prospect theory model for travel choice analysis. Transportation Research Record, 2082(1):141–147.

Axhausen, K. W., Hess, S., König, A., Abay, G., Bates, J. J., and Bierlaire, M. (2008). Income and distance elasticities of values of travel time savings: New swiss results. Transport Policy, 15(3):173–185.

Bagarello, F. (2019). Quantum Concepts in the Social, Ecological and Biological Sciences. Cambridge University Press.

Batley, R., Bates, J., Bliemer, M., Börjesson, M., Bourdon, J., Cabral, M. O., Chintakayala, P. K., Choudhury, C., Daly, A., Dekker, T., et al. (2017). New appraisal values of travel time saving and reliability in great britain. Transportation, pages 1–39.

Batley, R. and Dekker, T. (2019). The intuition behind income effects of price changes in discrete choice models, and a simple method for measuring the compensating variation. Environmental and Resource Economics, 74(1):337–366.

Ben-Akiva, M., McFadden, D., Train, K., et al. (2019). Foundations of stated preference elicitation: Consumer behavior and choice-based conjoint analysis. Foundations and Trends® in Econometrics, 10(1-2):1–144.

Bierlaire, M., Thémans, M., and Zufferey, N. (2010). A heuristic for nonlinear global optimization. INFORMS Journal on Computing, 22(1):59–70.

Birkhoff, G. and Von Neumann, J. (1936). The logic of quantum mechanics. Annals of mathematics, pages 823–843.

Blumenson, L. E. (1960). A derivation of n-dimensional spherical coordinates. The American Mathematical Monthly, 67(1):63–66.

Broekaert, J., Basieva, I., Blasiak, P., and Pothos, E. M. (2017). Quantum-like dynamics applied to cognition: a consideration of available options. Philosophical Transactions of the Royal Society A, 375.

Broekaert, J., Busemeyer, J., and Pothos, E. (2020). The disjunction effect in two-stage simulated gambles. an experimental study and comparison of a heuristic logistic, markov and quantum-like model. Cognitive Psychology, 117:101262.

Brown, S. D. and Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. Cognitive Psychology, 57(3):153–178.

Bruza, P. D., Wang, Z., and Busemeyer, J. R. (2015). Quantum cognition: a new theoretical approach to psychology. Trends in cognitive sciences, 19(7):383–393.

Busemeyer, J. R. and Bruza, P. D. (2012). Quantum Models of Cognition and Decision. Cambridge University Press.

Busemeyer, J. R., Pothos, E. M., Franco, R., and Trueblood, J. S. (2011). A quantum theoretical explanation for probability judgment errors. Psychological review, 118(2):193.

Busemeyer, J. R. and Townsend, J. T. (1992). Fundamental derivations from decision field theory. Mathematical Social Sciences, 23(3):255–282.

Busemeyer, J. R. and Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. Psychological Review, 100(3):432.

Busemeyer, J. R., Wang, Z., and Lambert-Mogiliansky, A. (2009). Empirical comparison of markov and quantum models of decision making. Journal of Mathematical Psychology, 53(5):423–433.

Busemeyer, J. R., Wang, Z., and Townsend, J. T. (2006). Quantum dynamics of human decision-making. Journal of Mathematical Psychology, 50(3):220–241.

Chorus, C. G. (2010). A new model of random regret minimization. EJTIR, 10(2).

Chorus, C. G., Arentze, T. A., and Timmermans, H. J. (2008). A random regret-minimization model of travel choice. Transportation Research Part B: Methodological, 42(1):1–18.

Cohen, A. L., Kang, N., and Leise, T. L. (2017). Multi-attribute, multi-alternative models of choice: Choice, reaction time, and process tracing. Cognitive Psychology, 98:45–72.

Cunha-e Sá, M. A., Madureira, L., Nunes, L. C., and Otrachshenko, V. (2012). Protesting and justifying: a latent class model for contingent valuation with attitudinal data. Environmental and Resource Economics, 52(4):531–548.

Daly, A. and Hess, S. (2010). Simple approaches for random utility modelling with panel data. In European Transport Conference. Citeseer.

Dekker, T. (2014). Indifference based value of time measures for random regret minimisation models. Journal of choice modelling, 12:10–20.

Dugas, C., Bengio, Y., Bélisle, F., Nadeau, C., and Garcia, R. (2001). A universal approximator of convex functions applied to option pricing. Advances in Neural Information Processing Systems, 13:1–8.

Englert, B.-G. (1996). Fringe visibility and which-way information: An inequality. Phys. Rev. Lett., 77(11):2154–2157.

Feynman, R. P., Leighton, R. B., and Sands, M. (1965). The Feynman Lectures on Physics - III. Addison-Wesley.

Fuss, I. and Navarro, D. (2013). Open parallel cooperative and competitive decision processes: A potential provenance for quantum probability decision models. Topics in Cognitive Science, 5:818–843.

Giergiczny, M., Dekker, T., Hess, S., and Chintakayala, P. K. (2017). Testing the stability of utility parameters in repeated best, repeated best-worst and one-off best-worst studies. European Journal of Transport and Infrastructure Research, 17(4).

Greenberger, D. M. and Yasin, A. (1988). Simultaneous wave and particle knowledge in a neutron interferometer. Phys. Lett. A., 128(8):391–394.

Guevara, C. A. and Fukushi, M. (2016). Modeling the decoy effect with context-rum models: Diagrammatic analysis and empirical evidence from route choice sp and mode choice rp case studies. Transportation Research Part B: Methodological, 93:318–337.

Hahnloser, R., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J., and Seung, H. S. (2000). Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. Nature, 405:947–951.

Hancock, T. O. (2019). Travel behaviour modelling at the interface between econometrics and mathematical psychology. PhD thesis, University of Leeds.

Hancock, T. O., Hess, S., and Choudhury, C. F. (2018). Decision field theory: Improvements to current methodology and comparisons with standard choice modelling techniques. Transportation Research Part B: Methodological, 107:18–40.

Hancock, T. O., Hess, S., Marley, A. A. J., and Choudhury, C. F. (2020). An accumulation of preference: two alternative dynamic models for understanding transport choices. Submitted.

Hawkins, G. E., Islam, T., and Marley, A. (2019). Like it or not, you are using one value representation. Decision, 6(3):237–260.

Henningsen, A. and Toomet, O. (2011). maxlik: A package for maximum likelihood estimation in R. Computational Statistics, 26(3):443–458.

Hess, S., Daly, A., and Batley, R. (2018). Revisiting consistency with random utility maximisation: theory and implications for practical work. Theory and Decision, 84(2):181–204.

Hess, S., Daly, A., Dekker, T., Cabral, M. O., and Batley, R. (2017). A framework for capturing heterogeneity, heteroskedasticity, non-linearity, reference dependence and design artefacts in value of time research. Transportation Research Part B: Methodological, 96:126–149.

Hess, S. and Palma, D. (2019). Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. Journal of Choice Modelling, 32:100170.

Hess, S., Rose, J. M., and Hensher, D. A. (2008). Asymmetric preference formation in willingness to pay estimates in discrete choice models. Transportation Research Part E: Logistics and Transportation Review, 44(5):847–863.

Hotaling, J. M., Busemeyer, J. R., and Li, J. (2010). Theoretical developments in decision field theory: comment on Tsetsos, Usher, and Chater (2010). Psychological Review, 117(4):1294–1298.

Hughes, R. I. (1992). The structure and interpretation of quantum mechanics. Harvard university press.

Kempe, J. (2003). Quantum random walks: an introductory overview. Contemporary Physics, 44(4):307–327.

Krajbich, I., Lu, D., Camerer, C., and Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. Frontiers in Psychology, 3:193.

Kvam, P. D., Pleskac, T. J., Yu, S., and Busemeyer, J. R. (2015). Interference effects of choice on confidence: Quantum characteristics of evidence accumulation. Proceedings of the National Academy of Sciences, 112(34):10645–10650.

Leong, W. and Hensher, D. A. (2014). Relative advantage maximisation as a model of context dependence for binary choice data. Journal of choice modelling, 11:30–42.

Lipovetsky, S. (2018). Quantum paradigm of probability amplitude and complex utility in entangled discrete choice modeling. Journal of choice modelling, 27:62–73.

Mahieu, P.-A., Crastes, R., Louviere, J., Zawojska, E., et al. (2016). Rewarding truthful-telling in stated preference studies. Technical report, Faculty of Economic Sciences, University of Warsaw.

Martínez-Martínez, I. (2014). A connection between quantum decision theory and quantum games: The hamiltonian of strategic interaction. Journal of Mathematical Psychology, 58:33–44.

Masiero, L. and Hensher, D. A. (2010). Analyzing loss aversion and diminishing sensitivity in a freight transport stated choice experiment. Transportation Research Part A: Policy and Practice, 44(5):349–358.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behaviour. In Frontiers in Econometrics, ed. P. Zarembka. (New York: Academic Press).

Moreira, C. and Wichert, A. (2016). Quantum probabilistic models revisited: The case of disjunction effects in cognition. Frontiers in Physics, 4:26.

Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th international conference on machine learning (ICML-10), pages 807–814.

Pothos, E. M. and Busemeyer, J. R. (2009). A quantum probability explanation for violations of ‘rational’decision theory. Proceedings of the Royal Society B: Biological Sciences, 276(1665):2171–2178.

Pothos, E. M., Busemeyer, J. R., and Trueblood, J. S. (2013). A quantum geometric model of similarity. Psychological Review, 120(3):679.

Roe, R. M., Busemeyer, J. R., and Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. Psychological Review, 108(2):370.

Seetharaman, P. (2003). Probabilistic versus random-utility models of state dependence: an empirical comparison. International Journal of Research in Marketing, 20(1):87–96.

Stathopoulos, A. and Hess, S. (2012). Revisiting reference point formation, gains–losses asymmetry and non-linear sensitivities with an emphasis on attribute specific treatment. Transportation Research Part A: Policy and Practice, 46(10):1673–1689.

Train, K. (2003). Discrete choice with simulation.

Trueblood, J. S., Brown, S. D., and Heathcote, A. (2014a). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. Psychological review, 121(2):179–205.

Trueblood, J. S. and Busemeyer, J. R. (2011). A quantum probability account of order effects in inference. Cognitive science, 35(8):1518–1552.

Trueblood, J. S. and Busemeyer, J. R. (2012). Quantum information processing theory. In Encyclopedia of the Sciences of Learning, pages 2748–2751. Springer.

Trueblood, J. S., Pothos, E. M., and Busemeyer, J. R. (2014b). Quantum probability theory as a common framework for reasoning and similarity. Frontiers in psychology, 5:322.

Turner, B. M., Schley, D. R., Muller, C., and Tsetsos, K. (2018). Competing theories of multialternative, multiattribute preferential choice. Psychological Review, 125(3):329.

Tversky, A. (1977). Features of similarity. Psychological Review, 84(4):327.

van Cranenburgh, S., Guevara, C. A., and Chorus, C. G. (2015). New insights on random regret minimization models. Transportation Research Part A: Policy and Practice, 74:91–109.

van Rijsbergen, C. (2004). The Geometry of Information Retrieval. Cambridge.

Vitetta, A. (2016). A quantum utility model for route choice in transport systems. Travel Behaviour and Society, 3:29–37.

White, L. C., Pothos, E. M., and Busemeyer, J. R. (2014). Sometimes it does hurt to ask: The constructive role of articulating impressions. Cognition, 133(1):48–64.

Yu, J. G. and Jayakrishnan, R. (2018). A quantum cognition model for bridging stated and revealed preference. Transportation Research Part B: Methodological, 118:263–280.

Zheng, H., Yang, Z., Liu, W., Liang, J., and Li, Y. (2015). Improving deep neural networks using softplus units. In 2015 International Joint Conference on Neural Networks (IJCNN), pages 1–4. IEEE.

APPENDIX: COMPARATIVE MODELS

The logit model. We test our quantum models against standard choice models (McFadden, 1974; Train, 2003; Ben-Akiva et al., 2019) using the same value functions. For these models, we define the utility for an alternative i (dropping the indices for individual (n) and choice task (t)) as:

$$V_i = \varepsilon_i + \sum_{j \neq i} \Delta_{ij}, \tag{A1}$$

where Δ_{ij} is defined using one of the four value functions and ε_i is the unobserved portion of the utility.[27] The assumption of type I extreme value distributions results in typical probabilities:

$$P_i = \frac{e^{V_i}}{\sum_{j=1}^{J} e^{V_j}}, \tag{A2}$$

for each individual (n) and choice task (t). Using this function together with the regret-based value functions would of course result in the wrong signs for the β-coefficients, thus we use $-V_i$ and $-V_j$ instead of V_i and V_j for these models. As we do not have this transformation in the quantum models, they instead use Δ_{ji} in place of Δ_{ij} to ensure the correct sign for the β-coefficients.

Decision Field Theory. We also test DFT, which was originally developed within mathematical psychology (Busemeyer and Townsend, 1992, 1993), thus is very different to models based on econometric theory. The key assumption under a DFT model is that each alternative has a preference value that updates over time within a single choice context. The decision-maker considers the alternatives until they reach some internal threshold (similar to the concept of satisficing, where one of the alternatives is deemed 'good enough') or an external threshold (i.e. some time constraint, where a decision-maker stops deliberating on the alternatives as a result of running out of time to make the decision). An example of a decision process under DFT is given in Figure A1.

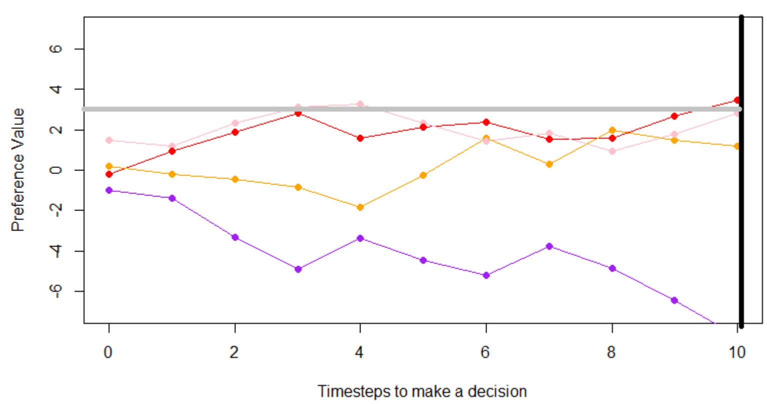

FIGURE A1 : An example of a decision-maker stopping upon reaching either an internal or external threshold

In the example given in this figure, the decision-maker chooses different alternatives if they make their choice after reaching an internal threshold (which is represented by the horizontal line) on the 4th preference updating step or if they conclude after 10 steps upon reaching a time threshold.

[27] Note that this is equivalent to the specification of the probability amplitude, ψ_i in Equation 3.

Mathematically, DFT was originally operationalised for internal thresholds, with a full account of this variation of DFT given by Busemeyer and Townsend (1993). However, the DFT models we use in this paper is based on DFT with external thresholds (c.f. Roe et al. (2001) for the first version of DFT with external thresholds for multiple alternatives). For DFT with an external threshold, the preference values update stochastically as a result of the assumption that a decision-maker compares the alternatives using just a single attribute at each preference updating step. Consequently, the preference values for the alternatives update iteratively:

$$P_t = S \cdot P_{t-1} + V_t, \tag{A3}$$

where P_t is a column vector containing the preference values of each alternative i at time t. S is a feedback matrix with memory and sensitivity parameters (detailed in Equation A4) and V_t is a valence vector (Equation A5), which varies depending on which attribute is attended to at time t, and is equivalent to a 'momentary utility'. The feedback matrix we use is based on the definition by Hotaling et al. (2010):

$$S = I - \phi_2 \times \exp(-\phi_1 \times D^2), \tag{A4}$$

where I is an identity matrix of size n, where n is the number of alternatives. The feedback parameter has two free parameters. The first, ϕ_1, is a 'sensitivity' parameter, which allows for competition between alternatives that are more similar (in terms of attribute values). This is the driving force that results in DFT being able to account for contextual effects (Roe et al., 2001). The second parameter, ϕ_2 is a 'memory' parameter, which captures whether attributes considered at the start of the deliberation process or attributes considered at the end are more important. Crucially, a value of $\phi_2 = 0$ results in the feedback matrix collapsing to an identity matrix, meaning that 'no memory loss' results in it not being possible for ϕ_1 to have an impact. This means that ϕ_2 has an important mathematical role in the model and thus cannot be purely treated as a psychological parameter, which is especially the case when DFT is applied to choice-only data. Finally, D is some measure of distance between the alternatives. In this paper, we use the Euclidean distance for simplicity. Next, the valence vector can be described:

$$V_t = C \cdot M \cdot W_t + \varepsilon_t, \tag{A5}$$

where C is a contrast matrix used to rescale the attribute values such that they total zero, M is a matrix containing the attribute values for all of the alternatives, $W_t = [0..1..0]'$ is a column vector and ε_t is an error term. W_t defines which attribute is being attended to by the decision-maker at preference updating step t, with entry $k = 1$ if and only if attribute k is the attended attribute. Note that the DFT models in this paper follow the new attribute scaling method developed by Hancock et al. (2020). Instead of estimating attribute importance weights, w_k, that corresponds to the likelihood of a decision-maker attending to that attribute k, we estimate 'attribute scaling coefficients'. These have many benefits (see Hancock et al. (2020) for a detailed explanation of these), including, most importantly, avoiding the limitation of having to sum to one. By instead assuming that each attribute is attended to with the same likelihood (all weights, $w_k = 1/n$), the relative importance can instead enter as a set of scaling coefficients, β_k, which are applied to the attributes before they are entered (through M in Equation A5) into the calculation of the valence vector at each preference updating step.

Finally, the error term is drawn from independent and identically distributed normal draws with mean 0 and a standard deviation, σ_ε, which is an estimated parameter. Consequently, the preference values P_t converge to a multivariate normal distribution (Roe et al., 2001). To calculate the probability with which each alternative is chosen under decision field theory, we simply require the expectation and covariance of P_t (ξ_t and Ω_t, respectively, detailed in Hancock et al. 2018). Hence the probability of choosing alternative j from a set of J alternatives at time t is:

$$Pr_j\left[\max_{i\in J} P_t[i] = P_t[j]\right] = \int_{X>0} \exp\left[-(X-\Gamma)'\Lambda^{-1}(X-\Gamma)/2\right]/(2\pi|\Lambda|^{0.5})dX, \tag{A6}$$

with X the set of differences between the preference value for the chosen alternative and each other alternative, $X = [P_t[j] - P_t[1], \ldots, P_t[j] - P_t[J]]'$. Additionally, we require transformations of the expectation and covariance, $\Gamma = L\xi_t$, $\Lambda = L\Omega_t L'$, with L a matrix comprised of a column vector of 1s and a negative identity matrix of size $J-1$ where J is the number of alternatives. The column vector of 1s is placed in the i^{th} column where i is the chosen alternative.

Prior to this paper (as far as the authors are aware), DFT has always been implemented using linear attribute differences, which are enforced by the contrast matrix, C. This results in element j of the matrix $C \cdot M \cdot W_t$ taking the form:

$$CMW_t[j] = \sum_1^n \frac{\beta_k(x_{jk} - x_{ik})}{n}, \tag{A7}$$

where n is the number of alternatives as before and k is the attribute being attended to at preference updating step t. Given that the value functions incorporating a softplus function do not result in $\Delta_{ij} = \Delta_{ji}$, they are not appropriate functions to be used within a DFT model. This is because DFT models use just a single difference Δ_{ij}, thus it is unclear whether Δ_{ij} or Δ_{ji} should be used. However, our ADLD value function can be configured such that $\Delta_{ij} = \Delta_{ji}$, if $\lambda_1 = \lambda_2$. Thus our ADLD DFT models require just a single λ parameter. The element $CMW_t[j]$ can thus have an updated numerator based on the ADLD value function:

$$CMW_t[j] = \sum_1^n \frac{\exp\left(-\lambda \cdot \beta_k \cdot |x_{jk} - x_{ik}|\right) \cdot \beta_k \cdot (x_{jk} - x_{ik})}{n}. \tag{A8}$$

It is worth noting here that we do not sum across attributes for each preference updating step, though to calculate the expectation of the preference values after t steps, a summation is required.