



DOI 10.22363/2313-1683-2017-14-3-255-272

UDK 159.9.072

NON-TRIVIALITY OF THE RESULTS OF MILGRAM FIELD EXPERIMENT IN MOSCOW AND NEW YORK SUBWAY

Olga V. Mitina¹, Alexander Ya. Voronov²

¹ Lomonosov Moscow State University
Mokhovaya str., 11 (Bld 9), Moscow, Russia, 125009

² State Academic University for the Humanities
Yaroslavskaya str., 13, Moscow, Russia, 129366

Abstract. Non-triviality of the results of the field experiment conducted on Stanley Milgram's methodology in the New York and Moscow subway have been studied. The statistical significance of the difference between empirical and predicted results has been taken as the non-triviality criterion. 208 respondents (psychologists and students studying psychology) were asked to predict an experimental result in dependence on an experimenter's and subject's gender, a subject's age, and a city where the experiment was carried out. The obtained results have confirmed our hypothesis on non-triviality of the experiments in subways: it has been showed that there is a statistically significant difference between real behavior of subway passengers (in New York and in Moscow) and predictions made by Moscow and Tashkent respondents. Practically in most cases the predicted probability that a subject gives a seat after request of experimenter (young woman or young man) is much less than in reality. The structural equation modeling (SEM) has been used to analyze the data by constructing the model taking account of all factors mentioned above. The model fit the experimental data well (CFI = 0.919). It has been found that predicted results depend not only on gender, age, and residence of a respondent but also on the degree of familiarity with the research. The obtained data give an important material for a further study of the role of situational (an experiment design) and individual (respondent characteristics) factors in predicted results; they contribute to further understanding of the problem of creation and support of non-formal social norms in various cultures and show new aspects of research carried out on experimental methodology of Stanley Milgram.

Key words: non-triviality of empirical fact, experimental social psychology, prediction of research result, Milgram subway experiment, identification, cross-cultural differences, gender differences, age differences, non-formal social norms, structural equation modeling (SEM)

Introduction

In the field experiment in the New York subway (Takooshian, 1972; Milgram, Sabini, 1978; Milgram, 2001; Luo, 2004; Milgram, 2010) healthy looking psychology students (experimenters) asked random passengers (subjects) to give up their seat in a jam-packed car. The first original research was carried out by student Harold Takooshian (Takooshian, 1972) under supervising by his professor Stanley Milgram. In the experimental situation "Without motivation" the experimenters did not explain the subjects why they wanted to take the seat breaking informal social norms of behavior (Cialdini, Kallgren, and Reno,

1991; Bicchieri, 2006) in a subway. The result was unexpected for the scientists: more than half of the subjects gave up their seats.

Unbearable stressful responses of experimenters during collection of empirical data were also unpredictable (Takooshian, 1972; Milgram, Sabini, 1978; Luo, 2004; Cliffs, Schultz, Cialdini, Goldstein, Griskevicius, 2007). All of the experimenters complained about unexpectedly high intensity of negative emotions experienced during this experiment.

During 2008–2010 these results were confirmed by experiments in Moscow and Sankt-Petersburg subways (Al'-Batal, 2009; Voronov, Al'-Batal, 2010; Voronov, Mitina, Al'-Batal, Gorbunova, 2011; Mitina, Voronov, 2012).

The unexpectedness of results of all these experiments in the subway allowed us to propose a hypothesis on non-triviality of the scientific fact under consideration.

The important criterion of a true value of the obtained scientific knowledge is the statistically significant difference between the experimental result and its prediction. This criterion was suggested and justified (Mitina, Petrovskii, 2001).

The respondents predicting the result may be people:

— who are not experts in the field and they did not hear anything about such experiments;

— who are experts and they know about such experiments (while they studied psychology), but they do not remember results;

— who are experts in the field and say that they remember the results.

In the first part of this paper the results of analysis of an accuracy of predicted results made by psychologists for subway experiments are presented. Our independent respondents were citizens of Moscow and Tashkent. They predicted results of experiments carried out in New York, Moscow, and Tashkent. In truth the real experiment was not conducted in Tashkent but the respondents did not know about it (they were told that the experiment took place in Tashkent also). It was studied how much the place of an experiment conducting, the gender of an experimenter, the gender and age of a subject, the gender, age and place of residence of a respondent and his (or her) awareness about subway studies affect such predictions.

Unfortunately, we were not able to use for comparison the predicted results obtained in the USA by Harold Takooshian (Takooshian, 1972; Milgram, Sabini, 1978) because of undifferentiatedness (these predicted results were not separated according mentioned parameters; however, we showed that these parameters influence significantly on the prognoses).

In the second part of our paper we used SEM for cross-cultural analysis of the data obtained in Moscow and Tashkent. We have constructed the model taking into account all possible combinations of the factors mentioned above and their influence on predictions. The model fits data with CFI = 0,919. We have identified the statistical significance of 1) factor loadings, 2) correlations between latent variables determining an experimental scheme, 3) correlations between latent variables corresponding to the situations in the experimental design and characteristics of our respondents.

The obtained information is important both for the study of non-triviality of empirical facts and as an additional material for understanding how non-formal social norms in various cultures are formed and maintained (Scott, 1971).

Procedure and methods of research

The respondents (psychologists and students studying psychology) from Tashkent (April, 2011) and Moscow (October, 2011 — end of June, 2012) filled in the prognostic questionnaire in university classrooms during their lectures or seminars.

Here we quote the introductory part of the questionnaire:

“In 1972 a famous American psychologist Stanley Milgram carried out experiments in New York subway. A healthy looking young man or a woman asked a passenger to give him (or her) a seat in a subway car full of people without explaining why. In 2008—2010 such an experiment was carried out in Moscow and Tashkent subways. The results of all these experiments were subjected to statistical analysis. The subjects-passengers were divided in four groups (men at the age before and after 40 and women at the age before and after 40).

What do you think:

How many of 100 passengers chosen randomly in each gender-age group gave their seat to a young man and how many to a young woman in New York, Moscow, and Tashkent?

While answering these questions, in the end of each from 24 lines (after the colon) write down a number of passengers (a range of numbers is not allowed.)”

Following this introduction there were 24 lines for 24 different experimental situations (for 24 predicting answers) which consists of 8 similar situations in 3 cities: $2 \times 2 \times 2$, i.e. an experimenter’s gender multiplied by a subject’s gender multiplied by a subject’s age (younger or older 40).

Table 1 presents the results of real experiments in Moscow and New York.

Table 1

The results of experiments in New York (1972) and Moscow (2008—2010)

City	Gender of		Age of subject	The number of subjects (= the number of experi- ments)	% of subjects having given a seat*
	experimenter	subject			
New York	Male	Male	< 40 years	10	60.0
			> 40 years	9	44.4
		Female	< 40 years	8	50.0
			> 40 years	7	42.9
	Female	Male	< 40 years	6	83.3
			> 40 years	5	80.0
		Female	< 40 years	8	87.5
			> 40 years	3	66.7
Moscow	Male	Male	< 40 years	12	83.3
			> 40 years	10	30.0
		Female	< 40 years	11	72.7
			> 40 years	13	53.9
	Female	Male	< 40 years	21	95.2
			> 40 years	21	66.7
		Female	< 40 years	21	90.5
			> 40 years	17	70.6

* Those who gave the seat and those who just shifted a bit to give some room were considered together.

Our database was supplemented by the US data as professor Takooshian kindly gave us “Sheets” of manual recordings made by experimenters and observers during experimental study in New York City (Takooshian, 1972). In 1972 the experiment in New York was carried out in 4 variations. Table 1 presents the summarized results only for 2 of them, i.e. with “Verbal request without motivation” and with “Written request without motivation”; there were no statistically significant differences between the results only of these two variations (Milgram, Sabini, 1978). Thus, all 56 cases (subjects-passengers) of these two variations in the New York City experiment (having analyzed the experiments, we excluded several ambiguously interpreted cases of subject’s behavior from the US data) may be considered as one homogeneous group of data obtained in the version “Without motivation”.

In Moscow (and according to the experiment legend in Tashkent) all experimenters worked in the version “Verbal request without motivation”; Moscow experimenters asked 126 subjects-passengers.

Each experimenter in Moscow and New York asked several subjects-passengers.

Table 2 presents the gender and age of experimenters. All experimenters were students getting their master or doctoral degree of about the same age. Therefore, it was not possible to differentiate experimental situation in dependence on their age.

Table 2

Description of the experimenters’ group in New York (1972) and Moscow (2008–2010)

City	Number of experimenters	Females from them	Average age
New York	10	4	24.0
Moscow	15	9	21.1

Table 3 presents the description of respondents’ group. They are mainly Psychology students or the people who have already received Ph.D. in Psychology.

All the respondents got an additional question: how much do they know about these experiments? Among 175 Moscow respondents only 48 gave a positive answer but they added that they did not remember exact figures and they gave wrong prognosis very often.

Table 3

Description of the respondents’ group in Tashkent (2011) and Moscow (2012)

Respondents	Moscow		Tashkent (uninformed about the experiment)	Total
	Uninformed about the experiment	Informed about the experiment		
Total	127	48	55	230
Males	29 (23%)	10 (21%)	10 (18%)	49 (21%)
Females	98 (77%)	38 (79%)	45 (82%)	181 (79%)
28 years old and younger	110 (87%)	44 (91%)	54 (98%)	208 (90%)

We have compared these two respondents groups (familiar and not familiar with the experiments) and found out that those who knew something about experiments predicted higher percentage of getting seats in a subway car than those who did not know anything about experiments. The difference was significant. Thus, we have revealed a degree of latent knowledge. Recall that the experimental results have shown suddenly high percent of those who gave their seats.

In Tashkent the respondents were Psychology students of the 2nd and 3rd year of training when they did not yet take the course which would describe these experiments and were not familiar with them.

So we had three groups of respondents: one group (in Moscow) who knew about experiments and two groups (in Moscow and Tashkent) did not know about them.

87% of Moscow respondents and 98% of Tashkent respondents were not older than 28. At this age people do not fully understand that standing in a public transport could be a physical problem. For them, the resolution of the situation realized in the experiment is most likely determined by the force of the informal norms of behavior in a subway. To provide the homogeneity we restricted the respondent selection by this age. Thus we had the final sample of 208 subjects: 1) 110 Moscow respondents and 54 Tashkent respondents who were not familiar with experiments and 2) 44 Moscow respondents who knew about them. Cross-tabulation analysis with calculation χ^2 statistics and Gilford ϕ -coefficient showed the balance of respondents among each group according gender ($p = .408$). There were about in 4 times more females than males in both groups.

Results and Discussion

Study of non-triviality of the results of Milgram field experiment in the subway of New York and Moscow

In the Table 4 all prognoses which were made for all experimental situations differentiated by city (New York/Moscow/Tashkent), by gender of an experimenter and a subject, and by the subject's age are presented (24 situations in total). Because all respondents giving prognoses were differentiated on the subgroups according their gender, place of living (Moscow/Tashkent) and a prior familiarity with the experiment (6 subgroups in total) the table 4 contains $24 \times 6 = 144$ prognoses numbers and 16 real results of experiments (8 in Moscow plus 8 in New York; as it was said in Tashkent a real experiment was not conducted). So only 96 prognoses from 144 can be compared with the real results.

Table 4 shows that all predictions of positive response were lower than real results and these differences were significant for uninformed respondents and partly for respondents who had known about the experiment beforehand in the part concerning Moscow (because information about the experiments in Moscow is not included in traditional courses of social psychology yet). This is most likely due to the powerful informal social norm of passenger's behavior in the subway rooted in the minds of respondents (even informed respondents!). Perhaps, the respondents, being psychology students, put themselves in the place of experimenters (also psychology students) (Scott, 1971). They were fearful of misconduct. The strength of this norm can manifest itself (from respondent's point of view) in low percent of positive behavioral responses made by all kinds of subjects on the experimenter request.

There are only 3 cases of prognosis which are higher than the experimental results. All these 3 cases are statistically insignificant and belong to group of informed respondents. In these three cases an experimenter asked an old man to give (to him or her) a seat. And male respondents overestimated the situation concerning a male experimenter (in New York and Moscow) and female respondent overestimated the situation concerning a female experimenter (in Moscow). Because of insignificance of differences in these cases we will not to discuss the possible reasons of such results.

Table 4

Comparison of experimental results with the results of predictions (% of subjects who gave their seats)

City of the experiment	Experimenter's gender	Subject's gender	Subject's age	Real results of the experiments	Prognoses of the experiment's results					
					Respondents uninformed about the experiment			Respondents informed about the experiment		
					Moscow		Tashkent		Moscow	
		Male	Female	Male	Female	Male	Female	Male	Female	
New York	Male	Male	< 40 years old	60.00	28.29**	26.82**	32.00*	15.33**	36.86	35.89
		Female	> 40 years old	44.44	24.46	18.59**	24.56	11.07**	44.71	32.58
	Female	Male	< 40 years old	50.00	22.75*	25.31	20.22**	14.14**	37.71	33.92
		Female	> 40 years old	42.86	16.96*	15.58**	18.33*	9.28**	19.71	24.36
Moscow	Male	Male	< 40 years old	83.33	64.71	67.27	53.33	27.53**	76.14	73.64
		Female	> 40 years old	80.00	54.42	49.65	39.67*	22.28**	52.86	62.47
	Female	Male	< 40 years old	87.50	29.92**	34.33**	39.44**	15.02**	38.14**	43.53**
		Female	> 40 years old	66.67	17.42**	23.66*	23.67*	14.00**	36.29	33.53
Tashkent	Male	Male	< 40 years old	83.33	21.54**	19.41**	24.22**	13.33**	36.29**	33.22**
		Female	> 40 years old	30.00	16.58	14.67	18.44	10.70**	31.14	27.17
	Female	Male	< 40 years old	72.73	17.08**	21.25**	19.33**	10.93**	25.57**	33.72**
		Female	> 40 years old	53.85	12.33**	15.39**	16.78**	8.49**	19.43**	24.92**
Moscow	Male	Male	< 40 years old	95.24	66.63**	66.90**	53.44**	32.74**	73.57**	70.17**
		Female	> 40 years old	66.67	54.54	51.55	37.33**	23.81**	60.86	70.28
	Female	Male	< 40 years old	90.48	28.08**	30.82**	33.44**	15.63**	34.71**	48.92**
		Female	> 40 years old	70.59	16.33**	20.41**	23.44**	11.98**	31.00**	37.53**
Tashkent	Male	Male	< 40 years old		25.58	25.57	24.44	29.44	26.86	45.56
		Female	> 40 years old		21.83	16.42	26.67	22.30	20.57	35.11
	Female	Male	< 40 years old		29.25	26.90	29.22	27.44	40.43	35.58
		Female	> 40 years old		15.42	18.35	16.67	19.47	31.43	27.58
Tashkent	Male	Male	< 40 years old	The experiment was not conducted	64.04	66.80	65.22	58.19	55.00	74.06
		Female	> 40 years old		48.58	53.89	47.44	42.12	42.43	64.64
	Female	Male	< 40 years old		37.58	33.34	41.89	33.86	33.57	47.50
		Female	> 40 years old		23.25	23.08	30.78	25.33	15.71	34.53

Note. Statistical significance of prognoses differences from real results was tested using binomial statistics: * .05 > p > .01; ** p < .01; all other prognoses are not significantly differing from real results.

As it has been expected, the most distant from reality results came out in the group of uninformed respondents: 79.7% (51 of 64) of their prognoses differ significantly from experimental results; the levels of significance of these differences tested by using binomial statistics (here and below “binomial statistics” just will be designated in abbreviated form as “binst”) were $p < .05$ for 7 of 51 (13.7%) predictions and $p < .01$ for 44 of 51 (86.3%) ones; young women from Tashkent (subgroup of uninformed respondents), in contrast to all of other five subgroups, made wrong predictions for all 16 positions (8 in New York and 8 in Moscow) with the lowest p -values $< .01$ (binst). Each of these 16 predictions was the lowest among all 6 subgroups of respondents. At the same time their analogous 8 predictions for Tashkent are considerably higher than for Moscow and New York, i.e. for these respondents the location of the experiment was the most important factor.

For comparison, in the group of informed respondents we have only 43.8% (14 of 32) of predictions differing significantly from experimental results; all (14 of 14) of the levels of significance of these differences had $p < .01$ (binst). We haven't found distinctions between gender subgroups of informed respondents concerning distribution of statistically significant and non-significant differences between all of predictions and corresponding frequencies of positive responses of subjects in all of 16 experimental situations.

The level of significance of the total distinction in prognoses exactness between uninformed respondents and informed ones is $p = .004$ (Mann-Whitney U Test).

Perhaps, during thinking about their prognostic decisions respondents identify themselves not only with experimenters but with subjects as well. One can think about this kind of identification in the cases when gender, age, and residence of respondent coincide with the corresponding characteristics of a subject. We call it complete identification. In the case of partial coincidence of these characteristics we could talk about partial identification of a respondent with a subject. In these two cases (complete and partial identification of a respondent with a subject) we could expect higher predictions as compared with predictions made by respondents free from identification with a subject: the phenomenon of the fundamental error of attribution (I am more personally oriented than other typical person of my gender, age, place of living etc.). This effect must be more pronounced when identification is more complete. Our data confirm these suggestions in 8 of 12 cases of complete identification and in 34 of 60 cases of partial identification.

The situation when a young man (experimenter) asks a woman (subject-passenger) over 40 to yield him a seat is the strongest misconduct in a subway. Such a situation is the most stressful and therefore it is the most difficult for the participants (for the young male experimenter and for the female subject-passenger). If our respondents thought similarly while making their predictions then we could assume that the minimal predicted frequencies of passengers' positive responses should be for such a situation; our assumption was confirmed by all 6 groups of respondents for New York subway and by 5 from 6 groups for Moscow subway.

In the real experiment this minimum of passenger response in such a situation was observed only in New York — 42.9%. In Moscow, such a minimum is 30.0% occurred in another “scene” when the young male experimenter made his unexpected request not to a female passenger over 40 years of age, but to a male passenger over 40 years of age. Such an experimental result in Moscow can be explained by the fact that men older than 40 years are the least healthy part of the Moscow population; another explanation can be revealed from breaking within gender “concurrent” subordination (analogically it was

expected that older women will not yield a seat to younger female experimenters, but really it did not happen). Only the subgroup of uninformed (!) female respondents from Moscow (i.e., only one of all 6 subgroups of our respondents) correctly predicted the result of this “real Moscow minimum”.

In a group of 8 real Moscow experiments the result (53.9%) of the theoretically minimal in passengers’ response to the experimental situation (a young male experimenter and a female passenger over 40 years of age) follows immediately after 30% response in the situation of the “real Moscow minimum”.

In New York, 44.4% of the response of male passengers over the age of 40 to the request of the young male experimenter (the “real Moscow minimum” situation) follows immediately after the theoretical and practical minimum of the American passengers’ responses (see 42.9% above).

Now let us discuss the predictions about experiments in Tashkent (which has never been carried out) made by the same 6 subgroups of respondents. The respondents of 4 subgroups predicted minimal number of positive responses for the situation when an experimenter is a young man and a subject is a woman older than 40 years old. Almost all respondents (except informed women from Moscow) predicted sufficiently low percent positive results for the situation when the experimenter is a young man and the subject is a man older than 40.

Some gender aspects of the predictions of the results obtained in subway experiments

In 70 of total 72 prognoses (6 subgroups of respondents \times 3 cities \times 4 categories of subjects-passengers) of the frequencies of positive reactions on the request to give a seat to the young female experimenter these frequencies are higher than the analogical prognostic frequencies but related to the young male experimenter. This result can be explained by the fact that a such behavior of a young woman (an experimenter) breaks informal social norms of behavior much less than a similar request made by a young man (such a request of a young man is a challenge to accepted norms of behavior of men in a public place). One can think that this gender effect reveals itself when a respondent makes predictions. Really (in the experiments), this gender effect manifests in both cities (Moscow and New York) for all 4 categories of subjects-passengers.

Uninformed respondents predicted the behavior of female subjects less accurately than of male subjects: the group of these respondents gave 31 predictions statistically significantly differing from reality from total 32 predictions of behavior of female subjects (compare similar values for male subjects: 20 from 32). The level of significance of this gender distinction is $p < .001$ (φ^* -criterion with $\varphi^*_{emp} = 3.86$).

The same gender phenomenon was also observed for the group of informed respondents: these respondents gave 10 predictions statistically significantly differing from reality from total 16 predictions of behavior of female subjects in comparison with 4 from 16 for male subjects; the level of significance is $p = .014$ (φ^* -criterion with $\varphi^*_{emp} = 2.195$).

This second gender effect (the first one was described above) may be probably explained by unexpectedly high frequencies of real positive responses of female subjects after the experimenter’s request.

*The cross-cultural aspect of the accuracy of prediction
of the results of subway experiments*

The integrated sample of all 6 subgroups of respondents gave more precise predictions of results of the New York experiments than of the Moscow ones. They gave (from total 48 for each town) 22 (45.8%) and 9 (18.8%) predictions non-significantly statistically differing from reality for New York and for Moscow, respectively. The level of significance of this cross-cultural distinction is $p < .001$ (φ^* -criterion with $\varphi^*_{emp} = 2.89$).

This cross-cultural difference between the accuracy of prognoses data (New York vs. Moscow) was more obviously [$p < .001$ (φ^* -criterion with $\varphi^*_{emp} = 3.881$)] for the group of informed respondents: they gave 14 (87.5%) and 4 (25%) predictions (from total 16 for each town) statistically non-significantly differing from reality for New York and for Moscow, respectively; probably the awareness of this group was based only on the published US data [for example, in English (Milgram & Sabini, 1978) and in Russian (Milgram, 2001)].

We found no distinctions between gender subgroups of the group of informed respondents concerning distribution of statistically significant and non-significant prognoses. Concretely: in each of these gender subgroups statistically significant distinctions of predicted results from real ones were noticed in the identical 1 of 8 (12.5%) and in the identical 6 of 8 (75%) cases for New York and Moscow, respectively; statistical significance of each of these 14 (7+7) distinctions had also the same level: $p < .01$ (bins).

The group of uninformed respondents gave 8 (25%) and 5 (15.6%) predictions (from total 32 for each town) statistically non-significantly differing from reality for New York and for Moscow, respectively. However, this cross-cultural effect was statistically non-significant: $p > .1$ (φ^* -criterion with $\varphi^*_{emp} = .94$).

The distinction between cross-cultural (Moscow vs. Tashkent) subgroups of uninformed respondents concerning analogous distribution of statistically significant (10 vs. 3) and non-significant (22 vs. 29) differences amongst all of their 64 predictions (32 for Moscow respondents plus 32 for Tashkent ones) was at the level of statistical significance $p = .012$ (φ^* -criterion with $\varphi^*_{emp} = 2.256$).

The using SEM for analyzing all of the results

Using SEM (Bentler, 2000) allows us to summarize separate findings made in previsions section, to bring together all differences and build a complete picture of individual fragments just as a large mosaic is made.

The predictions made under various combinations of factors determining an experimental situation were considered observable variables. The latent dummy variables were created corresponding to each possible value of all variables determining the experimental situations. For variable city latent dummy variables Moscow, New York, Tashkent were created, for variable experimenter gender latent dummy variables Female and Male were created and so on.

According to the theoretical model, each latent dummy variable, being the level of realization of the factor, determines observable variables (situations) with a corresponding level of realization of this factor (see Fig. 1).



Fig. 1. The theoretical model describing the influence of all possible combinations of levels of realization of factors determining an experiment on the results of predictions

Thus, all measured variables (the predictions about behavior of subjects) in New York are determined by the latent variable “New York”. All predictions about behavior of subjects with respect to an experimenter being a young man are determined by the latent variable “young man” and so on.

Since each situation is determined by four factors, in the full model there are four arrows from any four latent variables entering in each rectangle (a dependent variable). The calculated model can differ from the full model because some determinations can be insignificant.

The analysis will allow us to talk about convergent influence of the latent dummy variable corresponding to this or that value of each variable deterring experimental situation if the majority of predictions corresponding to this latent variable is confirmed. At the same time a “non-relevant” latent variable for this or that measured variable should not give significant predictions. Apart from this, the variables should not be completely synonymous or antonymous (i.e. the absolute value of the coefficient of correlation between them has to be less than 1).

The structural model includes correlations between all latent dummy variables: characteristics of experimental situations, characteristics of respondents making predictions (their gender, age, residence).

Having used the SEM we have found out:

- which of the factor loadings are significant;
- which of the correlations between factors determining an experimental situation are significant;
- which of the correlations between factors determining an experimental situation and characteristics of respondents are significant.

We should note that there are no cases when any loading should not be according the theoretical model but they are according to the data.

The model is good (CFI = .919).

If a latent variable loads a measured variable significantly, this means that the corresponding information (the value of this latent variable) affects significantly the corresponding prediction.

The analysis of Table 5 allows us to make the following conclusions.

The predictions are mostly dependent on the information about a city where an experiment was carried out. However, in Tashkent the most important information was the fact that the experimenter was a young man and the place of the experiment affected the predictions much less. Therefore, we can assert that there are convergent and divergent impacts of these latent variables. First, almost all measured variables were loaded by corresponding latent variables. Second, the latent variables corresponding to the cities where the experiments were carried out correlate between themselves, however, this correlation differs from 1 significantly (See Table 6).

The information that the experimenter is a young man affects significantly the predictions in Tashkent, while in New York the predictions are affected significantly by the fact that the experimenter is a young woman.

The latent variable “a subject is younger than 40” affects the predictions significantly in Moscow and New York, while the latent variable “a subject is older than 40” is very important for predictions in Tashkent.

Table 5

Significant factor loadings for each latent variable according to the experimental data with *p*-values < ,05

City	Experimenter	Subject		City			Experimenter		Subject's gender		Subject's age	
		Gender	Age	New York	Moscow	Tashkent	Young man	Young woman	Male	Female	Younger 40	Older 40
New York	Young man	Male	Younger 40	.624					.254		.477	
		Female	Older 40	.682			.119		.477			
	Young woman	Male	Younger 40	.771							.293	
		Female	Older 40	.921								
Moscow	Young man	Male	Younger 40	.672			.669				.223	
		Female	Older 40	.779			.660					.102
		Male	Younger 40	.676			.146		.444			
		Female	Older 40	.712					.310			.264
	Young woman	Male	Younger 40		.616				.244		.508	
		Female	Older 40		.729			.527				
		Male	Younger 40		.843						.329	
		Female	Older 40		.896					-.164		
Tashkent	Young man	Male	Younger 40	.664			.654			.148		
		Female	Older 40	.867			.688					
	Young woman	Male	Younger 40	.682					.682			
		Female	Older 40	.628					.346		.384	
Tashkent	Young man	Male	Younger 40			.737			.205	.464		
		Female	Older 40			.548	.125					
	Young woman	Male	Younger 40				.985					
		Female	Older 40				.671				.417	
Tashkent	Young man	Male	Younger 40			.757		.648				
		Female	Older 40			.825		.627			.130	
	Young woman	Male	Younger 40			.705			.513			
		Female	Older 40			.577			.241		.520	

Notes. 1. If, according to the theoretical model (Fig. 1), loading of a latent dummy variable on the measured variable has to exist and according to calculation on the experimental data this loading is significant we put the value of the loading in the corresponding cell of the table on the white background [all these loadings "on the white background" are positive as were expected; the only exclusion is significantly negative factor loading by female gender of the subject in the situation when young man asks older woman to give him a sit in Moscow subway (the value of this unique negative factor loading is highlighted in the corresponding cell in bold italics); the answer the question why in this situation the factor played opposite role is the subject of the future study]. 2. If the corresponding factor loading should not exist in the theoretical model and it is not significant according to the experimental data the corresponding cell is empty and white. 3. If the loading should be significant (according to theoretical model) but is not such according the empirical model the corresponding cell is grey color

Table 6

Correlation of independent variables (of the factors fixing the experiment and of respondent's characteristics)

	City			Experimenter		Subject			Respondent			
	New York	Moscow	Tashkent	Young man	Young woman	Male	Female	< 40	From Tashkent	Female	Age	Prior knowing about the experiment
City												
New York												
Moscow	.832											
Tashkent	.698	.713										
Male	.634	.654	.616									
Female	-.304	-.324	-.310	-.384								
Male				.100	.352							.249
Female				.237	.185	.463			-.126			.132
< 40				.225	-.147							.118
> 40	.288	.335	.261					-.307				

Note. All correlations are significant: p -value < .05.

It is worth mentioning that the information about a subject's gender positively affects predictions if the experimenter's gender is the same as the subject's gender: 5 times from 6 possible cases for men and 6 times for 6 possible cases for women. Perhaps, it is a manifestation of gender solidarity.

The positive response was predicted more often for young subjects in Moscow and New York (6 from 8), while in Tashkent for older subjects (3 from 4).

The first part of Table 6 (the correlations of latent variables determining the experiment design) allows us to find correlations between attitudes to give a positive prediction. It is possible to assume that there is a general attitude to think that people will give a seat in a subway car if one is asked. This could happen in any city and in a majority of situations. So it reflects the attitude that people can break social norms and behave as persons (not members of a group). The latent variable "the experimenter is a young woman" stands out from this pattern. The correlations of this variable with all other latent variables are negative. Most probably this fact can be interpreted as a manifestation of a general custom to give a seat to a woman, i.e. we met the situation when respondent predicted the giving a seat not because of breaking social norm, but because of substitution of one social norm by another. Respondents thinking so give much higher predictions when the experimenter is a young woman.

The right-hand side of Table 6 presents correlations between respondent's characteristics and their attitudes toward results.

We can say that respondents in Moscow are more likely to give positive prognoses (that a subject will give a seat) than respondents in Tashkent. And if to say about Tashkent's respondents only that thinking about people who live in the same city with them better than about people from other cities.

Female respondents expect more positive subject's reactions in Tashkent. Relatively older respondents consider Moscow situations positively. But they are wary of men and younger subjects.

Prior knowledge of the experiment makes predictions more accurate.

Conclusion

The results of the study confirmed our hypothesis about the non-triviality of the Milgram field experiment in the subway and indicated significant scientific value of this experiment.

It is so surprising for the common understanding that even knowing about it can be implicitly corrected by an informed respondent (in these cases prognoses are lower than real result).

Our assumption that respondents put themselves in the experimenter's place needs to be checked. However, let us recall that the experimenters (according to their reports) felt uneasily to ask a subject to give them a seat. In some way it can explain the situation with understating results by respondents. As usual we think about someone sitting in subway as about the crowd and cannot image that asking him or her personally reveal personality attitudes and behavior (not group ones).

Two gender differences were revealed and discussed during the analysis of the prognostic assessments: 1) prognostic frequencies of positive reactions on the request to give a seat

to the young female experimenter are higher than the analogical prognostic frequencies but related to the young male experimenter, 2) this finding is less trivial one: our respondents (especially, uninformed ones) predicted the behavior of female-subjects in subway experiment less accurately than of male-subjects.

The main cross-cultural aspect of our analysis (many informed respondents gave more precisely predictions of results of the New York experiments than of the Moscow ones) may be explained by the awareness of these respondents based only on the publications of US experiments.

The using of SEM allowed us to summarize separate findings obtained in the course of the test analysis of our hypothesis of non-triviality to build a complete picture of interrelationships between all situational and individual characteristics of the survey and of the experiment.

The data got during the survey gives material for future studying of the role of situational (experimental design) and individual (respondents' variables) factors in prognosis estimations of empirical results of the experiments, makes impact on the understanding of the problem of forming and maintaining of informal social norms in different cultures and highlights new aspects of studies of these phenomena.

Acknowledgments

We thank Professor H. Takooshian (Fordham University, NYC, USA) for unpublished data reports of experimenters and observers who took part in the research in New York in 1972 (Takooshian, 1972).

Also we thank Moscow psychologists G. Larina (Higher School of Economics — National Research University), I. Novikova (Peoples' Friendship University of Russia), and V. Spivakovskiy (Lomonosov Moscow State University) for recruiting respondents into prognostic survey.

The study was supported by grant Russian National Fond "Psychosemantic analysis of consciousness and the unconscious (on the material of general psychology, political psychology, psychology of art)" No. 17-18-01610.

REFERENCES

- Al'-Batal, T.M. (2009). Problema soblyudeniya social'nyh norm: eksperimenty v moskovskom i sankt-peterburgskom metro (po Milgremu i Sabini). *Prakticheskaya psihologiya obrazovaniya 21 veka: problemy i perspektivy*. Conference Proceedengs (pp. 218—226). Kolomna. (In Russ.).
- Bentler, P.M. (2000). *EQS 6 structural equations program manual*. Encino CA: Multivariate Software Inc.
- Bicchieri, C.(2006). *The grammar of society: the nature and dynamics of social norms*. New York: Cambridge University Press. doi: 10.1017/CBO9780511616037.
- Cialdini, R.B., Kallgren, C.A, & Reno R.R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. *Advances in Experimental Social Psychology*, 58, 1015—1026. doi: 10.1016/S0065-2601(08)60330-5.
- Cliffs, N.J., Schultz, N., Cialdini R.B., Goldstein, N.J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18 (5), 429—434.
- Luo, M. (2004, 14 Sept.). Revisiting a social experiment, and the fear that goes with it. *New York Times*, 153 (52972), B1-B8.
- Milgram, S. (2001). *Eksperiment v social'noi psihologii*. St. Petersburg: Piter. (In Russ.).

- Milgram, S. (2010). *The individual in a social world: Essays and experiments*. London: Pinter & Martin Ltd.
- Milgram, S., Sabini, J. (1978). On maintaining social norms: a field experiment in the subway. In A. Baum, J.E. Singer, S. Valins (Eds.). *Advances in environmental psychology* (vol. 1, pp. 31—40). Hillsdale, NJ: Erlbaum.
- Mitina, O.V., Petrovskii, V.A. (2001). Netrivial'nost' nauchnogo fakta i psihologicheskaya ekspertiza psihologov kak ekspertov. *Mir psihologii*, (4), 243—248. (In Russ.).
- Mitina, O.V., Voronov, A. Ya. (2012). Netrivial'nost' polevogo eksperimenta v metro (paradigma Stenli Milgrema): sopostavlenie prognoznyh i real'nyh rezul'tatov. In V.A. Barabanshikov (Ed.). *Eksperimental'nyi metod v strukture psihologicheskogo znaniya* (pp. 656—663). Moscow: Institut psihologii RAN Publ. (In Russ.).
- Scott, J.F. (1971). *Internalization of norms*. Prentice Hall: Englewood.
- Takooshian, H. (1972). An observational study of norms in the New York City subway. *Report on a class field experiment* (Unpublished manuscript).
- Voronov, A. Ya., Al'-Batal, T.M. (2010). Soblyudenie neformal'nyh social'nyh norm: polevye eksperimenty v metro. In V.A. Barabanshikov (Ed.), *Eksperimental'naya psihologiya v Rossii. Tradicii i perspektivi* (pp. 781—786). Moscow: Institut psihologii RAN Publ. (In Russ.).
- Voronov, A. Ya., Mitina, O.V., Al'-Batal, T.M., Gorbunova, A.I. (2011). Soblyudenie neformal'nyh social'nyh norm: polevye eksperimenty v paradigme S. Milgrema. In V.A. Barabanshikov (Ed.). *Sovremennaya eksperimental'naya psihologiya* (vol. 2, pp. 375—392). Moscow: Institut psihologii RAN Publ. (In Russ.).

© Mitina, O.V., Voronov, A. Ya., 2017

Article history:

Received 8 April 2017

Revised 6 June 2017

Accepted 14 June 2017

For citation:

Mitina, O.V., Voronov, A. Ya. (2017). Non-Triviality of the Results of Milgram Field Experiment in Moscow and New York Subway. *RUDN Journal of Psychology and Pedagogics*, 14 (3), 255—272. DOI 10.22363/2313-1683-2017-14-3-255-272

Bio Note:

Olga V. Mitina — Ph.D. in Psychology, leading scientific fellow of Faculty of Psychology of Lomonosov Moscow State University (Moscow, Russia). E-mail: omitina@inbox.ru

Alexander Ya. Voronov — Ph.D. in Biology, associate professor of Faculty of Psychology of State Academic University for the Humanities (Moscow, Russia). E-mail: a_voronov@inbox.ru

НЕТРИВИАЛЬНОСТЬ РЕЗУЛЬТАТОВ ПОЛЕВОГО ЭКСПЕРИМЕНТА МИЛГРЭМА В МОСКВЕ И НЬЮ-ЙОРКЕ

О.В. Митина¹, А.Я. Воронов²

¹ Московский государственный университет им. М.В. Ломоносова
Моховая ул., 11, стр. 9, Москва, Россия, 125009

² Государственный академический университет гуманитарных наук
Ярославская ул., 1, Москва, Россия, 129366

Исследовали нетривиальность результатов полевого эксперимента, проведенного по методологии Стэнли Милгрэма в нью-йоркском и московском метро. Критерием нетривиальности служила статистическая значимость различия между эмпирическим и прогнозируемым результатом. 208 респондентам (психологам и студентам, изучающим психологию) было предложено предсказать результат исследования в зависимости от пола экспериментатора и испытуемого, возраста испытуемого и города, в котором проводили эксперимент. Полученные результаты подтвердили гипотезу авторов о нетривиальности экспериментов в метро: было показано, что существует статистически значимое различие между реальным поведением пассажиров метро (в Нью-Йорке и в Москве) и прогнозами московских и ташкентских респондентов. В большинстве случаев предсказанная вероятность того, что испытуемый уступит свое место по ничем не мотивированной просьбе экспериментатора (девушки или юноши), была намного меньше, чем в эксперименте. Для анализа данных путем построения модели с учетом всех упомянутых факторов применили метод структурного моделирования (*SEM*). Модель хорошо соответствует экспериментальным данным ($CFI = 0,919$). Было обнаружено, что предсказанные результаты зависят не только от пола, возраста и места жительства респондента, но и от степени его знакомства с исследованием. Полученные данные дают важный материал для дальнейшего изучения роли ситуационных (дизайн эксперимента) и индивидуальных (характеристик респондента) факторов в прогнозируемых результатах; они способствуют дальнейшему пониманию проблемы формирования и поддержания неформальных социальных норм в различных культурах и освещают новые аспекты исследований, проводимых по экспериментальной методологии Стэнли Милгрэма.

Ключевые слова: нетривиальность эмпирического факта, прогноз результатов исследования, экспериментальная социальная психология, эксперимент Милгрэма в метро, идентификация, гендерные различия, возрастные различия, кросс-культурные сравнения, неформальные социальные нормы, метод структурного моделирования (*SEM*)

Благодарности и финансирование:

Исследование было поддержано грантом Российского Национального Фонда (РНФ) «Психосемантический анализ сознания и бессознательного (на материале общей психологии, политической психологии, психологии искусства)» № 17-18-01610.

© Митина О.В., Воронов А.Я., 2017

История статьи:

Поступила в редакцию: 8 апреля 2017

Принята к печати: 14 июня 2017

Для цитирования:

Митина О.В., Воронов А.Я. Нетривиальность результатов полевого эксперимента Милгрэма в Москве и Нью-Йорке // *Вестник Российского университета дружбы народов. Серия: Психология и педагогика*. 2017. Т. 14. № 3. С. 255—272. DOI 10.22363/2313-1683-2017-14-3-255-272

Сведения об авторах:

Митина Ольга Валентиновна — кандидат психологических наук, доцент, ведущий научный сотрудник факультета психологии Московского государственного университета им. М.В. Ломоносова (Москва, Россия). E-mail: omitina@inbox.ru

Воронов Александр Ярославич — кандидат биологических наук, доцент факультета психологии Государственного академического университета гуманитарных наук (Москва, Россия). E-mail: a_voronov@inbox.ru