Analysis of Main Influencing Factors of Scientific Data Sharing in Universities

Hongwei Yuan[1,2]*, Yan Liu[1], Song Huang[1]

[1]Library, Nanjing University of Aeronautics and Astronautics, Nanjing, China
[2]Evaluation Center for Industrial and Informatization Think Tanks, Nanjing University of Aeronautics and Astronautics, Nanjing, China
Email: *yhwlib@nuaa.edu.cn

Abstract

At present, China is in the critical period of implementing innovation-driven development strategy and building a strong country in science and technology. Strengthening and standardizing scientific data management is an important means to strengthen China's scientific and technological innovation capacity building. As an important base for scientific data generation, what are the attitudes, cognitions, behavioral characteristics and influencing factors of researchers on scientific data sharing? Based on the theory of technology acceptance model and planning behavior, this paper constructs the influencing factors model of scientific data sharing willingness in universities, discusses the factors affecting the scientific data sharing of universities, and draws three main factors affecting the scientific data sharing of universities: individual characteristics, subjective norms, systems characteristics. Finally, from the aspects of service content, incentive mechanism, literacy education, platform architecture, etc., scientific data sharing services of university libraries are carried out.

Subject Areas

Library, Intelligence and Philology

Keywords

Scientific Data Sharing, Willingness to Share, University Researchers, Influencing Factors Model, Countermeasures

1. Introduction

Scientific data can also be called research data. It refers to the original basic data

that reflect the nature, characteristics, and changing laws of the objective world obtained through scientific and technological activities (as experiment, observation, detection, investigation, etc.) or through other means. And various data are set for systematic processing according to the needs of different scientific and technological activities. According to the life cycle theory, whether it is the early, middle and late stage of research activities, scientific data are an indispensable and important resource for researchers to complete their results. It can ensure their experimental repeatability. One of the basic methods for effective development and rational use of scientific data is the sharing of scientific data. This kind of breakthrough sharing behavior makes the scientific resources continue to be recycled, breaking the phenomenon that individual researchers monopolize information resources, and also improves the optimal allocation of digital resources. It has an important and far-reaching impact on improving the level of global science and technology and promoting the sustainable development of the economy and society.

At the beginning of the 21st century, Microsoft Research Institute proposed that big data should be used as an important tool for academic activities and communication in the future. The analysis and processing of a large amount of accessible data received more attention, and the research on scientific data sharing was more specific. In this context, government agencies in many countries and international organizations have shown an emphasis on scientific data sharing. For example, the International Council of Scientific Unions has established the Committee on Data for Science and Technology (CODATA) and the World Data Center (WDC). UNESCO, the International Council of Scientific Unions and the Organization for Economic Co-operation and Development (OECD) funded the Global Science Information Sharing Initiative (GICSI). Government agencies in India, Japan and Australia which have established a shared strategic development plan at the national level and promulgated regulations to promote scientific data sharing.

Compared with the rapid development of scientific data sharing in other countries, China has been gradually developing scientific data sharing projects in the field. From the National Science and Technology Basic Condition Platform to the Regional Data Center, we have continuously implemented important strategic deployments such as China Mid-and Long-term Science and Technology Planning and The Twelfth Five-Year Science and Technology Development Plan. The introduction of these policy outlines means that China has made scientific data sharing a determination as an important development strategy. Universities are the cradle of the motherland's talents. It is a fertile ground for academic development. It is the source of the birth of scientific data. It conducts investigations and studies on the main factors of scientific data sharing in colleges and universities. It can deeply understand the more realistic ideas and scruples of researchers, and find the key factors affecting their willingness to share. Promote the process of scientific data sharing projects.

2. Theoretical Basis

2.1. Theory of Planned Behavior

In the late 1980s, the famous American social psychologist Ajzen [1] proposed the theory of planned behavior (TPB). The theory is based on the theory of reasoned action (TRA) proposed by Ajzen and Fishbein. The theory emphasizes that individual attitudes and subjective norms are two important factors that influence the individual's willingness to act. However, after constant research and practice, Ajzen found that personal behavior is controlled without being influenced by one's own will. Therefore, he extended the theory of reasoned action and added the variable factor of "perceived behavior control" to put forward the theory and proposed the theory of planned behavior. Subsequent research further describes the relationship between attitudes, subjective norms, and perceived behavioral control. Make the theory of planned behavior gradually mature and perfect. The research focus of this paper is the influence of attitude as individual characteristics, subjective norm and perceived behavior control as the impact of risk factors on the willingness of scientific data sharing in universities.

2.2. Technology Acceptance Model

The Technology Acceptance Model (TAM) was proposed by Davis [2] in 1989. It uses the theory of reasoned action to study the degree of user acceptance of information systems. The model proposes that the usefulness and ease of use are the two main factors that determine the degree of user acceptance. The usefulness of perception reflects the extent to which individuals believe that a particular system is being used to improve their performance. The ease of use of perception reflects how easily people think a particular system is being used. The technology acceptance model assumes that the use of the system is determined by the willingness. The willingness to act depends on the usefulness of the attitude and perception of use, and the attitude of use depends on usefulness and ease of use. The focus of this study is to explore the usefulness and ease of use, that is, the characteristics of the system, and to understand its impact on the willingness to share scientific data in universities.

2.3. Social Cognitive Theory

Social cognition theory is an educational theory advocated by the famous American psychologist Albert Bandura [3] in the late 1970s. Cognitive factors are added to the traditional behavioral personality theory, forming the concept of social cognitive theory. The theory treats human thinking and behavior as the result of constant interaction between individuals, the environment, and behavior. Social cognition theory emphasizes the role of self-efficacy. Self-efficacy is a judgment of the individual's interaction with the environment. People with strong self-efficacy will be interested in new problems, fully engaged and struggle for it. Until they overcome the difficulties. This paper studies the self-efficacy into individual characteristic variables and explores its impact on the willingness

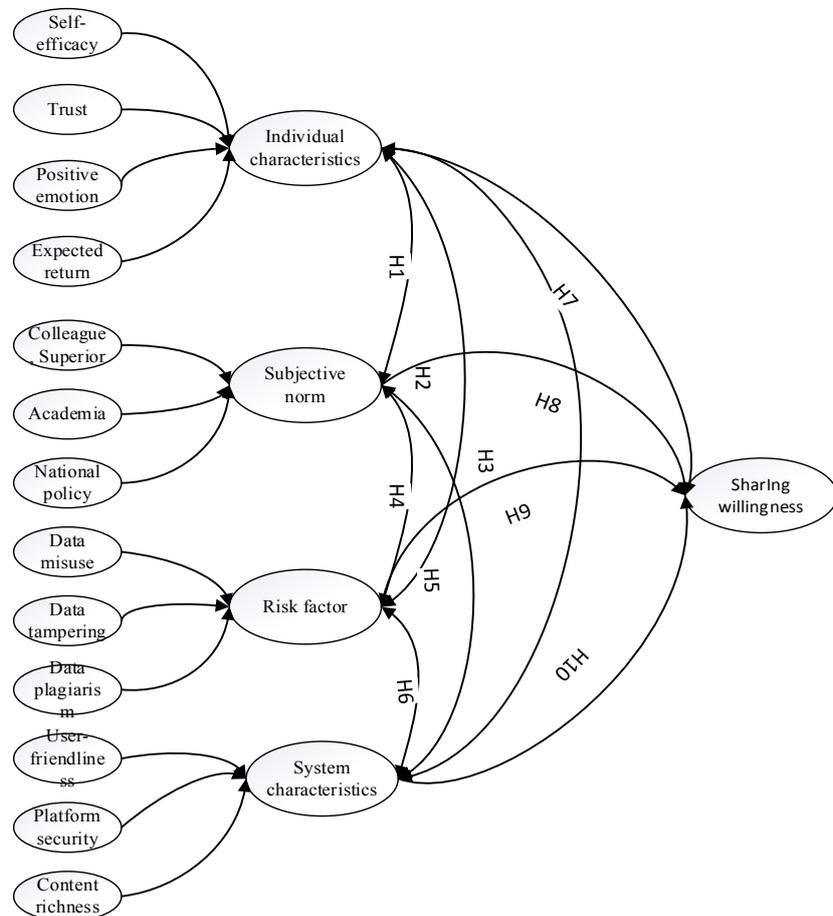of scientific data sharing in universities.

2.4. Social Exchange Theory

The social exchange theory founded by the famous American sociologist Homans [4] in the early 1960s emphasized psychological factors in human behavior, and is also called the theory of behavioral psychology. This theory holds that any form of interpersonal relationship can be called an exchange relationship. Only after the mutual psychological and material exchange process between people can we maintain and coordinate interpersonal relationships. This paper studies the exchange reciprocity mentioned in the social exchange theory as the expected return of data sharing among university researchers is material reward and spiritual reward, which are summarized into individual characteristic variables, and study its influence on the willingness of scientific data sharing in universities.

3. Hypothesis and Model Building

This paper studies the related literature about factors affecting the willingness to share scientific data in universities. Selecting the individual characteristics of university researchers, the subjective norms, risk factors, and the system characteristics of the scientific data sharing platform as external variables of research, combined with the theory of technology acceptance model and planning behavior, explores the impact of these factors on the willingness of scientific data sharing in universities. The following research hypotheses are proposed: H1: individual characteristics are significantly positively correlated with subjective norms, H2: individual characteristics are significantly negatively correlated with risk factors, H3: individual characteristics are significantly positively correlated with system characteristics, H4: subjective norms are significant negative correlation with risk factors, H5: subjective norms are significantly positively correlated with system characteristics, H6: system factors are significantly negatively correlated with risk factors, H7: individual characteristics are significantly positively correlated with sharing intentions, H8: subjective norms are significantly positively correlated with sharing intentions, H9: system characteristics are significantly positively correlated with sharing intentions, H10: risk factors are significantly negatively correlated with sharing intentions.

The individual characteristics of university researchers are mainly influenced by four factors: self-efficacy, trust, positive emotion and expected return. Subjective norms are mainly influenced by colleagues and superiors of university researchers, the academic circles in which they are located, and the national policies and regulations. Risk factors are mainly affected by three factors: data misuse, data tampering, and data plagiarism. The system characteristics of the scientific data sharing platform are mainly affected by consultation timeliness, platform security, interface friendliness and content richness.

In summary, the main factors affecting the scientific data sharing of universities are constructed, as shown in Figure 1.

Figure 1. Main influencing factors model of scientific data sharing in universities.

4. Research Method

4.1. Survey Design

In the design process of each research variable, many excellent and mature scales in China and abroad have been referred to. The individual characteristics are divided into four categories, among which self-efficacy, expected return, and positive emotions refer to the research of Albert Bandura, Hoffman, Ajzen, Li Mengjun [5], etc. Trust refers to the influence of Li Xueqin and Zhou Huaiying [6] on the shared willingness of cognitive trust and emotional trust. Subjective norms are adapted from the theory of reasoned action of Ajzen and Fishbein [7]. The risk factors refer to the research on the behavioral control of Ajzen's theory of planned behavior, and the domestic scholar Zhang Jinchao's [8] research on the extent of the impact of knowledge loss on the scientific data sharing attitude of researchers. The System Characteristics Scale refers to the extent to which Davis' perceived usefulness and perceived ease of use affect individual behavioral willingness. The sharing willingness scale refers to the scale study of knowledge sharing by Taylor, Todd [9], etc. This paper converts its content into scientific data sharing. The questionnaire used the Likert 5 scale to divide the answers selected by the respondents, from 5 points (very agree) to 1 point (very disagree).

4.2. Issuance and Collection

After completing the preliminary design of the questionnaire, 20 university teachers of information science were invited as pre-research objects. Based on the feedback information they gave, modified some of the vague or ambiguous topics. The content of the questionnaire was adjusted and supplemented reasonably to make the structure of the questionnaire more perfect. The officially issued questionnaires are sent online (Questionnaire Star Website, WeChat, QQ, E-mail, etc.) and offline (paper questionnaire). Over the past three weeks, 427 questionnaires were sent online and offline, 392 were collected, 364 valid questionnaires, and the effective rate of the questionnaire was 92.9%. Among them, 232 were masters and above, which guarantees the reliability of the sample.

5. Data Analysis

The data analysis part of this paper uses SPSS21.0 software to analyze and verify the proposed main impact model of scientific data sharing in universities. First, the questionnaire was tested for reliability and validity.

The reliability test is judged by the Cronbach coefficient (α value), and the closer the α value is to 1, the higher the reliability. Generally speaking, as long as the α value reaches 0.6, the reliability of the questionnaire is considered acceptable. As can be seen from Table 1, the overall α value of the questionnaire is 0.946, and the α values of each study variable are also above 0.8, indicating that the reliability of the measurement scale studied in this paper is very good.

Validity is divided into content validity, criterion validity and structural validity. This paper chooses to test the structural validity of the model, because in the actual operation process, the content validity requires experts to make conformity judgments on the measurement, and the criterion validity requirements have recognized benchmarks, which is more difficult to achieve. Therefore, structural validity is easier to implement. The structural validity of the questionnaire was measured by principal component analysis.

First pass the KMO and Bartlett spherical test, then factor analysis. KMO is an indicator for comparing the simple correlation coefficient and the partial correlation coefficient between variables. Its value range is between 0 - 1, and it is generally considered that the KMO value is above 0.7. As can be seen from Table 2,

Table 1. Questionnaire reliability test results.

Latent variable	Number of items	Cronbach's Alpha
Individual characteristics	8	0.937
Subjective norm	4	0.843
Sharing willingness	3	0.942
Risk factor	4	0.853
System characteristics	4	0.883
Total	23	0.946

Table 2. Test results of KMO and Bartlett.

Variable	Sampling enough Kaiser-Meyer-Olkin metrics	Bartlett's sphericity test	
Individual characteristics	0.922	Approximate chi square	2382.027
		df	28
		Sig.	0.000
Sharing willingness	0.788	Approximate chi square	3120.822
		df	3
		Sig.	0.000
Subjective norm	0.807	Approximate chi square	584.969
		df	6
		Sig.	0.000
Risk factor	0.811	Approximate chi square	668.122
		df	6
		Sig.	0.000
System characteristics	0.822	Approximate chi square	845.345
		df	6
		Sig.	0.000

the KMO values of the variables are all above 0.7, and the associated probability value P of the Bartlett spherical test is 0, which is lower than the significant probability level. It is indicated that the factor analysis of the questionnaire is appropriate.

The principal factors are then selected and rotated by principal component analysis. The cumulative variance contribution rate obtained is 75.626%. When the cumulative variance contribution rate of the general situation factor is 60% or more, it indicates that the scale has good structural validity. This shows that the research scale of this paper is more effective.

Analysis Table 3 shows that there is a positive correlation between the variables and a significant correlation at the 0.01 level. The correlation coefficients between individual characteristic, subjective norm, system characteristic variable and shared willing variable are 0.615, 0.507 and 0.501. Except for risk factors, the correlation coefficients of other variables and sharing willingness are greater than 0.4, indicating that the larger the values of these variables are. The higher the willingness to share. In line with the research hypothesis of this paper. The correlation coefficient between risk factors and willingness to share is 0.158, which indicates that the correlation between it and the willingness to share is weak, and the coefficient is positive. The same risk factor and the coefficients of the other three variables are also positive, which is not consistent with the hypothesis.

According to the range of r in the Pearson correlation coefficient (r value between 0.6 and 0.8, the variable is considered to be strongly correlated. When the r value is between 0.4 and 0.6, it is considered to be moderately correlated, and the r value is between 0 and 0.2, it is considered that the correlation is very weak or irrelevant.) It can be seen that the individual characteristics are moderately and positively correlated with the subjective norms, the individual characteristics

Table 3. Total variance of interpretation.

	Ingredient initial eigenvalue			Extract square sum loading			Rotation square sum loading		
	total	variance	accumulation	total	variance	accumulation	total	variance	accumulation
1	11.187	48.637	48.637	11.187	48.637	48.637	10.197	44.336	44.336
2	2.701	11.744	60.381	2.701	11.744	60.381	2.878	12.513	56.849
3	1.226	5.332	65.713	1.226	5.332	65.713	1.599	6.954	63.803
4	1.188	5.167	70.880	1.188	5.167	70.880	1.471	6.396	70.199
5	1.092	4.746	75.626	1.092	4.746	75.626	1.248	5.427	75.626
6	0.637	2.771	78.397						
7	0.616	2.677	81.074						
8	0.492	2.138	83.212						
9	0.476	2.068	85.280						
10	0.437	1.900	87.181						
11	0.419	1.821	89.002						
12	0.389	1.693	90.694						
13	0.340	1.478	92.172						
14	0.303	1.316	93.488						
15	0.277	1.206	94.695						
16	0.266	1.159	95.854						
17	0.227	0.986	96.839						
18	0.219	0.952	97.791						
19	0.206	0.895	98.687						
20	0.168	0.731	99.418						
21	0.119	0.516	99.934						
22	0.010	0.042	99.976						
23	0.005	0.024	100.000						

and the system characteristics are moderately and positively correlated, individual characteristics and the sharing willingness are highly significantly positively correlated, subjective norms and system characteristics are moderately positively correlated, subjective norms and sharing willingness are moderately positively correlated, and system characteristics and sharing willingness are moderately positively correlated. There was no significant correlation between the risk factors and the sharing willingness and the other two variables. The final research conclusions are the main influencing factors of scientific data sharing in universities: 1) individual characteristics of university researchers, 2) subjective norms, external environment and policy orientation, 3) system characteristics of data sharing platform.

Based on the predecessors TAM, TPB, social cognition and social exchange theory, this paper constructs a scientific data sharing willingness model for uni-

versity researchers, and analyzes and verifies the model through SPSS21.0. The overall fit of the model is good. The research results show that individual characteristics, subjective norms and system characteristics directly affect the scientific data sharing willingness of university researchers, and the risk factors have no significant impact on the scientific data sharing willingness of university researchers (Table 4).

6. Analysis of Research Results

1) The results of data analysis show that H7 is established, that is to say, individual characteristics have a direct positive impact on the scientific data sharing willingness of university researchers. Moreover, the individual characteristics have the most obvious influence on the scientific data sharing willingness of university researchers, compared with other three control variables. So we can be sure that the realization of the self-worth of university researchers, the trust in others and the shared environment, the willingness to give, and the spiritual and material rewards can strengthen the data sharing willingness of researchers. The input-output structure of scientific research should be optimized. Carry out data literacy education for researchers, cultivate them data sharing capabilities, and give certain incentives to those who actively participate in sharing.

2) Assuming H8 is established, it means that subjective norms have a direct positive impact on the scientific data sharing of university researchers. It can be

Table 4. Correlation coefficient and significance test of each variable.

		Individual characteristics	Sharing willingness	Subjective norm	Risk factor	System characteristics
Individual characteristics	Pearson correlation	1	0.615**	0.449**	0.184**	0.472**
	Significant (bilateral)		0.000	0.000	0.000	0.000
Sharing willingness	Pearson correlation	0.615**	1	0.507**	0.158**	0.501**
	Significant (bilateral)	0.000		0.000	0.000	0.000
Subjective norm	Pearson correlation	0.449**	0.507**	1	0.260**	0.450**
	Significant (bilateral)	0.000	0.000		0.000	0.000
Risk factor	Pearson correlation	0.184**	0.158**	0.260**	1	0.306**
	Significant (bilateral)	0.000	0.000	0.000		0.000
System characteristics	Pearson correlation	0.472**	0.501**	0.450**	0.306**	1
	Significant (bilateral)	0.000	0.000	0.000	0.000	

**Significantly correlated at 0.01 level (both sides).

seen that the national policy, the requirements of the academic community, and the mutual influence of peers can stimulate the willingness of researchers to participate in data sharing activities. The promotion of scientific data sharing behaviors can be enhanced through research and interpretation of national scientific data sharing policies, university appeals, and interactions between peers or colleagues. This will dispel the doubts of researchers on data sharing, and realize that data sharing is very necessary and important, enhance the enthusiasm of data sharing, and standardize scientific data sharing behavior. The specific methods that can be adopted are as follows: a) Research institutes such as the Science and Technology Department and libraries should regularly organize large-scale scientific data sharing publicity lectures to attract researchers' attention to data sharing. The content mainly focuses on the newly introduced national scientific data management and sharing policies, let researchers realize that scientific data sharing is the only way to develop in the future. b) Adopting the method of top-level design, popularizing and popularizing scientific data sharing through the decision-making level of university leaders. Then they will greatly affect the willingness of scientific researchers to share scientific data. c) Select active participants in data sharing in various subject areas in universities, and use them as excellent models to encourage other researchers to share.

3) Assume that H9 is established, indicating that the system characteristics have a direct positive impact on the scientific data sharing willingness of university researchers. It is concluded that the data sharing system is safe, complete and easy to operate, which can stimulate the data sharing of scientific researchers. Scientific data sharing platform construction is needed. Only when there have a real data sharing platform, researchers will consider the next sharing behavior. It can be designed with reference to the shared platform of existing universities in China. For example, Dspace open source software is used as the foundation to build a metadata description framework, select corresponding metadata standards for each subject area, or obtain a description by letting researchers fill out online forms. Pay attention to the friendly design of the platform interface, establish a good information navigation index, pay attention to the timeliness of consultation, the background staff should respond to the questions generated by the users in the sharing process on time, pay attention to the security of the platform, and set certain restrictions on the visiting users. And regularly maintain and update the platform.

4) It is assumed that the H10 data analysis results are not valid, which indicates that the risk factors do not have a direct negative impact on the scientific data sharing of universities. That is to say, university researchers are not deterred by the misuse, malicious tampering and plagiarism of scientific data. They dare to share the selfless dedication of scientific research. They believe that scientific data sharing will become a trend in academia. They believe that scientific data sharing will be in a normative and supervisable overall environment. This belief that is not afraid of risk and has the courage to explore should be

passed down. According to the statistical data feedback, although the results of the surveyed researchers show that their data sharing willingness is not affected by risk factors, the older researchers have a clear attitude towards risk factors, most of them think risk factors affect their willingness to share. Therefore, the author believes that it is necessary to focus on the data ethics of scientific researchers while strengthening the data skills training of researchers, and standardize data references.

7. Insufficient and Prospect

Although the research in this paper generally follows the scientific research methods, due to the influence of individual scientific research level, interpersonal skills, and scientific research conditions, the research has more or less deficiencies. The sample data collected in the first study mainly rely on the network to fill in, and lack face-to-face communication with the survey object, which makes the survey data have certain errors. Secondly, among the researchers who participated in the questionnaire, undergraduate and graduate students accounted for the vast majority, considering that they had just contacted scientific research projects or had never been exposed. There is also a certain bias in the cognition and understanding of scientific data sharing, which in turn affects subsequent data analysis and results. In addition, due to limited time, there is no research on the data sharing willingness of researchers in other universities, so the representativeness of the sample is not strong enough. The research methods for future research can be combined with questionnaires and expert interviews without relying on a single questionnaire survey. The analysis methods are not limited to basic analysis of variance. Future research can also be more accurate in the selection of samples. The scope of the sample can also be restricted to other scientific research units within the university. Researchers in the research institute are included in the discussion. If there is capacity to expand the sample capacity, consider using big data analysis tools like patron. It is believed that with the unremitting efforts of domestic researchers, China's scientific data sharing business will develop rapidly, narrowing the gap with western developed countries, and realizing the goal of China's scientific and technological power and cultural power.

Financed

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Ajzen, I. (1991) The Theory of Planned Behavior. *Organizational Behavior and*

Human Decision Processes, **50**, 179-211.
https://doi.org/10.1016/0749-5978(91)90020-T

[2] Davis, F.D. (1989) Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, **13**, 319-340.
https://doi.org/10.2307/249008

[3] Albert, B. (1986) Social Foundation of Thought and Action: A Social Cognitive Theory. Prentice Hall, New York.

[4] Homans, G.C. (1961) Social Behavior: Its Elementary Forms. *Revue Française De Sociologie*, **3**, 479-502.

[5] Li, M.J. (2014) Research on the Influencing Factors of the Implicit Knowledge Sharing Willing and Perceptual Sharing Ability of Enterprise Team Members. Xidian University.

[6] Li, X.Q., Zhou, H.Y. and Cai, X. (2010) An Empirical Study on the Relationship Between Trust and Knowledge-Sharing in University Research Team. *Library and Information Service*, **54**, 92-95.

[7] Fishbein, M. and Ajzen, I. (1974) Attitudes towards Objects as Predictors of Single and Multiple Behavioral Criteria. *Psychological Review*, **81**, 59-74.
https://doi.org/10.1037/h0035872

[8] Zhang, J.C. (2013) Research on the Willingness of Scientific Data Sharing of Researchers in Chinese Universities. *Information Studies: Theory & Application*, **36**, 25-30.

[9] Taylor, S. and Todd, P. (1995) Decomposition and Crossover Effects in the Theory of Planned Behavior: A Study of Consumer Adoption Intentions. *International Journal of Research in Marketing*, **12**, 137-155.
https://doi.org/10.1016/0167-8116(94)00019-K