From Cognitive to Computational Modeling: Text-based Risky Decision-Making Guided by Fuzzy Trace Theory

Jaron Mar
The University of Auckland
`jaron.mar@auckland.ac.nz`

Jiamou Liu
The University of Auckland
`jiamou.liu@auckland.ac.nz`

Abstract

Understanding, modelling and predicting human risky decision-making is challenging due to intrinsic individual differences and irrationality. Fuzzy trace theory (FTT) is a powerful paradigm that explains human decision-making by incorporating gists, i.e., fuzzy representations of information which capture only its quintessential meaning. Inspired by Broniatowski and Reyna's FTT cognitive model, we propose a computational framework which combines the effects of the underlying semantics and sentiments on text-based decision-making. In particular, we introduce Category-2-Vector to learn categorical gists and categorical sentiments, and demonstrate how our computational model can be optimised to predict risky decision-making in groups and individuals.

1 Introduction

Imagine that your town is preparing for a viral outbreak which is projected to kill 600 people. Two alternative programs to combat the virus have been proposed. Assume that the exact scientific estimates of the consequences of the programs are **Program A:** "*200 people will be saved*"; and **Program B:** "*1/3 probability that all 600 lives will be saved; 2/3 probability that no lives will be saved*". Given these choices, which program would you choose? Alternatively, if choices were presented as follows, which program would you choose? **Program C:** "*400 people will die*"; and **Program D:** "*1/3 probability that no one will die and a 2/3 probability that all 600 will die*". This problem is a modified version of the *Asian disease problem* (ADP) (Tversky and Kahneman, 1981), a well-studied *risky decision-making problem* (RDMP) in psychology where decisions are made under risk or include probabilistic outcomes (Edwards, 1954). In this RDMP, programs A and B form the *gain frame* where choices are worded in a positive and optimistic manner, whereas programs C and D form

the *loss frame* where choices are written in a negative and pessimistic manner. Studies have validated that in the gain frame, humans overwhelmingly prefer the safe choice A (72%), whereas in the loss frame, they overwhelmingly prefer the risky choice D (78%) even though the choices and outcomes in both frames are equivalent (Tversky and Kahneman, 1981). This phenomenon, known as the *Allais paradox*, implies that observed human choices are inconsistent with predictions based on expected utility alone, thereby confirming the influence of language, wording of choices, and sentiments on human decision-making.

Being able to understand, model and predict human decision-making leads to many real-world applications, from predicting election results (Hillygus and Shields, 2005), to improving user experience in recommender systems (Chen et al., 2013). However, Allais paradox means that understanding the integral but complex cognitive process of decision-making, particularly in humans, is extremely challenging due to our diverse characteristics, beliefs and experiences. Furthermore, human decision-making is often fraught with irrationality even in the presence of overwhelming evidence against some choice or beliefs (Simon, 1993). This brings into question how human decision-making can be modelled with these complexities and nuances involved. This is an especially important task when considering current approaches to decision-making, such as utility theory which typically lacks any behavioural basis and ignore human sentiments during human decision-making (Lerner et al., 2015).

Our goal is to develop a model of automated human decision-making that bridges current decision-making techniques with *fuzzy trace theory* (FTT), an established cognitive theory to predict group and individual decision making outlined in sections 2.2. Originally proposed by Brainerd and Reyna in the 1990s, FTT aims to explain cognitive phenomena

in memory and reasoning ([Brainerd and Reyna](), [1990]()). In a nutshell, FTT posits that humans form two types of mental representations, known as *verbatim* which are detailed representations and *gist* which are fuzzy representations that only capture the most quintessential meanings, and people prefer to make decisions based on gist rather than verbatim representations.

In contrast with alternative cognitive and decision-making theories such as expected utility theory ([Friedman and Savage](), [1952]()) and prospect theory ([Kahneman and Tversky](), [1980]()), we adopt FTT for two reasons. Firstly, FTT is the most holistic cognitive model which encompasses theories of how information is stored in memory and how memory plays an important role in our decision-making rather than treating decision-making as an isolated process. Because of this, FTT provides us with an extensive set of tools to explain and evaluate decision-making. Secondly, is FTTs suitability for computational modelling as conceptual parallels can be drawn between representation learning, particularly in neural-based language modelling, and the process of creating gist representations by distilling the quintessential information. For example, popular embedding methods for words, sentences and documents in NLP aim to create fuzzy semantic representations through dimensionality reduction of language to semantic vectors which can be viewed as gist representations of the original language ([Liu et al.](), [2020]()).

Contributions: We investigate two levels of text-based risky decision prediction tasks, group and individual-level prediction from a computational standpoint and incorporating state-of-the-art methods in NLP, we further investigate:

- *How do gist representations of choices give rise to decisions?* We present a framework of decision-making based on gist representation learning.

- *How can we computationally encode gist representations based on the language of choices?* We outline how gist representations can be computationally encoded using techniques in NLP and propose Category-to-Vector (Cat2Vec), to learn and predict categorical embeddings of choices.

- *How can we extract the underlying sentiments of gist representations?* By extending Cat2Vec, we show how sentiments can be learnt at a categorical level; this differs from traditional approaches

of sentiment analysis in NLP that examine sentiments at a text level.

- *How can individual differences of individuals and groups be modelled, what impact do these differences have on decision-making?* We propose that individual differences are mechanisms that can encode errors at various points in the decision-making process and propose an optimisation method to infer these individual differences.

- Finally, we demonstrate in experiments that our proposed model achieves state-of-the-art performance in predicting group and individual-based risky decision-making compared to baselines.

2 Task Formulation and Related Work

Risky decision-making has been studied in many different contexts. Here we formulate *n-choice decision-making problem* (nDMP): Taking as input natural language descriptions of n possible choices/outcomes $\mathcal{O}$, choose the most preferred outcome from the set of choices $\mathcal{O}$. We focus on a sub-problem known as a *n-choice risky decision-making problem* (nRDMP) which is an nDMP where there is some risk or probabilistic outcomes associated with choices in $\mathcal{O}$, e.g., programs B and D in the ADP. Specifically, we investigate the *gain-loss framing problem* which is comprised of two nRDMPs, nRDMP$_{gain}$ where choices are written as *gain frames* which accentuate the positive features of the text, e.g., programs A and B form a 2RDMP$_{gain}$ where 'saving people's lives' is the accentuated feature. Conversely, 2RDMP$_{loss}$ where choices are written as *loss frames* which accentuate the negative features of the text, e.g., programs C and D form a 2RDMP$_{loss}$ where 'people dying' is the accentuated feature. Additionally, choices have equivalent outcomes across both 2RDMPs.

2.1 Classical decision theory

Classical decision theory abstracts the outcomes using utilities, which are numerical values that reflect desirability. For example, *expected utility theory* (EUT) identifies the choice that maximises the expected utility assuming the axioms of rationality ([Von Neumann and Morgenstern](), [2007]()). However, in human decision-making these axioms are often violated, giving rise to, e.g., Allais ([Allais](), [1953]()) and Ellsberg ([Segal](), [1987]()) paradoxes. Generalised EUT such as *uncertain utility theory* ([Gul et al.](),

2008), *cumulative prospect theory* (CPT) (Tversky and Kahneman, 1992), and *multiple-criteria decision-making* (MCDM) (Zeleny, 2012) were proposed to resolve these discrepancies. However, these classical approaches not only fail to take into account semantic information given by the working of choices which is important contextually for decision making, but they also ignore cognitive processes such as sentiments of decision-makers.

Recent breakthroughs in NLP have led to a revolution in the breadth and robustness of problems that can be solved involving natural language by successfully capturing the underlying semantics and relationships of language. For example, neural language models have found resounding success in representation learning (Mikolov et al., 2013; Devlin et al., 2018), the task of uncovering feature representations of language which are useful for downstream NLP tasks. One such downstream NLP task, sentiment analysis, has benefited largely from the application of language models such as XLNet (Yang et al., 2019) and ULMFiT (Howard and Ruder, 2018). Rapid advancements thus give hope for the development of sophisticated computational decision-making models.

2.2 Group/Individual-level Tasks

In this paper, we consider two specific 2RDMP, *group level risky decision making* (GL-RDM) which the majority of psychological studies focus on and the novel task of *individual level risky decision making* (IL-RDM), defined as follows.

GL-RDM: Given a set of observed outcomes from human RDM experiments, each of which is described by a 5-tuple (2RDMP$_{gain}$, 2RDMP$_{loss}$, P_{gain}, P_{loss}, category), where 2RDMP$_{gain}$ is the gain frame of a 2RDMP, P_{gain} is the proportion of individuals in the gain frame who chose the *risky* choice, and $category$ is a grouping of similar experiments based on design and participants described in Section 6.1. 2RDMP$_{loss}$ and P_{loss} can be defined similarly by replacing $gain$ with $loss$. GL-RDM's objective is to predict the distribution of choice between P_{gain} and P_{loss} and for unseen human experiments within the same category.

IL-RDM: Given a set of nRDMPs, RDP = $\{rdp_1, rdp_2, \ldots, rdp_n\}$ where gain/loss frames of the same problem can appear as separate RDPs rdp_i, a set of individuals $Ivd = ivd_1, ivd_2, \ldots, ivd_m$ and a function which maps individuals and RDPs to their preferred choice

$PC(id_i, rdp_j) = pc_{i,j}$ where $pc_{i,j}$ is individual id_i's preferred choice for rdp_j. The objective for IL-RDM is to learn a model/mapping function for each individual which can predict an individual's preferred choice for unseen nRDMPs.

3 FTT-guided Risky Decision-making

The BR model. Broniatowski and Reyna laid out four main FTT principles in developing a cognitive model, i.e., the *BR model*, for the GL-RDM task (Broniatowski and Reyna, 2018, 2014). These principles are: (**C1**) Decision choices are encoded in different levels of gist representations, e.g., *categorical-* and *interval-*levels based on the psychological notion of levels of measurement (Stevens et al., 1946). (**C2**) Categorical gist representations of choices are distinguished based on binary (positive/negative) sentiments and decision-makers will prefer options with positive associations. In the BR model, sentiments of categories are drawn upon social and moral principles which are stored in long-term memory, e.g., saving lives is fundamentally good. (**C3**) When comparisons of categorical gist representations do not arrive at a conclusive result, the decision-maker will revert to more precise gist representations. In the BR model, gist representations compete and combine such that the simplest gist representation is chosen. (**C4**) Categorical gist is encoded based on the decision-maker's prior experiences and individual differences, i.e., *need for cognition* (NFC), *numeracy* (NUM), and *risk sensitivity* (RS).

Human experiments have provided evidence that the BR model is capable of explaining GL-RDM. However, being a box-arrow model, the BR model is comprised of hypothesized concepts or processes that lack precise definitions. Hence applying the model requires human interpretations and judgements on, e.g., notions such as gist lattices of each RDMP, the acquisition of sentiments, and individual differences. This informal nature, along with the inflexibility of the model being unable to be easily adapted to IL-RDM prevents the model from being used as an automated predictive tool.

Our model. Towards a fully automated tool for the RDMP tasks, we propose a computational model of risky decision-making that takes the input text descriptions of an RDMP and solves the GL-RDM and IL-RDM tasks automatically. The model is depicted in Figure 1. The key features of our model include: (1) All model components are automated,

i.e., gist representations are extracted through NLP as categorical embeddings. (2) Categorical and interval representations are formally defined as hierarchical with the interval level encapsulating the properties and information of the categorical-level (Stevens et al., 1946). (3) Individualistic differences, NFC, NUM and RS (see below) directly affect decision-making at a representational level and errors in judgement can propagate through the model adding more expressive.

4 Computing Gist Representations

The first challenge we tackle is the computational encoding of gist representations and how individual differences encode *error*, a mechanism to model variations in human decision-making at a representation level to perform GL-RDM and IL-RDM.

4.1 Categorical Representations

Categorisation is the act of grouping documents into categories based on semantic or sentiment similarity. For example, the `entertainment` category in the news dataset (see Section 6.1) comprises of various articles spanning multiple topics such as games, movie reviews, and celebrity gossip. The ability to categorise and recall the underlying sentiments of categories is an important prerequisite for FTT decision making asserted by principle C2 where humans prefer choices associated with categories with positive connotations over negative connotations. For example, the sentiments of the `travel` category in the news dataset has a strong negative sentiment due to the news articles being collected during the outbreak of COVID-19. Given this negative sentiment, people would be dissuaded from travelling. Current sentiment analysis methods focus on granular extraction of sentiments from text rather than categories where words that are highly indicative of a category do not necessarily reveal any insights into their sentiment.

Vector representations of words (Mikolov et al., 2013), sentences (Devlin et al., 2018), and documents (Le and Mikolov, 2014) capture the semantic relationships between entities. At a higher level, a *categorical embedding* should capture semantic relations between categories of documents. To our knowledge, no such representation has been proposed. To fill this gap, we propose *Category-2-Vector* (Cat2Vec) and a sentiment based extension, *sentiment-Cat2Vec*. Cat2Vec aims to find model-agnostic categorical representations that facilitates

the prediction of categories from text and sentiments from categories. More formally, given a set of M categories $C = \{k_1, k_2, \ldots, k_M\}$, a training set contains a number N of document-category pairs $\{(d_1, c_1), \ldots, (d_N, c_N)\}$, where each d_i is a document and $c_i \in C$ is the (ground truth) category of d_i. The objective is to maximise the average log probability $\frac{1}{N} \sum_{i=1}^{N} \log P(c_i|d_i)$ where $P(c_i|d_i)$ is the probability that document d_i belongs to category c_i. In sentiment-cat2vec, $P(c_i|d_i)$ is replaced by $P(s_i|c_i)$, the probability that category c_i belongs to a certain sentiment class. Here, we consider only binary positive and negative sentiments.

Cat2Vec extends a contrastive learning via negative sampling approach by simultaneously maximising the similarity between document encodings, v_{d_i}, with true category embeddings, v_{c_i}, by minimising the similarity between v_{d_i} and K negative category embeddings defined by the objective:

$$
\log \sigma \left((v_{c_i} \odot v_{i_i}) \cdot v_{d_i} \right)
$$
$$
+ \sum_{j=1}^{K} \mathbb{E}_{k_j \sim P_{\text{noise}}(C)} [\log \sigma((-v_{k_j} \odot v_{i_j}) \cdot v_{d_i})]
$$
$$
(1)
$$

where $\odot$ represents element-wise multiplication, $P_{\text{noise}}(C)$ is a noise distribution that dictates how categories are sampled, we select a uniform distribution, $\sigma(x) = 1/(1 + \exp(-x))$ and v_{i_i} is a category importance vector for category i which is learned simultaneously with the category embedding which provides an attention-like effect over category by accentuating or diminishing certain features in the category embedding when multiplied together. Furthermore, $v_{d_i} = Enc(d_i)$ where Enc is a document vector encoding function. In this paper we adopt a bi-directional LSTM with self-attention such that $v_{d_i} = (\overrightarrow{\alpha} \odot \overrightarrow{v}_{d_i} \| \overleftarrow{\alpha} \odot \overleftarrow{v}_{d_i})$ where $\overrightarrow{\alpha}, \overleftarrow{\alpha}$ are self-attention weights of the forwards and backwards LSTMs, resp., and $\overrightarrow{v}_{d_i}, \overleftarrow{v}_{d_i}$ are the hidden state output vectors for document d_i of the forward and backwards LSTMs, respectively. However, the encoder is interchangeable in Cat2Vec e.g. pretrained transformers like BERT (Devlin et al., 2018) can be used.

The novelty of our model lies in two main aspects, the introduction of a category importance vector to improve the ability of the model to learn relations between categories and the ability of our model to estimate $P(s_i|c_i)$ given labelled text doc-
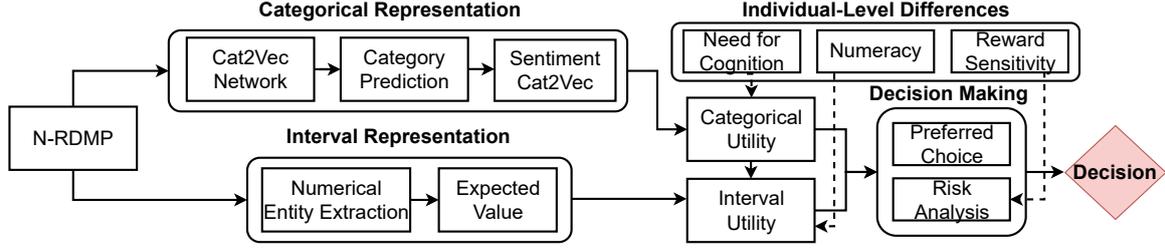
Figure 1: The main architecture of our computational model for decision making based on FTT

uments. To estimate $P(s_i|c_i)$ we introduce an extra dense output layer (D) in figure 2 which predicts $P(s|d_i)$, the probability that document i belongs to a certain sentiment class. In the case of binary sentiments, the joint loss becomes the binary cross-entropy loss of predicting the correct sentiment of a document plus the negative sampling loss in equation 1. After training the model we can estimate $P(s_i|c_i)$ by feeding the learned category embeddings, $(v_{c_i} \odot v_{i_i})$ into the output layers (D). Although these output layers are trained to learn $P(s|d_i)$, since the learned category embeddings are based on the document embeddings and are learned in the same semantic space, this approach gives us good estimates of $P(s_i|c_i)$.
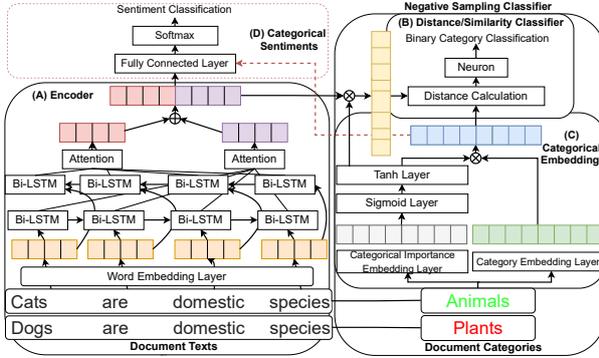

Figure 2: Cat2Vec Model

Equation 2 shows how error encoded categorical utility (CU) is calculated from categorical representations where Category(choice) is a function that takes RDM choices as inputs and outputs the underlying category related to the choice, Sentiment(category) is a function which takes a category as input and outputs the underlying sentiments related to that category as $\text{pos}_{category} - \text{neg}_{category}$. Categorical error is encoded based on NFC, an individual's tendency to engage in and enjoy cognitive activities (Cacioppo et al., 1996), can introduce error at a categorical level to account for individualism. To calculate an error encoded CU, we

sample from a logistic distribution which is consistent with existing literature in qualitative discrete choice models (McFadden, 2001). Formally, $\text{NFC} \in (0,1)$ and $CU \sim \text{Logistic}(\mu, s)$ where $\mu = \mathbb{E}[X]$ is the expected or true utility value and $s(\text{NFC}) = |\text{NFC} - 1| \times \mathbb{E}[X]$.

$$CU = \text{Logistic}(\text{Sentiment}(\text{Category}(\text{choice})), \text{NFC}) \quad (2)$$

4.2 Interval Representations

Interval representations are a more precise representation than categorical representation. It encodes the calculation of the expected value (EV) and utility of choices. Numerical information from text can be extracted using simple text extraction or *named entity recognition* (NER) (Nadeau and Sekine, 2007) where probabilities and their associated quantities can be extracted as arrays, e.g., in program B of the RDMP in the introduction the probabilities would be $[1/3, 2/3]$ and their corresponding quantities would be $[600, 0]$.

Equation 3 outlines the process to generate error-encoded interval utilities where CU is the categorical utility defined in equation 2 and EV is an expected value function which takes an input RDM choices and outputs the corresponding expected value associated with probabilities and quantities in choices which can extracted using standard text identification techniques such named entity recognition. Error is encoded based on NUM (Kahneman, 2003), which measures a person's ability to interpret and work with numbers to account for individualism is calculated as $IU \sim \text{Logistic}(\mu, s)$ where $\text{NUM} \in (0,1)$, $\mu = \mathbb{E}[X]$ is the true expected value and $s(\text{NUM}) = Q|\text{NUM} - 1| \times \mathbb{E}[X]$ where Q is number of quantities in the choice to account for error involving multiple calculations.

$$IU = \text{Logisitic}(\text{EV}(\text{choice}), \text{NUM}) \cdot CU \quad (3)$$

4.3 Representations for Decision Making

Finally, combining these representations allows us to derive the most beneficial choice in an nRDMP. The *preferred categorical*, Pref_{Cat} and *preferred interval*, Pref_{Int} choices are calculated based on which choice maximises categorical and interval utilities, respectively. If $\text{Pref}_{\text{Cat}} = \text{Pref}_{\text{Int}}$, there is a consensus on the best choice. If $\text{Pref}_{\text{Cat}} \neq \text{Pref}_{\text{Int}}$, there is no clear best choice. In this case, RS, a person's preference towards pursuing riskier but more rewarding decisions (Kacelnik and Bateson, 1997), is adopted as in the BR model. Risk sensitivity influences the probability of choosing the *safest* or *riskiest* choice in an nRDMP as $P(\text{risky}) = 1/(1 + e^{-RS})$ where $RS \in (-3, 3)$. The safest choice is one that involves the least probabilistic outcomes, whereas conversely, the riskiest choice involves the most probabilistic outcomes e.g., in the ADP in the introduction, program A is the safest as it involves one certain outcome while program B is the riskiest with two probabilistic outcomes.

4.4 Decision Making: A Worked Example

To demonstrate the fluidity of our model we apply our model to the ADP from the introduction. In the gain frame, the predicted category of programs A and B using the pretrained Cat2Vec from the experiments predicts the `life` category for both programs. The sentiments of the `life` category predicted by Sentiment-Cat2Vec is 0.9999 positive and 0.0001 negative giving categorical utility defined as $\text{pos}_{\text{category}} - \text{neg}_{\text{category}}$ for both programs equal to 0.9998. Taking into account numerical information, the expected value of programs A and B is 200 people being saved, the interval utility is thus the expected value times the categorical utility which is 199.6 for both programs. No consensus between categorical or interval choices can be made due to unclear preferred choices for both. Thus, the final choice is decided by risk sensitivity.

Individual differences encode error and preferences into choices allowing for consensus to arise, e.g., a person with low numeracy will sample interval utilities from a logistic distribution with a larger spread than someone with high numeracy who samples utility close to the true utility. Because the error encoded utilities are sampled, the preferred choice can change on different runs of the problem; however, individual differences influence the average choice. Figure 3 shows a snapshot when one parameter was altered while the others

were fixed and how these parameters can alter utilities to prefer certain choices across frames in the ADP.
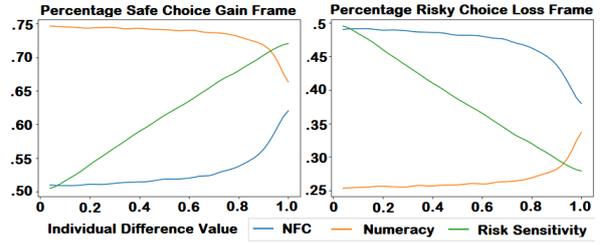

Figure 3: Effects of individual differences on the ADP.

5 Learning Individual Differences

The last challenge we explore is how optimal individual-level parameters, NFC, NUM and RS can be inferred in GL-RDM and IL-RDM by optimising the following objective functions.

$$\arg \min_{ivd} \sum_{P,\hat{P}}^{\text{Exp}} |P_{\text{gain}} - \hat{P}_{\text{gain}}(ivd)| + |P_{\text{loss}} - \hat{P}_{\text{loss}}(ivd)| \quad (4)$$

$$\hat{P}_{\text{gain}}(ivd) = \sum_{exp \in \text{Exp}} \frac{\mathbb{E}[\text{CDM}(\text{RDMP}_{gain}, ivd) = \text{risky}]}{|\text{Exp}|}$$
$$(5)$$

GL-RDM Objective. Is given by equation 4 where $ivd = (\text{NUM}, \text{NFC}, \text{RS})$ are individual-level parameters which represent the characteristics of the entire group and $\text{Exp} = \{e_1, e_2, \ldots, e_i\}$ is the set of results based on human psychological experiments where each e is a 5-tuple described in the task formulation in section 2.2.

IL-RDM Objective. Is given by equation 6 where $ivd_i = (\text{NUM}, \text{NFC}, \text{RS})$ is the parameters which characterises individual i, RDP is a set of nRDMPs and PC is the mapping function of individuals to their preferred choices defined in the task formulation in section 2.2. Thus, the goal is to learn optimal individual parameters for each individual which maximises the expectation that CDM chooses their true preferred choice over all RDPs.

$$\arg \max_{ivd} \sum_{j}^{|\text{RDP}|} \mathbb{E}[\text{CDM}(\text{rdp}_j, ivd_i) = PC(ivd_i, \text{rdp}_j)]$$
$$(6)$$

6 Experiments

6.1 Datasets

Categorical News. A dataset used for training/fine-tuning Cat2Vec and benchmark algorithms. The

dataset contains 22601 news articles with binary sentiments labelled from various news outlets dating from February to April 2020 using the Google News API spanning 46 news categories, e.g. `travel`, `entertainment` and `death`.

Group Risky Decision Making. A dataset of 88 psychological human experiments results grouped into categories used in the evaluation of the BR model. The categories represent differences in experiment controls and participants that undertook each experiment, e.g., 'ADP; within-subjects, low PISA'. The category outlines the risky decision-making problem; experimental design which can be grouped into *within*, where each participant is given both frames of a decision or *between* subject designs, where two independent groups answer each frame; and numeracy of participants, based on the performance of the country in which the experiment took place in the Program for International Student Assessment (PISA) (Stacey, 2015).

Individual 2-RDMP Prediction. A curated dataset of 38 unique 2-RDMPs selected from various psychological experiments regarding risky decision-making answered by 121 university students using a within-subject experimental design. Of the 38 2-RDMPs, most problems contain a corresponding gain and loss frame, e.g., the ADP in the introduction, each frame is considered a separate problem. Participants selected their preferred choice from the same pre-shuffled RDMP set[1] and no pre/post-processing of data was performed.

6.2 Evaluation Metrics

We apply different evaluation metrics suitable for each RDM task. For GL-RDM, we compare the true log-odds ratio (LOR) given by equation 7, between experimental results predicted by our model and the BR baseline model. Intuitively, the LOR measures the consistency of choices across the gain and loss frames.

$$
\text{LOR}(P_{\text{gain}}, P_{\text{loss}}) = \ln\left(\frac{P_{\text{gain}}(1 - P_{\text{loss}})}{P_{\text{loss}}(1 - P_{\text{gain}})}\right) \quad (7)
$$

To determine the goodness-of-fit between the predicted LOR, we apply a hypothesis test, the Wald statistic (χ^2) given by equation 9. The standard error (SE) is given by equation (8) where $n_{\text{safe,gain}}$ represents the number of individuals choosing the safe choice in the gain frame,

$n_{\text{safe,loss}}$, $n_{\text{risky,gain}}$, $n_{\text{risky,loss}}$ can be derived similarly. The standard error asymptotically approaches a normal distribution when n is sufficiently large; thus, the associated Wald statistic, equation (9), follows a chi-square distribution with one degree of freedom.

$$
\text{SE} = \sqrt{\frac{1}{n_{\text{safe,gain}}} + \frac{1}{n_{\text{safe,loss}}} + \frac{1}{n_{\text{risky,gain}}} + \frac{1}{n_{\text{risky,loss}}}}
$$
$$(8)$$

$$
\chi^2 = \left(\frac{\text{LOR}(P_{\text{gain}}, P_{\text{loss}}) - \text{LOR}(\hat{P}_{\text{gain}}, \hat{P}_{\text{loss}})}{\text{SE}}\right)^2 \quad (9)
$$

To compare the parsimony and implicitly the error between our the BR and null (Busemeyer et al., 2015) models, we use the Akaike information criterion (AIC) and Bayesian information criterion (BIC). For IL-RDM, we evaluate the accuracy of each model correctly predicting the true choices for each individual.

6.3 Benchmark Algorithms

In the paper, we use two different sets of baselines. For GL-RDM, we directly compare our model against the BR model. Due to the small number of experiments per grouping in the GL-RDM dataset, to maintain parity with the BR baseline, we apply the same jackknife-leave-one-out (JLOO) method used for parameter estimation in the BR baseline model to avoid post-hoc parameter estimation (Busemeyer and Wang, 2000). Formally, given m observed human risky decision making experiment results within a category of comparable RDPs, $\text{Exp} = \{e_1, e_2, \ldots, e_m\}$ as described in the task formulation in section 2.2. We wish to estimate m values of $G_i = (\text{NUM}_i, \text{NFC}_i, \text{RS}_i)$ where G_i is group-level differences relating to observed experiments i where $i = 1 \ldots m$. To achieve this, we apply the G_i can be estimated by equation (10) where ER_{-i} is the set of experimental results excluding er_i as not to use the result in the estimation.

$$
\arg\min_{G_i} \sum_{P_{\text{gain}}, P_{\text{loss}}}^{\text{ER}_{-i}} |P_{\text{gain}} - \hat{P}_{\text{gain}}(G_i)| + |P_{\text{loss}} - \hat{P}_{\text{loss}}(G_i)|
$$
$$(10)$$

For IL-RDM, due to a lack of existing benchmark algorithms we compare our model against two baselines (1) Naive binary model using pretrained transformer language models where all

---

[1]See appendix, section A.3 for questionnaire

RDM-choices are combined as a single input and outputs 0 or 1 corresponding to the safe or risky choice. (2) Sentiment analysis models as claim C2 asserts sentiments are highly influential in decision-making where decisions are based on choices with the highest positive sentiment. **Random:** Uniformly samples one of the available choices. **Vader:** A rule-based sentiment analysis for social media(Gilbert and Hutto, 2014). **XLNet:** SOTA pretrained autoregressive language model fine-tuned on the news dataset sentiments (Yang et al., 2019). **ULMFiT:** Pretrained language model fine-tuned on the news dataset sentiments using inductive transfer learning (Howard and Ruder, 2018).

6.4 Experiment Results

GL-RDM results. Table 1 (full table A.4 in Appendix) shows key discrepancies between our computational model, CDM, compared to both the actual LOR based on all 88 human experiments and those predicted by the BR model. Within each category, we find optimal group-level parameters which minimise (4) to calculate the predicted LOR of our model for each experiment using the jackknife-leave-one-out (JLOO) method to maintain comparability between the BR model.

Reference	Actual	BR	CDM	SE	χ^2
Standard ADP; one presentation, between-subjects, low PISA					
(1) Tversky and Kahneman (1981)	**2.20**	**1.65***	**1.86**	**.26**	**1.83**
(2) Mayhorn et al. (2002), Young	2.98	1.68	1.50	.58	5.89*
			TOTAL of 14 predicted: 13 (93%)		
Standard ADP; within-subjects, low PISA					
(3) LeBoeuf and Shafir (2003) Exp #2	**.57**	**1.05***	**.81**	**.17**	**.52**
			TOTAL of 3 predicted: 3 (100%)		
Standard ADP; multiple presentations, between-subjects, low PISA,					
(4) Jou et al. (1996)	**2.01**	**.87***	**1.13**	**.33**	**6.99***
			TOTAL of 6 predicted: 5 (83%)		
Other problems; multiple presentations, between-subjects, high PISA					
(5) Kühberger (1995) Plant #2	2.34	.7*	.34	.73	7.50*
			TOTAL of 4 predicted: 3 (75%)		
Zero-complement problems; multiple presentations, between-subjects					
(6) Kühberger and Tanner (2010) Crops	-.43	0	.34	.30	6.83*
(7) Kühberger and Tanner (2010) Fish disease	**.83**	**0***	**.37**	**.30**	**15.74***
			TOTAL of 7 predicted: 5 (71.4%)		
"400 not saved" certain-option problems; multiple presentations, between-subjects, high PISA					
(8) Kühberger (1995) Plant #1	**.49**	**-.88***	**.11**	**.58**	**.50**
(9) Kühberger (1995) Cancer #1	-1.36	-.74	.11	.60	6.21*
			TOTAL of 5 predicted: 6 (83.3%)		
		OVERALL TOTAL of 88 predicted: 82 (93.2%)			

* Indicates results with Wald statistics over 1 degree of freedom
Note: Bolded rows indicate results where CDM outperforms or is equal to the BR model

Table 1: Group Level Experiment Results.

Critically, our results show that our computational model is capable of automating the prediction of human risky decision making on a wide range of RDMPs by predicting 82 of 88 (93.2%) experiments based on the Wald statistic. These results hold even when RDMPs were manipulated to capture a wider gamut of decision making through variations on framing and truncation of choices

where options were removed (Reyna et al., 2014). These results are comparable to carefully crafted human conducted analysis using the BR model which also predicted 82 of 88 experiments.

To further demonstrate the parsimony of our model compared to the BR and null models by applying the AIC and BIC metrics under a null CDM model where parameters are set to 0, we get AIC=14941 and BIC=14950. Whereas under the null model of the BR model, AIC=14981 and BIC=14986. In the best cases, our model outperforms all variations of the BR baseline model, with our model attaining AIC=13374 and BIC=13383 compared to AIC=13409 and BIC=13510. Furthermore, taking into consideration the relative likelihood ratio (RLR) to compare models using the AIC scores, $\exp((13374 - 13409)/2) = 2.5 \times 10^{-8}$, yields a significant result where the BR baseline model is only 2.5×10^{-8} times as probable as our model to minimize the information loss. Thus, our model attains better goodness-of-fit compared to the BR model while using significantly fewer parameters, 3 compared to up to 172 in the BR model.

IL-RDM results. Table 2 displays the average 5-fold cross-validation result predicting all 121 individuals' decisions for all 38 questions. Our model with a modest 63.19% accuracy outperforms all benchmark algorithms which hover around 50% for sentiment and 60% for pretrained language model baselines. This reinforces that IL-RDM is a more challenging problem and although sentiment analysis is important for decision-making, current SOTA sentiment analysis is not suitable for IL-RDM and only performs comparably to random choice. It is worth noting that while pretrained language models can be naively applied to IL-RDM with competitive results, they can not be naively applied to GL-RDM which requires the simultaneous predictions of two distributions of choices across frames where often the same RDM and choices is used across all experiments within a category.

Also displayed in the table are results when using transformers as encoders within Cat2Vec and results from a minor ablation study. For IL-RDM, transformers do not significantly improve accuracy as the resulting predicted categories and sentiments of categories from RDM-choices are highly similar between encoders. In the ablation study where choices are derived based on preferred choices at different levels of representation, i.e., $\text{CDM}_{\text{Categorical}}$ and $\text{CDM}_{\text{Interval}}$, reinforces the full

expressiveness of our model comes from the consensus between levels of representation and influence of individual differences.

Algorithm	Accuracy (%)	Min (%)	Max (%)
Random	49.87 ± 0.4	48.99	51.23
Vader	51.37 ± 0.9	48.64	54.33
XLNet	49.83 ± 2.1	44.21	56.31
ULMFiT	47.08 ± 2.1	39.31	53.30
BERT	61.94 ± 0.7	**60.80**	64.70
GTP2	53.65 ± 1.1	51.44	57.95
roBERTa	60.83 ± 1.7	57.23	65.87
CDM (Bi-LSTM)	62.47 ± 0.6	60.43	64.11
CDM (BERT)	62.30 ± 1.2	59.54	65.87
CDM (GTP2)	$\mathbf{63.19 \pm 1.2}$	60.64	**68.00**
CDM (roBERTa)	62.68 ± 1.2	59.43	66.35
Sentiment Cat2Vec	53.72 ± 1.4	48.04	56.67
$CDM_{Categorical}$	51.86 ± 1.3	46.38	54.33
$CDM_{Interval}$	53.17 ± 1.2	48.40	55.84

Table 2: Individual-Level Experiment Results

6.5 Error Analysis and Discussion

To understand the shortfall of our model for both GL-RDM and IL-RDM, we analyse cases in which our model fails to predict human decision-making. In GL-RDM, of the 6 experiments that our model did not successfully predict, 3 of these ((4),(5) and (7) in table 1) were not predicted by the BR baseline model indicating problems with parameter estimation using JLOO as these experiments are outliers with relatively significant differences in LORs within their respective categories.

In IL-RDM, inconsistencies exist across 2-RDMPs due in part to the within-subject design as participants may notice the underlying problem causing them to compare between problems rather than independently (Kahneman and Frederick, 2002). For example, figure 4 shows loss frames where individuals overwhelming preferred the safe choice, e.g., Q2, Q4 and most gain-loss pairs do not show a clear distinction between safe and risky choices in opposing frames, e.g. (Q26, Q2). Both cases are inconsistent with psychological studies.
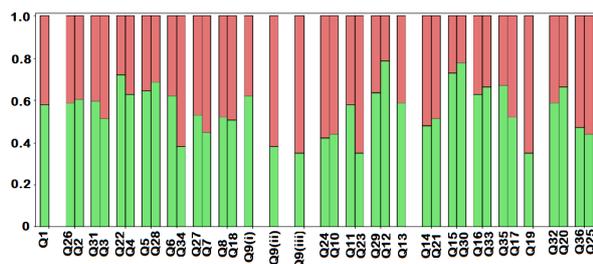

Figure 4: Ratio of choices for all questions with gain and loss frames grouped.

Quantity of data is also an issue. While the number of participants and questions answered were quite large for psychology experiments, this dataset is relatively small for machine learning tasks. Figure 5(A) shows the histogram of all 121 individuals' test accuracy on one fold. Overall, our model can predict most individual's choices accurately, but the average is lowered by some individuals our model cannot predict due to inconsistencies mentioned in (1). Also, since the size of each fold is relatively small, containing 7-8 test RDPs, any RDP not predicted correctly will cause a large decrease in accuracy. Figure 5(B) shows the percentage of correct choices our model predicts for all individuals from the combined 5-fold test questions. On average, our model predicts gain and loss RDMPs relatively equally with accuracy of 65.73% and 62.84% respectively. However, RDMPs with *"both frame"*, contains choices with combined gain and loss wording, cannot be predicted by our model due to this duality with an average of 42.73%.
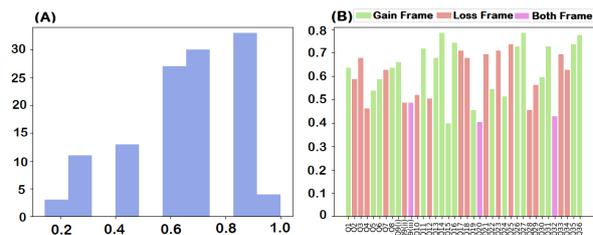
Figure 5: (A) Histogram of individual accuracy on one fold and (B) combined k-fold test accuracy per question

7 Conclusion and Future Work

This paper provides the first steps into a fully computational framework of risky decision-making, which adopts the cognitive and psychological basis of FTT with our model outperforming baselines in individual and group RDP prediction. Potential applications of our model are wide-ranging for scenarios in which predicting and understanding the characteristics of human risky decision-making is pivotal, e.g., the design of safety mechanisms based on how people make decisions in risky scenarios or in improving personalised recommendation systems based on understanding the users' personal traits and how they make decisions. Future work, therefore, involves adapting our model towards real-world applications, exploration of generalised decision-making and the design and evaluation of sophisticated end-to-end machine learning models for text-based decision-making.

References

Maurice Allais. 1953. Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica: Journal of the Econometric Society*, pages 503–546.

Charles J Brainerd and Valerie F Reyna. 1990. Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review*, 10(1):3–47.

David Broniatowski and Valerie Reyna. 2014. A mathematical formalization of fuzzy trace theory. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 36.

David A Broniatowski and Valerie F Reyna. 2018. A formal model of fuzzy-trace theory: Variations on framing effects and the allais paradox. *Decision*, 5(4):205.

Jerome Busemeyer, Zheng Wang, and Richard Shiffrin. 2015. Bayesian model comparison favors quantum over standard decision theory account of dynamic inconsistency. *Decision*, 2:1–12.

Jerome R Busemeyer and Yi-Min Wang. 2000. Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology*, 44(1):171–189.

John T Cacioppo, Richard E Petty, Jeffrey A Feinstein, and W Blair G Jarvis. 1996. Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological bulletin*, 119(2):197.

Paul S Carlin. 1990. Is the allais paradox robust to a seemingly trivial change of frame? *Economics Letters*, 34(3):241–244.

Li Chen, Marco De Gemmis, Alexander Felfernig, Pasquale Lops, Francesco Ricci, and Giovanni Semeraro. 2013. Human decision making and recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 3(3):1–7.

John Conlisk. 1989. Three variants on the allais example. *The American Economic Review*, pages 392–407.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

James N Druckman. 2001a. Evaluating framing effects. *Journal of economic psychology*, 22(1):91–101.

James N Druckman. 2001b. Using credible advice to overcome framing effects. *Journal of Law, Economics, and Organization*, 17(1):62–82.

James N. Druckman and Rose McDermott. 2008. Emotion and the framing of risky choice. *Political Behavior*, 30(3):297–321.

Ward Edwards. 1954. The theory of decision making. *Psychological bulletin*, 51(4):380.

N. Fagley and Paul Miller. 1990. The effect of framing on choice: Interactions with risk taking propensity, cognitive style, and sex. *Personality and Social Psychology Bulletin*, 16:496–510.

Peter Fischer, Eva Jonas, Dieter Frey, and Andreas Kastenmüller. 2008. Selective exposure and decision framing: The impact of gain and loss framing on confirmatory information search after decisions. *Journal of Experimental Social Psychology*, 44(2):312–320.

Milton Friedman and Leonard J Savage. 1952. The expected-utility hypothesis and the measurability of utility. *Journal of Political Economy*, 60(6):463–474.

CHE Gilbert and Erric Hutto. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf*, volume 81, page 82.

Faruk Gul, Wolfgang Pesendorfer, et al. 2008. Measurable ambiguity. *Levine's Bibliography*.

Thorvald Haerem, Bård Kuvaas, Bjørn T Bakken, and Tone Karlsen. 2011. Do military decision makers behave as predicted by prospect theory? *Journal of Behavioral Decision Making*, 24(5):482–497.

Scott Highhouse and Payam Yüce. 1996. Perspectives, perceptions, and risk-taking behavior. *Organizational Behavior and Human Decision Processes*, 65(2):159–167.

D Sunshine Hillygus and Todd G Shields. 2005. Moral issues and voter decision making in the 2004 presidential election. *PS: Political Science and Politics*, 38(2):201–209.

Jeremy Howard and Sebastian Ruder. 2018. Universal language model fine-tuning for text classification. *arXiv preprint arXiv:1801.06146*.

Steffen Huck and Wieland Müller. 2012. Allais for all: Revisiting the paradox in a large representative sample. *Journal of Risk and Uncertainty*, 44(3):261–293.

Jerwen Jou, James Shanteau, and Richard Harris. 1996. An information processing view of framing effects: The role of causal schemas in decision making. *Memory cognition*, 24:1–15.

Alex Kacelnik and Melissa Bateson. 1997. Risk-sensitivity: crossroads for theories of decision-making. *Trends in cognitive sciences*, 1(8):304–309.

Daniel Kahneman. 2003. Maps of bounded rationality: Psychology for behavioral economics. *American economic review*, 93(5):1449–1475.

Daniel Kahneman and Shane Frederick. 2002. Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and biases: The psychology of intuitive judgment*, 49:81.

Daniel Kahneman and Amos Tversky. 1980. Prospect theory. *Econometrica*, 12.

Anton Kühberger. 1995. The framing of decisions: A new look at old problems. *Organizational Behavior and Human Decision Processes*, 62(2):230–240.

Anton Kühberger and Patricia Gradl. 2013. Choice, rating, and ranking: Framing effects with different response modes. *Journal of Behavioral Decision Making*, 26(2):109–117.

Anton Kühberger and Carmen Tanner. 2010. Risky choice framing: Task versions and a comparison of prospect theory and fuzzy-trace theory. *Journal of behavioral decision making*, 23(3):314–329.

Anton Kühberger. 1995. The framing of decisions: A new look at old problems. *Organizational Behavior and Human Decision Processes*, 62(2):230 – 240.

Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *International conference on machine learning*, pages 1188–1196.

Robyn LeBoeuf and Eldar Shafir. 2003. Deep thoughts and shallow frames: On the susceptibility to framing effects. *Journal of Behavioral Decision Making*, 16:77 – 92.

Jennifer S Lerner, Ye Li, Piercarlo Valdesolo, and Karim S Kassam. 2015. Emotion and decision making. *Annual review of psychology*, 66:799–823.

Irwin Levin, Gary Gaeth, Judy Schreiber, and Marco Lauriola. 2002. A new look at framing effects: Distribution of effect sizes, individual differences, and independence of types of effects. *Organizational Behavior and Human Decision Processes*, 88:411–429.

Zhiyuan Liu, Yankai Lin, and Maosong Sun. 2020. Representation learning and nlp. In *Representation Learning for Natural Language Processing*, pages 1–11. Springer.

David R Mandel. 2001. Gain-loss framing and choice: Separating outcome formulations from descriptor formulations. *Organizational Behavior and Human Decision Processes*, 85(1):56–76.

David R Mandel. 2014. Do framing effects reveal irrational choice? *Journal of Experimental Psychology: General*, 143(3):1185.

Christopher B Mayhorn, Arthur D Fisk, and Justin D Whittle. 2002. Decisions, decisions: Analysis of age, cohort, and time of testing on framing of risky decision options. *Human Factors*, 44(4):515–521.

Daniel McFadden. 2001. Economic choices. *American economic review*, 91(3):351–378.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Paul Miller and N. Fagley. 1991. The effects of framing, problem variations, and providing rationale on choice. *Personality and Social Psychology Bulletin*, 17:517–522.

David Nadeau and Satoshi Sekine. 2007. A survey of named entity recognition and classification. *Lingvisticae Investigationes*, 30(1):3–26.

Hidetaka Okder. 2012. The illusion of the framing effect in risky decision making. *Journal of Behavioral Decision Making*, 25(1):63–73.

Valerie F Reyna and Charles J Brainerd. 1991. Fuzzy-trace theory and framing effects in choice: Gist extraction, truncation, and conversion. *Journal of Behavioral Decision Making*, 4(4):249–262.

Valerie F Reyna, Christina F Chick, Jonathan C Corbin, and Andrew N Hsia. 2014. Developmental reversals in risky decision making: Intelligence agents show larger decision biases than college students. *Psychological science*, 25(1):76–84.

Michael Rönnlund, Erik Karlsson, Erica Laggnäs, Lisa Larsson, and Therese Lindström. 2005. Risky decision making across three arenas of choice: Are younger and older adults differently susceptible to framing effects? *The Journal of general psychology*, 132:81–92.

Uzi Segal. 1987. The ellsberg paradox and risk aversion: An anticipated utility approach. *International Economic Review*, pages 175–202.

Herbert A Simon. 1993. Decision making: Rational, nonrational, and irrational. *Educational Administration Quarterly*, 29(3):392–411.

Kaye Stacey. 2015. The international assessment of mathematical literacy: Pisa 2012 framework and items. In *Selected regular lectures from the 12th International Congress on Mathematical Education*, pages 771–790. Springer.

Keith E. Stanovich and Richard F. West. 1998. Individual differences in framing and conjunction effects. *Thinking & Reasoning*, 4(4):289–317.

Jeannette M Stein. 2012. Framing effects: The influence of handedness and access to right hemisphere processing. *Laterality: Asymmetries of Body, Brain and Cognition*, 17(1):98–110.

Stanley Smith Stevens et al. 1946. On the theory of scales of measurement.

Kazuhisa Takemura. 1994. Influence of elaboration on the framing of decision. *The Journal of Psychology*, 128(1):33–39.

R Scott Tindale, Susan Sheffey, and Leslie A Scott. 1993. Framing and group decision-making: Do cognitive changes parallel preference changes? *Organizational Behavior and Human Decision Processes*, 55(3):470–485.

Amos Tversky and Daniel Kahneman. 1981. The framing of decisions and the psychology of choice. *science*, 211(4481):453–458.

Amos Tversky and Daniel Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4):297–323.

John Von Neumann and Oskar Morgenstern. 2007. *Theory of games and economic behavior (commemorative edition)*. Princeton university press.

Xiao Tian Wang. 1996. Framing effects: Dynamics and task domains. *Organizational behavior and human decision processes*, 68(2):145–157.

Xiao-Tian Wang and Victor S Johnston. 1995. Perceived social context and risk preference: A reexamination of framing effects in a life-death decision problem. *Journal of Behavioral Decision Making*, 8(4):279–293.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. *arXiv preprint arXiv:1906.08237*.

Milan Zeleny. 2012. *Multiple criteria decision making Kyoto 1975*, volume 123. Springer Science & Business Media.

Yin-ling Zhang, Li-jun Xiao, Yu Ma, and Dan-min Miao. 2008. Effect of framing on risky choice: One case study in china. *Social Behavior and Personality: an international journal*, 36(5):651–658.

Yinling Zhang and Danmin Miao. 2008. Social cues and framing effects in risky decisions among chinese military students. *Asian Journal of Social Psychology*, 11(3):241–246.

A Appendix

A.1 Final Decision Algorithm

Algorithm 1 corresponds to the algorithm mentioned in section 6.1 of the main paper.

A.2 Evaluation Metric Calculations

The calculations for the second type of evaluation metric we use to compare the parsimony of our model against baseline algorithms are the Akaike information criterion (AIC) and Bayesian information criterion (BIC) are given by equations (13) and (14) using the log-likelihood calculated by equations

---

Algorithm 1 Computational Decision Making

Input: n**RDMP, NUM, NFC, RS**
Output: Decision/Preferred Choice
1: **for** choice in n-RDMP **do**
2:     **if** CU(Pref$_{Cat}$, NFC) < CU(choice, NFC) **then**
3:         Pref$_{Cat}$ = choice
4:     **if** IU(Pref$_{Int}$, NUM) < IU(choice, NUM) **then**
5:         Pref$_{Int}$ = choice
6: **if** Pref$_{Cat}$ = Pref$_{Int}$ **then**
7:     **return** Pref$_{Cat}$
8: **else if** Uniform(0, 1) ≤ RiskSensitivity(RS) **then**
9:     **return** Riskiest Choice
10: **else**
11:     **return** Safest Choice

---

and (12). In these equations $n_{1,1}$ is the number of people who chose the first choice (safe choice) in the first problem (gain frame), $p_{1,1}$ is the predicted proportion of subjects who chose the first choice (safe choice) in the first problem (gain frame), etc. For the AIC, k is the total number of parameters of our model, 3 which correspond to each individual difference and in BIC, n is the total number of data points, 176 to represent the gain and loss frames in the 88 human experiments.

$$\ln[L(y_i)] = n_{1,1} \ln p_{1,1} + n_{1,2} \ln p_{1,12} \\ + n_{2,1} \ln p_{2,1} + n_{2,2} \ln p_{2,2} \quad (11)$$

$$\ln[L(y)] = \sum_i \ln[L(y_i)] \quad (12)$$

$$\text{AIC} = 2k - 2\ln[L(y)] \quad (13)$$

$$\text{BIC} = k\ln(n) - 2\ln[L(y)] \quad (14)$$

To compare models using AIC, the relative likelihood ratio (RLR) given in equation 15 can be applied which compares the probability that the BR baseline model minimises the estimated information loss compared to our CDM model given that $\text{AIC}_{CDM} \leq \text{AIC}_{BR}$ where AIC_{BR} and AIC_{CDM} are corresponding AIC scores of each model.

$$\text{RLR} = \exp\left(\frac{\text{AIC}_{CDM} - \text{AIC}_{BR}}{2}\right) \quad (15)$$

A.3 Individual Level Questionnaire

Full inventory of all 36 questions used in the Individual 2-RDMP Prediction dataset:
Q1: Which of the following options do you prefer?
    (a) A sure win of $30

(b) 80% chance to win $45

Q2: Imagine that 6000 pieces of precious paintings in a world-famous museum are accidentally exposed to a disastrous chemical pollution. Two alternative plans to rescue these art treasures have been proposed. Assume that the exact estimates of the consequences of the plans made by scientists are as follows:

(a) If plan A is adopted, 4000 pieces will be destroyed by the chemical pollution.

(b) If plan B is adopted, there is a one-third probability that none of these paintings will be destroyed, and two-thirds probability that all 6000 of these paintings will be destroyed.

Q3: A large car manufacturer has recently been hit with a number of economic difficulties and it appears as if three plants need to be closed and 6000 employees laid off. The vice-president of production has been exploring alternative ways to avoid this crisis and has developed two plans:

(a) Plan C: This plan will result in the loss of 2 plants and 4000 jobs.

(b) Plan D: This plan has a 2/3 probability of resulting in the loss of 3 plants and all 6000 jobs, but has a 1/3 probability of losing no plants and no jobs

Q4: Imagine you recieve a letter from the president of a subsidiary describing a dilemma concerning whether to fight an impending patent violation suit or settle out of court that reads: If we do not agree to this proposal, PMG will file their suit. Going to court would involve the possibility of losing $1,100,000 in damages and losing the Duraplast line. If we win in court, we will incur a small sum for legal expenses. Our corporate lawyer, Mr. Bell, and our outside law firm estimate that we have a 2 in 3 chance of losing the case.

(a) Agree to the proposal (no lawsuit)

(b) Disagree to the proposal: 2/3 chance of losing the lawsuit and incurring costs of $1100000

Q5: Imagine that you have lung cancer and you must choose between two therapies: surgery and radiation. Surgery for lung cancer involves an operation on the lungs. Most patients are in the hospital for two or three weeks and have some pain around their incisions; they spend a month or so recuperating at home. After that, they generally feel fine. Radiation therapy for lung cancer involves the use of radiation to kill the tumor and requires coming to the hospital about four times a week for six weeks. Each treatment takes a few minutes and during the treatment, patients lie on a table as if they were having an x-ray. During the course of the treatment, some patients develop nausea and vomiting, but by the end of the six weeks they also generally feel fine. Thus, after the initial six or so weeks, patients treated with either surgery or radiation therapy feel about the same.

(a) Surgery: Of 100 people having surgery, 90 live through the postoperative period, 68 are alive at the end of one year and 34 are alive at the end of five years.

(b) Radiation Therapy: Of 100 people having radiation therapy, all live through treatment, 77 are alive at the end of one year and 22 are alive at the end of five years.

Q6: Imagine that you brought $6000 worth of stock from a company that has just filed a claim for bankruptcy recently. The company now provides you with two alternatives to recover some of your money.

(a) You will save $2000 of your money

(b) You will take part in a random drawing procedure with exactly a one-third probability of saving all $6000 of your money, and two-thirds probability of saving none of your money.

Q7: Imagine that in one particular state it is projected that 1000 students will dropout of school during the year, two programs have been prosed to address this problem, but only one can be implemented. Based on other states experiences with programs, estimates of the outcomes that can be expected for each program can be made.

(a) Program 1: 600 of the 1000 students will drop out of school

(b) Program 2: 2/5 chance that none of the 1000 students will drop out of school and 3/5 chance that all 1000 students will drop out of school

Q8: Assume that you have just been given a gift of $1000.

(a) Taking an additional $500 for sure.

(b) Flipping a coin and winning another $1000

if heads comes up or getting no additional money if tails comes up.

Q9(i): Imagine that you face the following pair of concurrent decisions. First examine both decisions, then indicate the options you prefer.

(a) A sure gain of $240

(b) 25% chance to gain $1000, and 75% chance to gain nothing

Q9(ii): Imagine that you face the following pair of concurrent decisions. First examine both decisions, then indicate the options you prefer.

(a) A sure loss of $750

(b) 75% chance to lose $1000, and 25% chance to lose nothing

Q9(iii): Imagine that you face the following pair of concurrent decisions. First examine both decisions, then indicate the options you prefer.

(a) 25% chance to win $240, and 75% chance to lose $760

(b) 25% chance to win $250, and 75% chance to lose $750

Q10: You are staying in a hotel room on vacation. You paid $6.95 to see a movie on pay TV. After 5 minutes you are bored and the movie seems pretty bad. Would you continue to watch the movie or not?

(a) Continue to watch

(b) Turn it off and lose $6.95

Q11: Imagine that your country is preparing for the outbreak of an unusual disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:

(a) If Program A is adopted, 200 people will be saved

(b) If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved

Q12: Imagine that you have decided to see a play where admission is $10 per ticket. As you enter the theatre you discover that you have lost a $10 bill.

(a) Still pay $10 for a ticket for the play

(b) Don't pay $10 for a ticket for the play

Q13: Consider the following two stage game. In the first stage, there is a 75% chance to end the game without winning anything, and a 25% chance to move into the second stage. If you reach the second stage, you have a choice between: A sure win of $30 and 80% chance to win $45

(a) A sure win of $30

(b) 80% chance to win $45

Q14: Imagine that six people in your family, including both of your parents, your brothers and your sisters, are infected by a fatal disease. Two alternative medical plans to treat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the plans are as follows:

(a) If plan A is adopted, two of them will be saved.

(b) If plan B is adopted, there is a one-third probability that all six of them will be saved, and two-thirds probability that none of them will be saved.

Q15: Your are presented with the following report from the head of a special team assigned to investigate the prospects of a project in Arizona: Our new analysis indicates that, if we choose to compete with ATC, we would face the possibility of capturing only a small market share. This would give us an after-tax return on investment of as little as 10%, while capturing a large market share would give us a return of 22%. We estimate our chance of getting a small market share to be 2 in 3. If we were to team up with ATC on the terms proposed, our return would be 14% after tax, with the same total investment.

(a) Compete with ATC: 1/3 chance of gaining a large market share of 22% and 2/3 chance of gaining a small market share of 10%

(b) Don't compete with ATC: 100% chance of capturing 14% market share

Q16: A committee found a fish disease in a nearby lake. About 12 fish species (among them the most popular dining fish) have the Proliferative Kidney Disease (PKD). This is a chronically developing infectious disease which can have deadly consequences for the fish. Young fish are especially susceptible, while others seem to be immune against an infection. Experts suggest that PKD is one cause of declining fish catches. The researchers

assume human activities and water pollution foster the spread of the disease. They are considering releasing more fish into the lake to control the epidemic. Imagine that you are a government official of the adjacent village.

(a) Option A: If the release of fish is implemented, 4 fish species will survive.

(b) Option B: If the release of fish is implemented, there is 1/3 probability that all of the 12 fish species will survive, and 2/3 probability that none of them will survive.

Q17: Imagine a refinery that processes petroleum products. An investigation found that due to tank leaks, both soil and drinking water became contaminated. Due to this contamination 720 children from the adjacent village have a fatal disease. There is agreement among experts that children will not suffer health problems, provided they have a strong immune system. Otherwise, it is likely that children will have serious health problems. A vaccine against this disease has been developed and tested. However, the vaccine sometimes can cause side effects that can be fatal too. You are an environmental activist with much influence on the local hospital and you have to decide if you want to lobby for the vaccination or not.

(a) Option C: If the vaccination is adopted, the health of 480 children will be damaged for sure.

(b) Option D: If the vaccination is adopted, there is a one-third probability that the health of none of the 720 children will be damaged, and a two-thirds probability that the health of all 720 of them will be damaged.

Q18: Assume that you have just been given a gift of $2000. But you now are forced to choose between the following two alternatives:

(a) Losing $500 for sure

(b) Flipping a coin and losing $1000 if heads comes up or losing nothing if tails comes up

Q19: Which of the following options do you prefer?

(a) 25% chance to win $30

(b) 20% chance to win $45

Q20: Imagine that you are about to purchase a jacket for $125, and a calculator for $15. The calculator salesman informs you that the calculator you wish to buy is on sale for $10 at the other branch of the store, located 20 minutes drive away.

(a) Make the trip to the other store and save 5 dollars but lose 20 minutes

(b) Don't make the trip to the other store and save 20 minutes but lose 5 dollars

Q21: Imagine that six people in your family, including both of your parents, your brothers and your sisters, are infected by a fatal disease. Two alternative medical plans to treat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the plans are as follows:

(a) If plan A is adopted, four of them will die.

(b) If plan B is adopted, there is a one-third probability that none of them will die, and two-thirds probability that all six of them will die.

Q22: Imagine you recieve a letter from the president of a subsidiary describing a dilemma concerning whether to fight an impending patent violation suit or settle out of court that reads: If we do not agree to this proposal, PMG will file their suit. Going to court would involve the possibility of keeping the Duraplast line and incurring only a small sum for legal expenses. If we lose in court, we will incur $1,100,000 in damages. Our corporate lawyer, Mr. Bell, and our outside law firm agree that we have a 1 in 3 chance of winning the case.

(a) Agree to the proposal (no lawsuit)

(b) Disagree to the proposal: 1/3 chance of winning the case

Q23: Imagine that your country is preparing for the outbreak of an unusual disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:

(a) If Program C is adopted 400 people will die.

(b) If Program D is adopted there is 1/3 probability that no one will die, and 2/3 probability that 600 people will die.

Q24: You are staying in a hotel room on vacation. You turn on the TV and there is a movie on. After 5

minutes you are bored and the movie seems pretty bad. Would you continue to watch the movie or not?

(a) Continue to watch

(b) Turn it off

Q25: Imagine that six people are infected by a fatal disease. Two alternative medical plans to treat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the plans are as follows:

(a) If plan A is adopted, four people will die.

(b) If plan B is adopted, there is a one-third probability that none of them will die, and two-thirds probability that all six people will die.

Q26: Imagine that 6000 pieces of precious paintings in a world-famous museum are accidentally exposed to a disastrous chemical pollution. Two alternative plans to rescue these art treasures have been proposed. Assume that the exact estimates of the consequences of the plans made by scientists are as follows:

(a) If plan A is adopted, 2000 pieces will be saved from the chemical pollution.

(b) If plan B is adopted, there is a one-third probability that all the 6000 paintings will be saved, and two-thirds probability that none of these paintings will be saved.

Q27: Imagine that in one particular state it is projected that 1000 students will dropout of school during the year, two programs have been prosed to address this problem, but only one can be implemented. Based on other states experiences with programs, estimates of the outcomes that can be expected for each program can be made.

(a) Program 1: 400 of the 1000 students will stay in school

(b) Program 2: 2/5 chance that all 1000 students will stay in school and 3/5 chance that none of the 1000 will stay in school

Q28: Imagine that you have lung cancer and you must choose between two therapies: surgery and radiation. Surgery for lung cancer involves an operation on the lungs. Most patients are in the hospital for two or three weeks and have some pain around their incisions; they spend a month or so recuperating at home. After that, they generally feel fine.

Radiation therapy for lung cancer involves the use of radiation to kill the tumor and requires coming to the hospital about four times a week for six weeks. Each treatment takes a few minutes and during the treatment, patients lie on a table as if they were having an x-ray. During the course of the treatment, some patients develop nausea and vomiting, but by the end of the six weeks they also generally feel fine. Thus, after the initial six or so weeks, patients treated with either surgery or radiation therapy feel about the same.

(a) Surgery: Of 100 people having surgery, 10 die during surgery or the postoperative period, 32 die by the end of one year and 66 die by the end of five years.

(b) Radiation Therapy: Of 100 people having radiation therapy, none die during treatment, 23 die by the end of one year and 78 die by the end of five years.

Q29: Imagine that you have decided to see a play and paid the admission price of $10 per ticket. As you enter the theatre you discover that you have lost the ticket. The seat was not marked and the ticket cannot be recovered.

(a) Pay $10 for another ticket

(b) Don't pay $10 for another ticket

Q30: Your are presented with the following report from the head of a special team assigned to investigate the prospects of a project in Arizona: Our new analysis indicates that, if we choose to compete with ATC, we would have the possibility of capturing a large market share. This would give us an after-tax return on investment of as much as 22%, while capturing a small market share would give us a return of only 10%. We estimate a 1 in 3 chance of getting a large market share. If we were to team up with ATC on the terms proposed, our return would be 14% after tax, with the same total investment.

(a) Compete with ATC

(b) Don't compete with ATC

Q31: A large car manufacturer has recently been hit with a number of economic difficulties and it appears as if three plants need to be closed and 6000 employees laid off. The vice-president of production has been exploring alternative ways to avoid this crisis and has developed two plans:

(a) Plan A: This plan will save 1 plant and 2000

jobs

(b) Plan B: : This plan has a 1/3 probability of saving all 3 plants and all 6000 jobs, but has a 2/3 probability of saving no plants and no jobs

Q32: Imagine that you are about to purchase a jacket for $15, and a calculator for $125. The calculator salesman informs you that the calculator you wish to buy is on sale for $120 at the other branch of the store, located 20 minutes drive away.

(a) Make the trip to the other store and save 5 dollars but lose 20 minutes

(b) Don't make the trip to the other store and save 20 minutes but lose 5 dollars

Q33: A committee found a fish disease in a nearby lake. About 12 fish species (among them the most popular dining fish) have the Proliferative Kidney Disease (PKD). This is a chronically developing infectious disease which can have deadly consequences for the fish. Young fish are especially susceptible, while others seem to be immune against an infection. Experts suggest that PKD is one cause of declining fish catches. The researchers assume human activities and water pollution foster the spread of the disease. They are considering releasing more fish into the lake to control the epidemic. Imagine that you are a government official of the adjacent village.

(a) Option C: If the release of fish is implemented, 8 fish species will die.

(b) Option D: If the release of fish is implemented, there is 2/3 probability that none of the 12 fish species will die, and 1/3 probability that all of the 12 fish species will die.

Q34: Imagine that you brought $6000 worth of stock from a company that has just filed a claim for bankruptcy recently. The company now provides you with two alternatives to recover some of your money.

(a) You will lose $4000 of your money

(b) You will take part in a random drawing procedure with exactly a two-thirds probability of losing $6000 all of your money, and one-third probability of not losing any of your money

Q35: Imagine a refinery that processes petroleum products. An investigation found that due to tank leaks, both soil and drinking water became contaminated. Due to this contamination 720 children from the adjacent village have a fatal disease. There is agreement among experts that children will not suffer health problems, provided they have a strong immune system. Otherwise, it is likely that children will have serious health problems. A vaccine against this disease has been developed and tested. However, the vaccine sometimes can cause side effects that can be fatal too. You are an environmental activist with much influence on the local hospital and you have to decide if you want to lobby for the vaccination or not.

(a) Option A: If the vaccination is adopted, the health of 240 children will be saved for sure.

(b) Option B: If the vaccination is adopted, there is a one-third probability that the health of all of the 720 children will be saved, and a two-thirds probability that the health of none of them will be saved.

Q36: Imagine that six people are infected by a fatal disease. Two alternative medical plans to treat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the plans are as follows:

(a) If plan A is adopted, two people will be saved.

(b) If plan B is adopted, there is a one-third probability that all six people will be saved, and two-thirds probability that none of them will be saved.

A.4 Group Level Results

Table A.4 shows the full table corresponding to table 2 in the paper with extra information regarding the choices made in each human experiment. P represents the percentage of people that chose the risky choice given either the gain frame or loss frames, the true P_{gain} and P_{loss} values, and n represents the number of human subjects who answered each frame which corresponds to the number of decisions our model made to produce results in the table.

Reference	Gain Frame		Loss Frame		Actual LOR	BR Baseline			CDM Model		
	n	P	n	P		LOR	SE	χ^2	LOR	SE	χ^2
Standard ADP; one presentation, between-subjects, low PISA											
Tversky and Kahneman (1981)	**152**	**28**	**155**	**78**	**2.20**	**1.65***	**.26**	**4.34**	**1.86**	**.26**	**1.83**
Reyna and Brainerd (1991)	36	53	36	81	1.31	1.72	.54	.57	1.79	.54	.79
Tindale et al. (1993)	144	42	144	79	1.63	1.71	.26	.10	1.77	.26	.27
Wang and Johnston (1995)	50	40	50	68	1.16	1.73	.42	1.83	1.74	.45	1.71
Highhouse and Yüce (1996)	122	29	122	74	1.94	1.68	.29	.82	1.69	.28	.80
Wang (1996)	31	42	34	77	1.50	1.71	.54	.14	1.62	.55	.05
Stanovich and West (1998)	148	32	144	65	1.37	1.74	.25	2.34	1.78	.26	2.43
Druckman (2001a)	50	32	55	77	1.93	1.70	.44	.27	1.51	.43	.94
Druckman (2001b)	69	32	79	76	1.91	1.70	.37	.34	1.66	.36	.49
Mayhorn et al. (2002), Young	29	24	29	86	2.98	1.68	.69	3.52	1.50	.58	5.89*
Mayhorn et al. (2002), Older	29	21	29	69	2.14	1.70	.61	.53	1.54	.58	1.06
LeBoeuf and Shafir (2003), Exp #1	48	49	55	56	1.40	1.74	.25	1.77	1.54	.43	.12
LeBoeuf and Shafir (2003), Exp #2	147	25	146	57	1.47	1.71	.43	.32	1.58	.25	.13
(Stein, 2012)	47	40	57	68	1.16	1.73	.41	1.89	1.72	.44	1.55
									TOTAL of 14 predicted: 13 (93%)		
Standard ADP; one presentation, between-subjects, high PISA											
Takemura (1994)	45	20	45	69	2.18	1.39	.49	2.56	1.52	.47	2.02
Mandel (2001)	26	54	26	85	1.55	1.44	.67	.03	1.40	.61	.08
Fischer et al. (2008)	17	36	17	77	1.78	1.43	.76	.21	1.43	.78	.20
Zhang and Miao (2008) #1	65	66	68	87	1.21	1.47	.44	.34	1.44	.37	.29
Zhang and Miao (2008) #2	45	67	48	88	1.25	1.46	.54	.14	1.41	.45	.08
Zhang et al. (2008), Military	134	54	130	83	1.44	1.44	.29	.00	1.48	.27	.05
Zhang et al. (2008), Civilian	60	65	58	90	1.54	1.43	.51	.04	1.48	.40	.05
Haerem et al. (2011)	29	59	26	73	.65	1.48	.58	2.02	1.46	.59	1.93
Okder (2012)	52	37	53	76	1.68	1.42	.43	.34	1.66	.43	.01
Kühberger and Gradl (2013), Exp #1	63	32	63	68	1.53	1.43	.38	.06	1.45	.39	.02
Kühberger and Gradl (2013), Exp #2	14	57	15	73	.72	1.46	.80	.85	1.52	.86	.87
Mandel (2014), Exp #2	38	42	38	74	1.35	1.45	.49	.04	1.57	.51	.16
Mandel (2014), Exp #3	25	32	25	80	2.14	1.42	.66	1.20	1.54	.63	.92
									TOTAL of 13 predicted: 13 (100%)		
Standard ADP; within-subjects, low PISA											
Stanovich and West (1998)	292	32	292	54	.9	.94	.24	1.58	.96	.17	.08
Levin et al. (2002)	102	28	102	56	1.2	.92	.30	.94	.94	.29	.74
LeBoeuf and Shafir (2003) Exp #2	**287**	**25**	**287**	**46**	**.57**	**1.05***	**.17**	**7.86**	**.81**	**.17**	**.52**
									TOTAL of 3 predicted: 3 (100%)		
Standard ADP; multiple presentations, between-subjects, low PISA											
Fagley and Miller (1990), Exp #1	94	51	96	70	.79	.95	.30	.27	1.38	.31	3.01
Fagley and Miller (1990), Exp #2	54	39	55	73	1.43	.92	.41	1.55	1.28	.41	.16
Miller and Fagley (1991)	23	43	23	67	.89	.94	.61	.01	1.40	.65	.40
Jou et al. (1996)	80	35	80	80	2.01	.87*	.36	9.66	1.13	.33	6.99*
Rönnlund et al. (2005) Young	32	41	32	69	1.19	.93	.53	.23	1.30	.55	.07
Rönnlund et al. (2005), Older	32	28	32	56	1.17	.94	.52	.20	1.40	.55	.16
									TOTAL of 6 predicted: 5 (83%)		
Standard ADP; multiple presentations, between-subjects, High PISA											
Kühberger (1995), Exp #1	25	48	23	78	1.36	.73	.64	1.31	.86	.61	.64
Kühberger (1995), Exp #2	16	56	14	57	.04	.81	.74	.81	1.10	.80	1.76
Druckman and McDermott (2008)	101	45	113	67	.94	.71	.28	1.84	.46	.28	2.57
									TOTAL of 3 predicted: 3 (100%)		
Allais Paradox gambles; low PISA											
Conlisk (1989)	236	49	236	86	1.83	1.68	.23	.44	1.70	.21	.55
Carlin (1990)	65	40	65	78	1.7	1.71	.39	0	1.75	.45	.03
									TOTAL of 2 predicted: 2 (100%)		
Allais Paradox gambles; low PISA											
Huck and Müller (2012) Laboratory	70	66	70	87	1.26	1.73	.44	1.14	1.32	.46	.03
									TOTAL of 1 predicted: 1 (75%)		
Other framing problems; multiple presentations, between-subjects, low PISA											
Reyna et al. (2014) College Students	63	35	63	55	.85	.95	.37	.02	.99	.37	.21
Reyna et al. (2014) Adults	54	40	54	60	.8	.95	.39	.01	1.12	.41	.60
Reyna et al. (2014) Experts	36	38	36	71	1.37	.93	.5	.18	.69	.49	2.01
									TOTAL of 3 predicted: 3 (100%)		
Other framing problems; multiple presentations, between-subjects, high PISA											
Kühberger (1995) Plant #1	25	52	23	83	1.48	.73	.68	1.61	.97	.61	.79
Kühberger (1995) Cancer #1	24	38	25	48	.43	.8	.58	.2	.21	.60	.11
Kühberger (1995) Plant #2	16	19	17	71	2.34	.7	.83	4.43	.34	.73	7.50*
Kühberger (1995) Cancer #2	16	69	14	64	-.2	.82	.78	1.39	.97	.82	2.11
									TOTAL of 4 predicted: 3 (75%)		
Other framing problems; multiple presentations, between-subjects, mixed PISA											
Kühberger and Tanner (2010) Water contamination	93	33	93	73	1.69	1.27	.32	1.78	1 .26	.31	2.01
Kühberger and Tanner (2010) Crops	93	33	93	59	1.06	1.36	.3	.92	1.32	.31	.64
Kühberger and Tanner (2010) Fish disease	93	28	93	59	1.32	1.32	.31	0	1.20	.31	.12
Kühberger and Tanner (2010) Endangered forest	93	24	93	55	1.37	1.31	.32	.03	1.23	.31	.16
									TOTAL of 4 predicted: 4 (100%)		
Zero-complement truncated framing problems, one presentation; framing manipulated between-subjects											
Reyna and Brainerd (1991)	35	51	36	58	.28	0	.48	.34	.41	.49	.07
Mandel (2001) Exp1	23	48	25	72	1.03	0	.61	2.86	.30	.60	1.47
Mandel (2001) Exp2	36	64	38	63	-.03	0	.48	0	.35	.48	.70
									TOTAL of 3 predicted: 3 (100%)		
Zero-complement truncated framing problems; multiple presentations, framing manipulated between-subjects											
Kühberger and Tanner (2010) Water contamination	93	54	93	65	.45	0	.3	2.21	0	.3	2.21
Kühberger and Tanner (2010) Crops	93	54	93	43	-.43	0	.3	2.14	.34	.30	6.83*
Kühberger and Tanner (2010) Fish disease	93	63	93	43	-.83	0	.3	7.68*	.37	.30	15.74*
Kühberger and Tanner (2010) Endangered forest	93	40	93	43	.13	0	.3	.2	.23	.30	.13
Reyna et al. (2014) College Students	63	43	63	49	.25	0	.36	.47	.26	.36	0
Reyna et al. (2014) Adults	54	51	54	55	.18	0	.39	.23	.29	.39	.75
Reyna et al. (2014) Experts	36	52	36	62	.41	0	.48	.74	.20	.48	.19
									TOTAL of 7 predicted: 5 (71.4%)		
Nonzero-complement truncated framing problems; one presentation, between-subjects, low PISA											
Reyna and Brainerd (1991)	35	26	37	81	2.52	3.44	.57	2.59	2.29	.57	.13
									TOTAL of 1 predicted: 1 (100%)		
Nonzero-complement truncated framing problems; multiple presentations, between-subjects, low PISA											
Reyna et al. (2014) College Students	63	23	63	60	1.61	1.9	.4	.51	2.19	.42	1.90

Reyna et al. (2014) Adults	54	26	54	73	2.05	1.87	.44	.15	1.83	.44	0.23
Reyna et al. (2014) Experts	36	20	36	81	2.84	1.85	.59	2.72	1.86	.54	3.29
									TOTAL of 3 predicted: 3 (100%)		
Nonzero-complement truncated framing problems; one presentation, between-subjects, low PISA											
Kühberger and Tanner (2010) Water contamination	93	25	93	85	2.84	2.61	.38	.39	2.31	.35	2.24
Kühberger and Tanner (2010) Crops	93	29	93	78	2.19	2.7	.34	2.28	2.27	.35	.09
Kühberger and Tanner (2010) Fish diseas	93	22	93	83	2.87	2.61	.37	.49	2.21	.34	3.48
Kühberger and Tanner (2010) Endangered forest	93	15	93	65	2.33	2.68	.36	.96	2.30	.35	.02
									TOTAL of 4 predicted: 4 (100%)		
Certain-option disambiguated problems; single presentation, between-subjects											
Mandel (2001) Exp #1	23	52	22	50	-.09	0	.6	.02	.28	.62	.34
Mandel (2014) Exp #3	22	41	24	50	.37	0	.6	.38	.03	.60	.31
									TOTAL of 2 predicted: 2 (100%)		
Certain-option disambiguated problems; multiple presentations, between-subjects											
Kühberger (1995) ADP #1	26	62	23	57	-.21	0	.58	.13	.18	.59	.45
Kühberger (1995) Plant #1	26	46	23	52	.24	0	.57	.18	.11	.59	.05
Kühberger (1995) Cancer #1	24	50	23	35	-.63	0	.6	1.1	.35	.60	2.61
Kühberger (1995) ADP #2	22	41	19	37	-.17	0	.64	.07	.28	.65	.47
Kühberger (1995) Plant #2	13	31	19	37	.27	0	.77	.13	.17	.75	.02
Kühberger (1995) Cancer #2	7	71	13	62	-.45	0	1.01	.19	.13	1.02	.28
									TOTAL of 6 predicted: 6 (100%)		
"400 not saved" certain-option disambiguated and truncated problems; multiple presentations, between-subjects, high PISA											
Kühberger (1995) ADP #1	25	60	23	43	-.67	-.79	.59	.04	.07	.59	1.64
Kühberger (1995) Plant #1	27	44	23	57	.49	-.88	.57	5.68*	.11	.58	.50
Kühberger (1995) Cancer #1	24	75	23	43	-1.36	-.74	.63	.98	.11	.60	6.21*
Kühberger (1995) ADP #2	16	50	19	37	-.54	-.79	.69	.13	.12	.70	.85
Kühberger (1995) Plant #2	14	57	14	50	-.29	-.8	.76	.45	.15	.79	.29
Kühberger (1995) Cancer #2	14	50	16	44	-.25	-.8	.73	.56	.15	.76	.26
									TOTAL of 5 predicted: 6 (83.3%)		
Certain-option disambiguated, zero-complement truncated problems; single presentation, between-subjects, high PISA											
Mandel (2014), Exp #3	26	58	25	32	-1.06	-1.46	.58	.45	.02	.58	3.59
									TOTAL of 1 predicted: 1 (100%)		
"400 not saved vs. 2/3 chance that 600 not saved" truncation problem; single presentation, framing manipulated between-subjects											
Mandel (2001) Exp #1	23	57	24	58	.07	0	.59	.02	1.09	.64	.01
Mandel (2001) Exp #2	36	64	37	59	-.19	0	.48	.15	1.02	.63	.00
									TOTAL of 3 predicted: 3 (100%)		
									OVERALL TOTAL of 88 predicted: 82 (93.2%)		

* Indicates results with Wald statistics over 1 degree of freedom
Note: Bolded rows indicate results where CDM outperforms BR