A Review and Prospect of Cumulative Prospect Theory Research

Eho-Cheng Lo, Chinese Culture University, Taiwan

The Asian Conference on Psychology & the Behavioral Sciences 2022
Official Conference Proceedings

Abstract
In view of the wide adoption and various research extensions of Cumulative Prospect Theory (CPT), this paper represents an attempt to perform a systematic review of articles that have employed CPT so as to explore its research trajectories and trends over time. A literature retrieval from Web of Science (WOS) yields a corpus of 495 articles in relation to CPT spanning over 2001-2020. The topic modeling method featuring Latent Dirichlet Allocation (LDA) is performed to produce topic trends and prospects concerning the corpus. For this purpose, we make use of the RStudio implementation of relevant packages for data preprocessing, modeling and visualization. The results are mainly categorized by dividing the articles into types of CPT exploration and parameter elicitation, the interplay and comparison between CPT and other theories and methods, and domain-specific applications by utilizing CPT to expound decision behavior. The conclusion drawn from the findings suggests that the potential active and new lines of CPT research in the future could be aimed more at route choice in transportation networks as well as decision making on the trade-off associated with issues of energy and environment.

Keywords**:** CPT, LDA, Topic Models, Literature Analysis, RStudio

1. Introduction

Uncertainty is an inherent part of decision making. In view of the likelihood and probability of an outcome, the ongoing quest to ensure decision-makers' behavior during the process of decision making has been a continuing research in the social sciences for more than 280 years (Bernoulli, 1738/1954). In the context of bounded rationality theory (Simon, 1957) and rank-dependent expected utility (Quiggin, 1982), Tversky and Kahneman (1992) introduced Cumulative Prospect Theory (CPT) as an alternative to other normative and descriptive models of decision making under uncertainty.

Since its debut, CPT has been one of the most favorable descriptive decision-making models (Zhou et al., 2017). At the time of writing, the original paper, "*Advances in Prospect Theory: Cumulative Representation of Uncertainty*" (Tversky and Kahneman, 1992), has been cited 15,683 times (computed by Google Scholar). Widely employed in various domains like behavioral economics, policy formulation, behavioral finance, transportation, and energy management, CPT has been a theoretical lens by researchers conducting empirical studies of choice under uncertainty and risk. However, there is limited attention paid to reveal the underlying research inclination and preference. Such expanding and heterogenous contributions to CPT make obtain a general overview and perspective of the hidden research topics embedded in CPT literature a complex task.

To address the aforementioned shortcomings, in this paper we adopt topic modeling method based on Latent Dirichlet Allocation (LDA) algorithm, implemented by LDAvis and topic model packages in R, to unveil how CPT-related studies had developed in the first two decades of the twenty-first century, and to have a picture of the prospect of CPT-related research.

2. Background

2.1. Cumulative Prospect Theory (CPT)

Relying on economic experiments, CPT was initiated to address decision making, either uncertain or risky, pertaining to any number of outcomes (Tversky and Kahneman, 1992). Instead of a normative model, CPT is deemed a descriptive model of decision making, which makes it close to or in agreement with true behavior. To reflect the nature of a descriptive model, CPT features a subjective value function and a subjective probability function (probability weight function, PWF). That is, a PWF reflects probabilistic distortion. A subjective value function $v(x)$ can be depicted in the form of a two-part power function (Tversky and Kahneman, 1992):

$$v(x) = \begin{cases} v^{+}(x) = x^{\alpha},\ x \geq 0;\ \alpha > 0, \\ v^{-}(x) = -\lambda(-x)^{\beta},\ x < 0;\ \beta > 0;\ \lambda \geq 1 \end{cases} \qquad (1)$$

where $v(x)$ is the subjective utility with respect to option x, α is the concavity of the value function for gains ($x \geq 0$), and β is the convexity of the value function for losses ($x < 0$). $0 < \alpha < 1$ and $0 < \beta < 1$ suggest diminished sensitivity for losses and gains. λ denotes a loss-aversion coefficient. $\lambda \geq 1$ indicates greater preference for gain than for the same loss, which reflects that the loss region of subjective utility is steeper than the gain region. As suggested by function (1), the pattern of a subjective value function $v(x)$ is shown in Fig. 1.

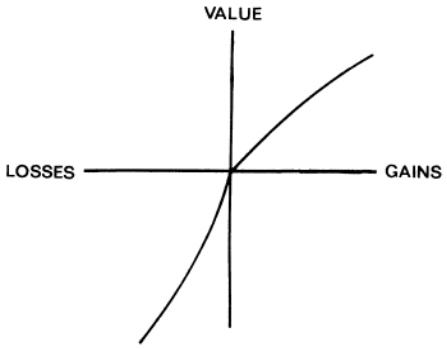


Figure 1: Shape of subjective value function (Kahneman, 1979).

The function form (1) tells that the reference point of gain and loss is 0. Let the reference point (determined by the decision maker) as x_0 ($x_0 \neq 0$), and the function (1) can be therefore equivalently written as below:

$$v(x) = \begin{cases} v^+(x) = (x - x_0)^\alpha, & x \geq 0;\ \alpha > 0, \\ v^-(x) = -\lambda(-x + x_0)^\beta, & x < 0;\ \beta > 0;\ \lambda \geq 1 \end{cases} \quad (2)$$

Similarly, a subjective probability function can also be described as a two-part power function as follows (Tversky and Kahneman, 1992):

$$\pi(p) = \begin{cases} \pi^+(p) = \dfrac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}}, & x \geq 0, \\ \pi^-(p) = \dfrac{p^\delta}{(p^\delta + (1-p)^\delta)^{1/\delta}}, & x < 0 \end{cases} \quad (3)$$

where $\pi^+(p)$ is the probability of subjective gains, $\pi^-(p)$ is the probability of subjective losses, p is the actual probability of gains and losses, γ and δ are the sensitivity of gains and losses, and $\gamma \leq 1$ and $\delta \leq 1$. Accordingly, the pattern of a subjective probability function $\pi(p)$ is shown in Fig. 2. Corresponding the so-called PWF, both function (3) and Fig. 2 show that moderate and high probabilities are under-weighted and low probabilities are over-weighted by decision makers.

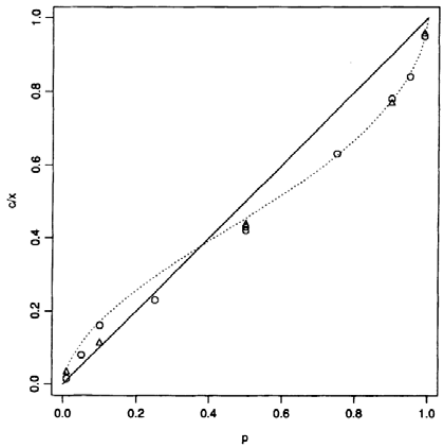


Figure 2: Shape (dotted curve when $\gamma \neq 1$and $\delta \neq 1$) of subjective probability function (Tversky and Kahneman, 1992).

By combining functions of (2) and (3), the prospect value of CPT can be described as the sum of the subjective gains and subjective losses as follows:

$$U = \sum (v^{+}(x) \cdot \pi^{+}(p)) + \sum (v^{-}(x) \cdot \pi^{-}(p)) \qquad (4)$$

According to functions (2), (3) and (4), it is obvious that CPT contains five parameters, and when $\gamma = 1$ and $\delta = 1$, we'll have the standard linear weighting. The function form of (4) indicates that one will pursue risks or avoid risks in the conditions of losses or gains respectively. It also suggests that one is more concerned with losses than with gains (Tversky and Kahneman, 1992).

As an alternative to normative models like Expected Utility Theory, CPT has garnered a great deal of support, applications and extensions. In the field of transportation, Schwanen and Ettema (2009), Gao et al. (2010), Li and Hensher (2011), Chow et al. (2010), and Zhang et al. (2018) used CPT to study route choice in traffic networks. Breuer and Perst (2007) employed CPT to analyze discount reverse convertibles and reverse convertible bonds. In view of portfolio optimization, Omane-Adjepong et al. (2019) adopted CPT to classify and select cryptocurrencies. Félix et al. (2019) made use of CPT to explain the overpricing of out-of-the money single stock calls. In the domain of energy management, researchers resorted to CPT to study the determination of the optimal photovoltaic/battery energy storage/electric vehicle charging stations portfolio (Liu and Dai, 2020), the site selection of photovoltaic power plants (Liu et al., 2017), and the capacity credit of wind power simulations for various wind time series interval lengths (Wilton el al., 2014). The trace of CPT relevant research can also be found in bidding decision on land auction (Peng and Liu, 2015), government purchase of home-based elderly-care services (Lu et al., 2020), the influence of emotions specific to risk on flood insurance demand (Robinson and Botzen, 2020), the selection of new product development concept (Wang et al., 2018), and addressing the risk decision-making problem in emergency response (Liu et al., 2014).

2.2. Literature analysis

On a given subject, the analysis of literature review allows research trends as well as potential gaps leading to new studies and findings to be uncovered (Levy and Ellis, 2006). Literature analysis is considered as an imperative basis and approach to reveal appreciation on new research themes. This connection is manifested by literature reviews of various publications regarding numerous sciences (Cronin et al., 2008; Jesson and Lacey, 2006).

Without the help of new information and communications technology (ICT), large amount of time and efforts are mandatory to conduct thorough and comprehensive literature analysis in pursuit of new research topics of a given subject. With the facilitation and advancement of ICT and without the limitation on time and locations, these days scholars and researchers of different academic disciplines are able to access multiple online libraries and databases so as to retrieve a pile of articles on a given research subject. Nevertheless, in order to extract useful insights and knowledge from high volumes of papers retrieved, digesting the contents as well as grasping the contexts require determination and great efforts. To deal with such a challenging task, the measures of text mining (TM) based on ICT are introduced and employed to investigate the literature and to uncover trendy studies over journals and time. In order to generate a body of knowledge of the literature in question, associated terms and vocabularies (a sequence of "n" words in the name of n-gram) need to be distilled and

classified from large texts (Delen and Crossland, 2008). TM is particularly applied to examine unstructured or semi-structured datasets like text documents (Fan et al., 2006). To reveal the academic research trends, a number of studies have adopted TM techniques to investigate papers in a variety of journal databases. Lee et al. (2010), Hung (2010), Sharma et al. (2018), Zhai et al. (2015), and Kim (2016) applied TM to research trends in the fields of information science (digital library), education (e-learning), machine learning, biomedicine, and medical informatics respectively.

As a specific type of algorithms applied to TM, by taking the number and distribution of terms into account to model a specific number of different topics from unstructured datasets, a method called Latent Dirichlet Allocation (LDA) was proposed by Blei et al. (2003). Such a manner can assist researchers to recognize topics related to the gap for prospect studies (Moro et al., 2015).

2.3. Topic Models and Latent Dirichlet Allocation (LDA)

In the setting of TM techniques, topic models are a sort of unsupervised statistical machine learning methods. The purpose of topic models and associated analytics is to summarize the topics from a corpus in a way of reduced human resources. The so-called topics are clusters designated by grouping the words, which are unknown beforehand. For unstructured datasets without reference to known outcomes, unsupervised machine learning is applied to infer underlying structure, resemblances and distinct patterns of data and therefore to make sense of data. Topic models came into being through the research on searching, indexing and clustering voluminous unlabeled and unstructured documents (Sun et al., 2017). Various applications, like social networks, images, genetic research (Blei, 2012), opinion classification, sentiment discovery, trend detection, and big data research (Shivashankar et al., 2011; Hu et al., 2014), communication similarity in political movements (Stier et al., 2017), gauging framing and meaning nuances in cultural sociology (DiMaggio et al., 2013), and contemporary art discourse (Roose et al., 2018) had employed topic models to reveal the main themes and patterns residing in a huge amount of data in those fields.

In the context of social sciences, LDA is the foremost and the most adopted variant of topic modeling methods (Zhao et. al, 2014; Saari 2019; Pääkkönen and Ylikoski, 2020). As a specific algorithm of topic modelling applicable for massive collections of documents (Blei, 2012), without training data, prior labeling and annotations, LDA can be utilized to investigate thousands or millions of documents where human annotation is impossible (Blei, 2012; Sun et al., 2017). LDA captures the perception of documents bearing multiple topics, that is, distinct topics “latently” existing in documents show in different proportions. In other words, each document can be regarded as a mixture of “latent” topics which expound common occurrence of words in documents. Likewise, each topic is treated as a mixture of words, which suggests a combination of ideas with a certain meaning inside the corpus. For instance, the words “decision, risk, uncertainty” and “text, word, document, corpus” usually appear together in CPT-related and LDA-related studies respectively. This means that a topic is a group of words that frequently appear in the documents of a corpus. Similarly, those clusters of words also have higher probabilities (weight assigned) in a topic, and those words can also have higher probability in some topics. As a statistical model, LDA therefore represents such a perception by an imaginary random (generative) process that produce documents. To be specific, it is assumed that topics are prescribed before associated documents and data. Here, a “Dirichlet distribution” reflects the distribution of those

prescribed topics, and it is applied to designate the words in a document with respect to different topics (Blei, 2012).

In essence, as a type of generative probabilistic models for a corpus, LDA is a Bayesian hierarchical modeling transcribed in three levels (Blei et al., 2003). It is suggested that topics and word mixtures are represented by Dirichlet distributions, which generate documents accordingly. As illustrated by Blei (2012), Fig. 3 represents the probabilistic graphical model for LDA algorithm. Random variables are denoted by a node. Unshaded and shaded (grey) nodes depict hidden (latent, i.e., existing but neither known nor seen directly) and observed random variables separately in the order given. Rectangular plates indicate the replication of variables. Plate K, M and N are the number of topics, document and word respectively. In other words, the rectangle M denotes the documents, the corpus, we are going to investigate. The rectangle N indicate the word positions within a certain document. As aforementioned, Fig. 3 also expresses that LDA features a three-level Bayesian hierarchy. Hidden random variable α and β , as input parameters, denote the topic distribution of each document and word distribution of each topic respectively. The observed random variable W, as the output, are the words that one can see and read. That is, given the words W, the conditional (posterior) distribution of the rest hidden variables are performed. The determination of each topic's word distribution φ in K is derived from the input β. In a similar fashion, each document's topic distribution θ in M is generated from the input α. The random variable Z refers to the topic assignment for each specific term in W, which is developed from θ. Different topics are described by the terms output from LDA. In a word, LDA assumes topics, which can be revealed by analytical measures, of a corpus exist in a latent space.

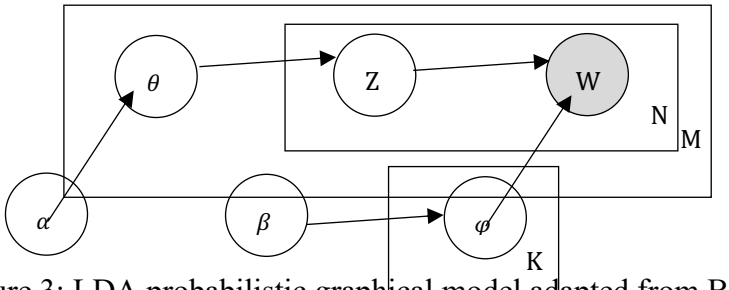


Figure 3: LDA probabilistic graphical model adapted from Blei (2012)

Equivalent to Fig. 3, as explained by Blei (2012), the probabilistic graphical model for LDA can be represented as equation (5):

$$p(\theta, z, w|\alpha, \beta) = p(\theta|\alpha)\prod_{n=1}^{N} p(z_n|\theta)p(w_n|z_n, \beta) \qquad (5)$$

During the past decade, LDA has been wildly applied in various academic disciplines. Fields like social network analysis (Weng et al., 2010), politics (Grimmer and Stewart, 2013), journalism (Rusch et al., 2013), cultural sociology (Mohr and Bogdanov, 2013), business intelligence (Moro et al., 2015), communication research (Maier et al., 2018) have employed LDA to the studies on research trends.

3. Methodolog

3.1. Literature acquisition

In this study, only “Web of Science (WOS)” was accessed as our literature (including journal articles and conference proceedings) source. The terms “cumulative prospect theory” were selected as our only query. With quotation marks to make the terms an exact phrase, the search in question was input by *Topic* inquiry of the *Basic Search*. As per the input, WOS sought the fields of literature title, abstract and author keywords. Published literature between the years 2001 and 2020 was examined. The distribution over the designated time is shown in Fig 4. This study only examined the title, abstract and author key words for our purpose. At the time of writing, the corpus composed of 495 research papers were retrieved.

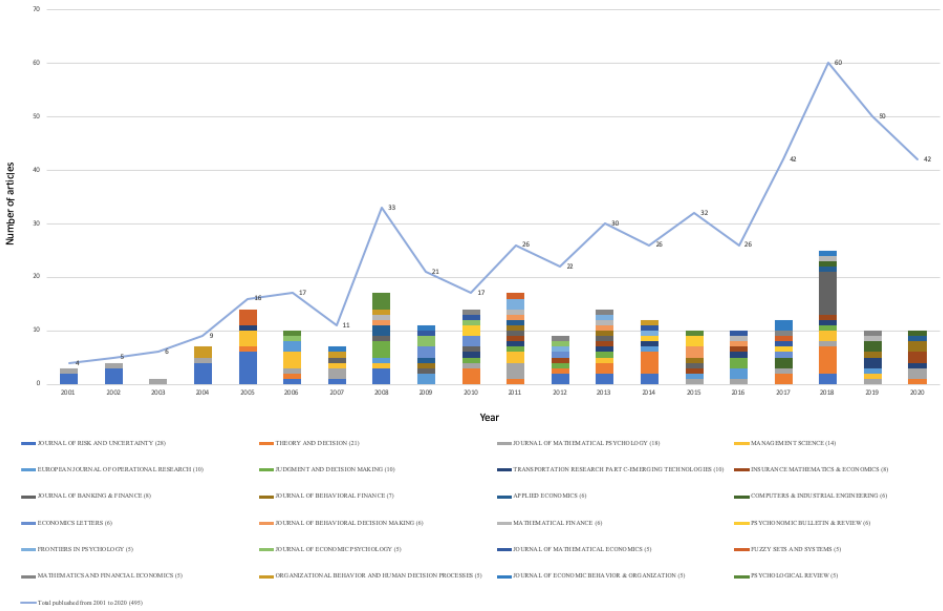


Figure 4: Distribution of number of articles over year (n=495).

In Fig. 4, the line graph shows the number of articles published by year. The corpus of 495 documents is contributed by 214 types of journals and conference proceedings. In terms of the release frequency over that period, the top three journals are the “Journal of Risk and Uncertainty (n=28; 5.66%),” “Theory and Decision (n=21; 4.24%),” and “Journal of Mathematical Psychology (n=18; 3.64%).” Though fluctuations begin in 2007 and thereafter, the number of circulated articles, peaking at 60 in 2018, increases in general in the timeframe. The bar chart in Fig. 4 illustrates 24 journals (11.21% of the journal source) that at least have five CPT-related articles over the period of 20 years, which accounts for 217 documents (43.84% of the corpus).

In view of the color varieties of the bar chart as well as the distances (gaps) between the apogees of the bars and corresponding points in line graph in Fig.4, it suggests that the source of the articles had been getting more diversified with the passage of time. As shown in Fig. 5, the trend of the portions (ranging between 87.5% in 2005 and 20.00% in 2019) attributed by these 24 journals had gone down as a whole. It also tells the concentration and dominance of journal sources had been towards lower. Fig. 4 and Fig. 5 both suggest that the spillover effects of CPT on other academic disciplines.

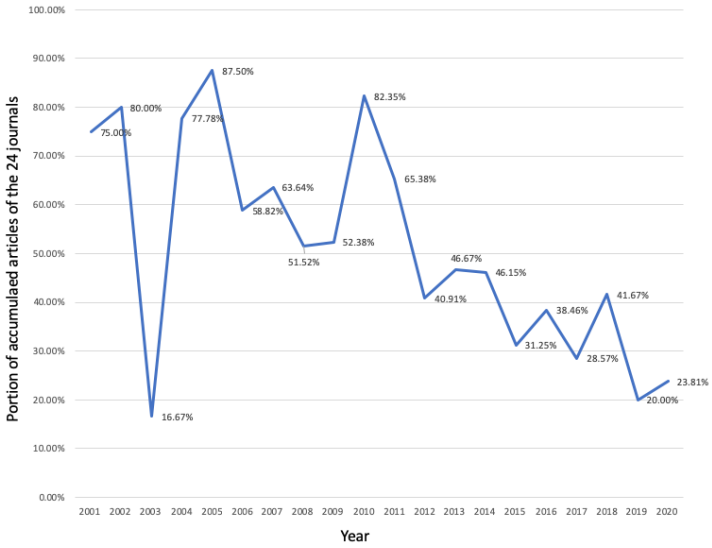


Figure 5: Dominance of the 24 journals (taken as a whole) in the corpus over year.

3.2 Data preprocessing

In this step, RStudio (version 4.0.3) was installed and run as the workspace. Commands first tokenized retrieved texts by splitting them into sentences, decomposing sentences into words, lowercasing the words, and removing punctuation. Next, stop words (e.g., “the”, “and”, “or”, “for”, etc.), numbers, non-alpha numeric characters were all removed. Finally, excutions lemmatized and stemmed various forms of a word to one single form or its root form (e.g., change "makes", "making", or "made" to the lemma "make"). The above tasks were performed by the “tm” package (version 0.7-8), a framework for text mining (Ingo and Kurt, 2008; Ingo et al., 2008), within RStudio.

For the scalar value of Diriclet distribution hyperparameter, 0.02 was set for the distributions over the vocabulary and topics. Gibbs’s sampling number was set to 5,000 for scanning the corpus. Fig. 6 demonstrates the data preprocessing executed by commands of “tm” package. As an example, Fig. 6 (a) and Fig. 6 (b), snapshotted from the raw data in CSV file and the console of RStudio, are the original text and the text after preprocessing respectively.

Performance on an intuitive symbolic number skills task-namely the number line estimation task-has previously been found to predict value function curvature in decision making under risk, using a cumulative prospect theory (CPT) model. However there has been no evidence of a similar relationship with the probability weighting function. This is surprising given that both number line estimation and probability weighting can be construed as involving proportion judgment, that is, involving estimating a number on a bounded scale based on its proportional relationship to the whole. In the present work, we re-evaluated the relationship between number line estimation and probability weighting through the lens of proportion judgment. Using a CPT model with a two-parameter probability weighting function, we found a double dissociation: number line estimation bias predicted probability weighting curvature while performance on a different number skills task, number comparison, predicted probability weighting elevation. Interestingly, while degree of bias was correlated across tasks, the direction of bias was not. The findings provide support for proportion judgment as a plausible account of the shape of the probability weighting function, and suggest directions for future work.	22 performance intuitive symbolic number skills tasknamely number line estimation taskhas previously found predict value function curvature decision making risk using cumulative prospect theory cpt model however evidence similar relationship probability weighting function surprising given number line estimation probability weighting can construed involving proportion judgment involving estimating number bounded scale based proportional relationship whole present work reevaluated relationship number line estimation probability weighting lens proportion judgment using cpt model twoparameter probability weighting function found double dissociation number line estimation bias predicted probability weighting curvature performance different number skills task number comparison predicted probability weighting elevation interestingly degree bias correlated across tasks direction bias findings provide support proportion judgment plausible account shape probability weighting function suggest directions future work
(a)	(b)

Figure 6: Text (research abstract) comparison before (a) and after (b) data preprocessing (entry 22 in the corpus as an example).

3.3 Topic modelling by LDA

In this research, due to a lack of knowledge about the trends of analyzed CPT-associated body of literature, the topic modeling features LDA is chosen. The R package of "topicmodels" (version 0.2-11) was selected and performed to carry out latent topic extraction (Hornik and Grün, 2011). The approach (metric) proposed by Arun et al. (2010) was employed to identify the number of topics of the corpus in question. At least 50 meaningful words were grouped to represent a topic, and the top 50 topics were visualized in the analysis radar.

3.4 Topic visualization

Names and interpretation were assigned to each extracted LDA topics. We utilized the visualization tool LDAvis (version 0.3.2) to generate a 2D topic map with axes based on interpreted topic grouping (Sievert and Shirley, 2014). LDAvis processes multidimensional scale analysis to bring about the axes of a bi-dimensional space. It depicts each topic by a circle distributed in the above-mentioned space. The area of a cycle denotes the relevance of the corresponding topic to the corpus. The distance between the centers of circles stands for topic similarity (the more similar the shorter).

4. Results

By deleting repetitions, 1,451 unique terms were obtained from the corpus. Terms not greater than 5-time repetitions were ignored. Accordingly, Fig. 7 demonstrates the 30 most frequently appearing topics of the corpus.

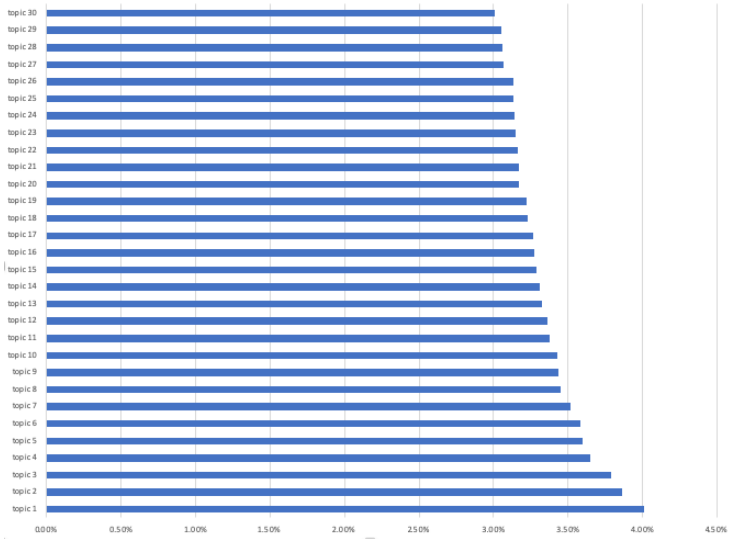
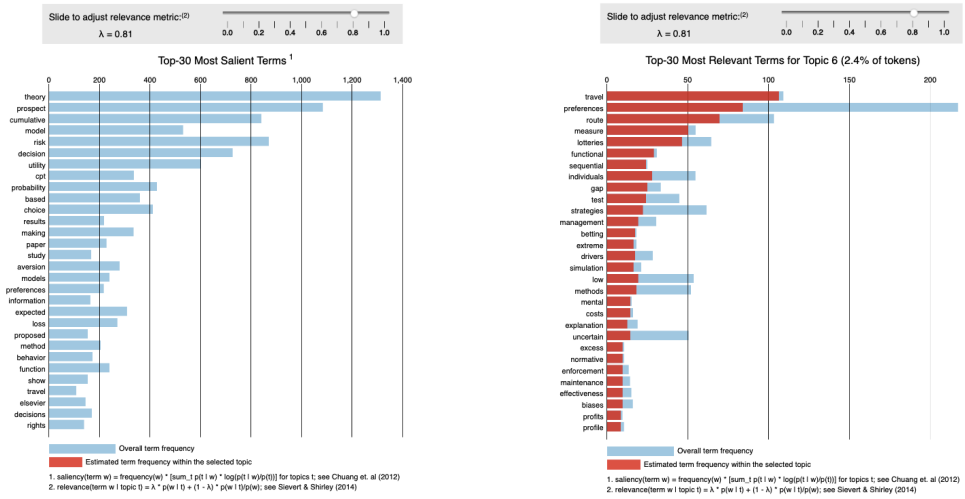


Figure 7: The distribution of top-30 most frequent topics of the corpus.

Given the value computed by the metric proposed by Arun et al. (2010), the inferred optimal number of topics is K=50. Gibbs's was applied for sampling to render the associated parameters and inference (Lynch, 2011). By LDAvis, the distributions of the corpus-wide and topic-specific terms are shown in Fig. 8.



(a) (b)

Figure 8: Distribution of the corpus-wide and topic-specific terms by LDA inference.

According to Fig. 8 (a), the top-30 most noticeable terms in the analyzed corpus are shown. In consonant with intuition, the terms that constitute "cumulative prospective theory" top the horizonal bar chart. The term "theory" ranking above the rest suggests that "theory" not only refers to CPT but also relates to other theories like bounded rationality theory, rank-dependent expected utility and expected utility theory, when it comes to behavioral decision-making research. The second highest term "prospect" can refer to CPT as well as its predecessor PT; additionally, it can be solely extracted to explain the meaning and concept of prospect itself. Noticeably, "risk" comes in third. It suggests that, aside from CPT itself tacking risky decision making, the majority of the studies deal with the contexts in which the probability of an uncertain outcome is well known. The acronym CPT, the initial letters of *c*umulative *p*rospect *t*heory, also ranks high. The rest terms, ranking above number ten, are those words commonly accompanied with CPT.

In Fig. 8 (b), the horizontal bars visualize the terms most highly associated with a certain topic (here, topic 6 as an example). The overlaid bar of a specific term indicates the topic-specific (red shaded) and corpus-wide frequencies. By adjusting the slider at the upper-right of Fig. 8 (b), to increase the lambda (λ) parameter will decrease the weight of the ratio of the word frequency of a given topic to the word overall frequency in the corpus. That is, important words for the given topic move downward.

The visualized relation (similarity) between topics and the prevalence of each topic are illustrated in Fig. 9. Each circle represented a topic, and the circle area tells the prevalence of the corresponding topic. The Euclidean distance between the centers of the circles tells the similarity (semantic relationship) between the topics. At an aggregate level, neighboring or overlapped circles are clustered to infer associated research themes denoted from C_1 to C_8. Isolated circles are denoted from T_1 to T_8.

The most favored topics from the analyzed corpus about CPT were C_1 comprising topic 1, 2 and 5 (in terms of prevalence, from high to low). C_1 is research about risky/stochastic decision, which is the exploration, extensions and arguments pertaining to CPT. C_2, made by topic 3 and 4, is relevant to the comparison and contract between CPT and other models. C_3 empirically address the parameters of PWF through experiments, that is, to discover the values of the abovementioned γ and δ in function (3). Through the lens of psychology, C_4

is the studies on decision making in the given-and-take context of environment protection. C_5 sheds light on the alignment and/or discrepancy of CPT with respect to experiments in certain conditions and domains. C_6 specializes in investment behavior of financial sector. C_7 is relevant to the integration of CPT and TODIM (TOmada de Decisao Interativa Multicriterio) for certain research purposes. Similar to C_7, C_8 is about the handshake between CPT and MCDM (Multi-Criteria Decision Making) method to realize various research attempts.

To generalize the topics of C_1 to C_8, C_1, C_2 and C_3 primarily examine the theory, CPT, itself. Obviously, the circle areas in C_1, C_2 and C_3 are self-explanatory per the criteria of literature selection set previously. C_4 and C_6 address CPT's application to certain research themes. C_5 focuses on the alignment with or disagreement with the arguments of CPT. C_7 and C_8 let CPT accommodate other methods in order to achieve particular research goals.

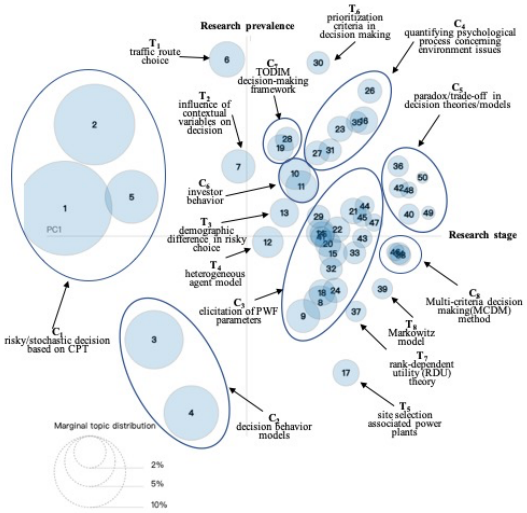


Figure 9: Visualization of the 50 most prevalent topics (rendered by LDAvis) and clustering.

For T_1 and T_5, one focuses on the traffic route choice and the other concentrate on the power plant site selection. Regarding T_2, T_3, T_4 and T_6, they attempt to study CPT from the perspectives of contextual variables, demography, agents, and criteria priority respectively. Finally, as clusters of C_7 and C_8, T_7 and T_8 are associated with taking other methods and theories, Rank-Dependent Utility theory and Markowitz model, into consideration.

5. Discussion and Conclusion

Research on CPT has made progress on its adoption and applicability over the last 20 years. Nevertheless, there is a lack of review studies that examine the development, trends, patterns and findings. In contributing to this gap in the literature, this research provides literature collection criteria, LDA-enabled topic modeling methods and practice. In this paper, it presents a holistic view of the current state, at the time of this writing, of CPT-related research by presenting the results of a systematic review of 495 articles put forward since 2001. This study provides an alternative way other than the traditional statistical analysis approach to discover research past and prospect trajectories in the field of behavioral economics, specifically those relevant to CPT.

With the implementation of topic modeling characterized by LDA as well as TM, data preprocessing and visualization packages in RStudio, this study reveals the implicit preferences and trends of CPT research. The papers from WOS are investigated in the first

two decades of the twenty-first century. Through this study, 50 topics, portrayed by some 30 words and grouped into 8 clusters (by overlapped and neighboring circles) and 8 isolated circles, are delivered and concluded in a 2D space. The horizontal and vertical axes delineating the 2D space represent popularity and research stage respectively in the corpus. Through this spatial representation and analysis, this study draws a generalized conclusion, in terms of CPT's applications, on route choice in transportation networks as well as decision making on the trade-off associated with issues of energy and environment that could be the coming foci of CPT in exploring decision behavior. In addition, to incorporate other theories and methods into CPT's terrain could be an inviting research approach to explore.

As an attempt to apply LDA to generate topic trends and prospects of CPT associated research, the proposed method bears several limitations. Those limitations arise from the statistical assumptions inhere in LDA. One assumption is that the order of the documents in the collection is not considered. This impedes the analysis of the track and change of a certain underlying theme of the corpus over time. Second, LDA relies on the premise that the words are exchangeable in the document. It is also known as the assumption of the "bag of words." In other words, this assumption does not take how the topics conditionally generate words on the previous word. However, this assumption has minor impact on the quality of this study since we only focus on the course semantic structure of the texts. The third limitation is that the number of topics is fixed and manifest assumed by LDA. It means that the number of topics is given rather than determined by posterior inference of the document collected. Additionally, the hierarchies of topics in question cannot be inferred accordingly.

The results are useful for academic stakeholders to formulate their prospect research lines and concentrations. For stakeholders in academia, research institutes, and governments, our findings also facilitate the decision on funding CPT-based research and applications. Two aspects of direct extension of this study include: On the one hand, the employment of derivative models of LDA that relax LDA assumptions previously mentioned. That is, to relax aforementioned assumptions of LDA could be the following directions for research in CPT. By doing so, future work may produce a more realistic posterior topical structure where a topic could be a sequence of distributions over words in place of a single distribution over words. It will therefore more accurately reflect the topical trace of collections spanning over years. On the other hand, the original paper on CPT by Tversky and Kahneman (1992) has been cited more than 15,000 times, and it suggests that to include more research articles, qualified and filtered by certain research purposes, into the analyzed corpus can be exploited. To combine both aspects with suitable analysis and visualization packages, inferences from and insights into CPT research can be explored and expected.

References

Arun, R., Suresh, V., Madhavan, C., & Murty, M. (2010). *On finding the natural number of topics with latent dirichlet allocation: Some observations*

Bernoulli, D. (1738/1954) Exposition of a New Theory on the Measurement of Risk. *Econometrica*, 22, 23-36. http://dx.doi.org/10.2307/1909829

Blei, D. M. (2012). Probabilistic topic models. *Commun.ACM, 55*(4), 77–84.

Blei, D., Ng, A., & Jordan, M. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*, 993-1022.

Breuer, W., & Perst, A. (2007). Retail banking and behavioral financial engineering: The case of structured products. *Journal of Banking & Finance, 31*(3), 827-844. doi:https://doi.org/10.1016/j.jbankfin.2006.06.011

Chow, J. Y. J., Lee, G., & Yang, I. (2010). Genetic algorithm to estimate cumulative prospect theory parameters for selection of high-occupancy-vehicle lane. *Transportation Research Record, 2157*(1), 71-77.

Cronin, P., Ryan, F., & Coughlan, M. (2008). Undertaking a literature review: A step-by-step approach. *British Journal of Nursing (Mark Allen Publishing), 17*, 38-43.

Delen, D., & Crossland, M. D. (2008). Seeding the survey and analysis of research literature with text mining. *Expert Systems with Applications, 34*(3), 1707-1720. doi:https://doi.org/10.1016/j.eswa.2007.01.035

DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics, 41*(6), 570-606. doi:https://doi.org/10.1016/j.poetic.2013.08.004

Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2006). Tapping the power of text mining. *Communications of the ACM, 49*, 76-82.

Félix, L., Kräussl, R., & Stork, P. (2019). Single stock call options as lottery tickets: Overpricing and investor sentiment.*20*(4), 385-407.

Gao, S., Frejinger, E., & Ben-Akiva, M. (2010). Adaptive route choices in risky traffic networks: A prospect theory approach. *Transportation Research Part C: Emerging Technologies, 18*(5), 727-740. doi:https://doi.org/10.1016/j.trc.2009.08.001

Grimmer, J., & Stewart, B. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis, 21*, 267-297.

Hornik, K., & Grün, B. (2011). Topicmodels: An R package for fitting topic models. *Journal of Statistical Software, 40*

Hu, Z., Fang, S., & Liang, T. (2014). Empirical study of constructing a knowledge organization system of patent documents using topic modeling. *Scientometrics, 100*(3), 787-799.

Hung, J. (2010). Trends of E-learning research from 2000 to 2008: Use of text mining and bibliometrics. *British Journal of Educational Technology, 43*, 5-16.

Ingo, F., and Kurt, H. (2008). *m: Text Mining Package*. R package version 0.7-8, https://CRAN.R-project.org/package=tm.

Ingo, F., Kurt, H., & David, M. (2008). Text mining infrastructure in R. *Journal of Statistical Software, 25*

Jesson, J., & Lacey, F. (2006). How to do (or not to do) a critical literature review. *Pharmacy Education, 6*

Kahneman, D. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263.

Kim, Y. (2016). Medical informatics research trend analysis: A text mining approach. *Health Informatics Journal, 24*

Lee, J., Kim, H., & Pan Jun, K. (2010). Domain analysis with text mining: Analysis of digital library research trends using profiling methods. *J.Information Science, 36*, 144-161.

Levy, Y., & Ellis, T. J. (2006). A systems approach to conduct an effective literature review in support of information systems research. *Informing Science: The International Journal of an Emerging Transdiscipline, 9*, 181+.

Li, Z., & Hensher, D. (2011). Prospect theoretic contributions in understanding traveller behaviour: A review and some comments.*31*(1), 97-115.

Liu, J., Xu, F., & Lin, S. (2017). Site selection of photovoltaic power plants in a value chain based on grey cumulative prospect theory for sustainability: A case study in northwest china. *Journal of Cleaner Production, 148*, 386-397.

Liu, Y., Fan, Z., & Zhang, Y. (2014). Risk decision analysis in emergency response: A method based on cumulative prospect theory. *Computers & Operations Research, 42*, 75-82. doi:https://doi.org/10.1016/j.cor.2012.08.008

Liu, Z., & Dai. (2020). Portfolio optimization of Photovoltaic/Battery energy Storage/Electric vehicle charging stations with sustainability perspective based on cumulative prospect theory and MOPSO. *Sustainability, 12*, 985.

Lu, J., He, T., Wei, G., Wu, J., & Wei, C. (2020). Cumulative prospect theory: Performance evaluation of government purchases of home-based elderly-care services using the pythagorean 2-tuple linguistic TODIM method. *International Journal of Environmental Research and Public Health, 17*(6), 1939.

Lynch, S. (2011). *Introduction to applied bayesian statistics and estimation for social scientists*

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., et al. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures, 12*, 1-26.

Mohr, J. W., & Bogdanov, P. (2013). Introduction—Topic models: What they are and why they matter. *Poetics, 41*(6), 545-569. doi:https://doi.org/10.1016/j.poetic.2013.10.001

Moro, S., Cortez, P., & Rita, P. (2015). Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent dirichlet allocation. *Expert Systems with Applications, 42*(3), 1314-1324. doi:https://doi.org/10.1016/j.eswa.2014.09.024

Omane-Adjepong, M., Ababio, K. A., & Alagidede, I. P. (2019). Time-frequency analysis of behaviourally classified financial asset markets. *Research in International Business and Finance, 50*, 54-69. doi:https://doi.org/10.1016/j.ribaf.2019.04.012

Pääkkönen, J., & Ylikoski, P. (2020). Humanistic interpretation and machine learning. *Synthese,* , 1-37.

Peng, Y., & Liu, X. (2015). Bidding decision in land auction using prospect theory.*19*(2), 186-205.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization, 3*(4), 323-343. doi:https://doi.org/10.1016/0167-2681(82)90008-7

Robinson, P. J., & Botzen, W. J. W. (2020). Flood insurance demand and probability weighting: The influences of regret, worry, locus of control and the threshold of concern heuristic. *Water Resources and Economics, 30*, 100144.

Roose, H., Roose, W., & Daenekindt, S. (2018). Trends in contemporary art discourse: Using topic models to analyze 25 years of professional art criticism. *Cultural Sociology, 12*(3), 303-324.

Rusch, T., Hofmarcher, P., Hatzinger, R., & Hornik, K. (2013). Model trees with topic model preprocessing: An approach for data journalism illustrated with the WikiLeaks afghanistan war logs. *The Annals of Applied Statistics, 7*

Saari, E. (2019). Trend analysis in AI research over time using NLP techniques. Tampere University.

Schwanen, T., & Ettema, D. (2009). Coping with unreliable transportation when collecting children: Examining parents' behavior with cumulative prospect theory. *Transportation Research Part A: Policy and Practice, 43*(5), 511-525. doi:https://doi.org/10.1016/j.tra.2009.01.002

Sharma, D., Kumar, B., & Chand, S. (2018). *Trend analysis in machine learning research using text mining*

Shivashankar, S., Srivathsan, S., Ravindran, B., & Tendulkar, A. V. (2011). Multi-view methods for protein structure comparison using latent dirichlet allocation. *Bioinformatics, 27*(13), i61-i68.

Sievert, C., & Shirley, K. (2014). *LDAvis: A method for visualizing and interpreting topics*

Simon, H. A. (1957). *Models of man; social and rational.* Wiley.

Stier, S., Posch, L., Bleier, A., & Strohmaier, M. (2017). When populists become popular: Comparing facebook use by the right-wing movement pegida and german political parties. *Information, Communication & Society, 20*, 1365-1388.

Sun, X., Liu, X., Bin, L., Li, B., Lo, D., & Liao, L. (2017). Clustering classes in packages for program comprehension. *Scientific Programming, 2017*, 1-15.

TVERSKY, A., & KAHNEMAN, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297-323.

Wang, Y., Li, Y., Ying, C., & Chin, K. (2018). A new product development concept selection approach based on cumulative prospect theory and hybrid-information MADM. *IOP Conference Series: Materials Science and Engineering, 381*, 012133.

Weng, J., Lim, E., Jiang, J., & qi, Z. (2010). *Twitterrank: Finding topic-sensitive influential twitterers*

Wilton, E., Delarue, E., D'haeseleer, W., & van Sark, W. (2014). Reconsidering the capacity credit of wind power: Application of cumulative prospect theory. *Renewable Energy, 68*, 752-760.

Zhai, X., Li, Z., Gao, K., Huang, Y., Lin, L., & Wang, L. (2015). Research status and trend analysis of global biomedical text mining studies in recent 10 years. *Scientometrics, 105*

Zhang, C., Liu, T., Huang, H., & Chen, J. (2018). A cumulative prospect theory approach to commuters' day-to-day route-choice modeling with friends' travel information. *Transportation Research Part C: Emerging Technologies, 86*, 527-548. doi:https://doi.org/10.1016/j.trc.2017.12.005

Zhao, W., Zou, W., & Chen, J. (2014). Topic modeling for cluster analysis of large biological and medical datasets. *BMC Bioinformatics, 15 Suppl 11*, S11.

Zhou, F., Lei, B., Liu, Y., & Jiao, R. J. (2017). Affective parameter shaping in user experience prospect evaluation based on hierarchical bayesian estimation. *Expert Systems with Applications, 78*, 1-15. doi:https://doi.org/10.1016/j.eswa.2017.02.003