

# ПСИХОЛОГИЯ И СОВРЕМЕННЫЕ ТЕХНОЛОГИИ

УДК 159.9

ГРНТИ 15.41.21

## ПСИХОЛОГИЧЕСКИЕ ФАКТОРЫ ЭФФЕКТИВНОСТИ ОПРОВЕРЖЕНИЯ ДЕЗИНФОРМАЦИИ В СОЦИАЛЬНЫХ СЕТЯХ<sup>1</sup>

©2022 г. Т.А. Нестик \*, Е.А. Михеев\*\*

*\*доктор психологических наук, профессор РАН, заведующий лабораторией социальной  
и экономической психологии,  
Институт психологии Российской академии наук, Москва;  
E-mail: nestik@ipras.ru*

*\*\*соискатель, Институт психологии Российской академии наук, Москва,  
E-mail: mih-news@mail.ru*

В статье анализируются психологические факторы эффективности противодействия распространению фейковой информации в социальных сетях. Выделяются психологические механизмы, снижающие действенность опровержения дезинформации (эффекты обратного действия, «консенсуса экспертов», эмоционального заражения, дезинформации и др.). Рассматриваются существующие подходы к противодействию дезинформации, в том числе: гибридный (сочетания метод проверки человеком и метод машинного обучения), функциональный (сочетание анализа учетных записей пользователей, их сообщений и контекстного метода) и сетевой (сочетание возможностей методов на базе искусственного интеллекта и экспертных оценок). Приводятся результаты апробации опросника «Отношение личности к дезинформации в социальных сетях». Показано, что характеристики отношения интернет-пользователей к дезинформации как социальному явлению и тем или иным способам борьбы с ней, действительно, связаны с доверием к опровержениям и способностью к распознаванию фейков. Приводятся результаты серии экспериментов, проведенных среди пользователей социальных сетей, склонных к пересылке фейковых сообщений (N = 1208) с целью

---

<sup>1</sup> Статья написана при поддержке гранта РФФИ №18-29-22046 мк.

изучения роли персонализации источника опровержения, конспирологичности опровергающего сообщения и его эмоциональности в противодействии дезинформации. Показано, что опровержение, облеченное в форму конспирологического предупреждения, является более эффективным, чем опровержение от официальной организации; а позитивно (юмористически) окрашенное опровержение более эффективно, чем нейтрально окрашенное. Одновременно опровержения менее эффективны, если они внушают страх. Это хорошо согласуется с проведенными ранее исследованиями, показавшими, что при нагнетании тревоги включаются защитные психологические механизмы, приводящие к недооценке угрозы и закреплению первоначальных заблуждений. Предложены рекомендации по повышению эффективности опровержения фейков, распространяемых в социальных сетях. Намечены перспективы дальнейших исследований.

*Ключевые слова: дезинформация, фейки, опровержение, эффект бумеранга, конспирология, юмор.*

## ПСИХОЛОГИЧЕСКИЕ МЕХАНИЗМЫ, СНИЖАЮЩИЕ ДЕЙСТВЕННОСТЬ ОПРОВЕРЖЕНИЯ ФЕЙКОВ

Как известно, противодействие распространению фейковой информации в социальных сетях связано с целым рядом трудностей как технического так и психологического характера (Нестик, Михеев, 2019).

Опровержение дезинформации связано с эффектом «обратного действия» или «эффектом бумеранга». Эффект обратного действия — это когнитивное искажение, которое заставляет людей при столкновении с фактами, несоответствующими их убеждениям, отвергать и игнорировать эти факты. По сути, эффект демонстрирует, что предъявление собеседнику доказательств, которые опровергают его позицию, часто неэффективно и приводит к обратной реакции, заставляя людей доказывать свое первоначальное мнение сильнее. Установлено, что влияние эффекта обратного действия зависит от условий информирования, а также эмоций, которые вызывает информация, сообщение (Trevors, 2016).

Респонденты еще более решительно начинают оспаривать искаженные или недостоверные сведения, в которые они верят, при предъявлении им аргументов и фактов, доказывающих их ложность. Попытки опровержения фейковой информации,

основанных, в том числе на фактчекинге, то есть проверке достоверности сведений по различным базам, не имеют решающего значения в формировании общественного мнения по тем или иным событиям. Это связано с тем, что опровержения часто вытесняются из информационного поля из-за роста информации представителей политических элит и СМИ, которые продвигают менее точные, но соответствующие их формату, стилю опровержения. В результате сформированные искаженные представления сохраняются в обществе долгое время.

По мнению многих исследователей, одним из заблуждений является убежденность в необходимости предоставления как можно большего числа достоверных сведений. Но если такая информация не соответствует убеждениям людей, то она будет вызывать обратную реакцию, эффект «обратного действия» (Kunda, 1990; Kahan, 2017). У людей будет формироваться стойкое неприятие к опровержениям, что будет способствовать к большему распространению дезинформации, увеличению числа людей с искаженным представлениями и убеждениями.

Эффект «консенсуса экспертов» заключается в том, что при обсуждении проблем в контексте научного консенсуса, то есть коллективного решения в определенной области науки, в среде друзей и родственников онлайн активизируется обратная связь, повышается рефлексия, влияющая на формирование убеждений пользователей. Это связано с тем, что пользователи более восприимчивы к информации, которая поступает от близких людей, чем от кого-то бы то ни было (Linden, et al. 2021).

Эффект эмоционального заражения также может влиять на восприятие опровержения. В январе 2012 г. исследователи из Корнельского университета провели эксперимент среди 689 003 пользователей сети Facebook. Для изучения эмоционального заражения они манипулировали доступностью контента пользователям: те, кому были более доступны негативные сообщения друзей, сами чаще писали негативно окрашенные тексты (Kramer et al., 2014). Поскольку интернет-пользователи не всегда в полной мере

отдают себе отчет в том, что являются участниками эксперимента, использование этого метода остается дискуссионным (Grimmelmann, 2015).

Устойчивость воздействия дезинформации, как уже было отмечено, связана с тем, что пользователи остаются нечувствительными к опровержениям, если вера в них заставляет изменить уже сложившиеся взгляды (Ecker et al., 2014). Кроме того, уровень влияния эффекта дезинформации на личность может иметь длительный характер (Johnson et al., 1994), т.к. забытая или устаревшая информация остается в памяти, несмотря на все попытки обновить ее (Ecker et al., 2010). При этом если забытая или устаревшая информация всплывает в памяти, то она воспринимается, как достоверная и не вызывает сомнений (Ecker et al., 1993; Lewandowsky et al., 2012).

Одним из методов повышения эффективности опровержений ложных сообщений в социальных сетях может быть пометка информации криптографическими маркерами, а также специальными тегами или ярлыками, например, «вызывает доверие», «официальная информация». Ряд исследований подтверждают эффективность этого метода. В частности, в ходе одного из онлайн-экспериментов, где новостям о вакцинах в Твиттере присваивались ярлыки достоверности, было зафиксировано повышение уровня доверия к вакцинам (Featherstone et al., 2019), а также повышение уровня доверия к самим новостям (Gao et al., 2018).

Важно учитывать, что в опровержении дезинформации важную роль играют субъекты опровержения и совершенствование алгоритмов, которые создают связанные сообщения в социальных сетях. Ряд исследований показывают, что эффективными субъектами опровержения дезинформации об эпидемиях могут выступать, например органы здравоохранения (Bode, Vraga, 2018).

Кроме того, необходимо при опровержении сосредоточиться на ослаблении связей между формированием групповой идентичности и фейками, а также уменьшением потока подкрепляющей эти фейки информации от представителей элит и СМИ. По мнению ряда экспертов, это потребует перехода от стратегии предоставления информации из

одного официального источника (СМИ, органов власти) к стратегии, учитывающей роль посредников в формировании и поддержании убеждений, взглядов и смыслов (Nyhan, 2021).

Повышение эффективности опровержений ложных сообщений в социальных сетях тесно связано с другими способами противодействия фейкам, основанных на комбинировании взаимодействия «человек-компьютер-искусственный интеллект». В частности, в международной практике используются следующие основные подходы по противодействию ложной информации: гибридный, функциональный, сетевой и основанный на знаниях.

В гибридном подходе использованы методы проверки человеком и метод машинного обучения (Yang et al., 2019). Функциональный подход предполагает использование функций, связанных с учетными записями пользователей, содержанием их сообщений и контекстный метод (Gilani et al., 2017). Сетевой метод включает методы обнаружения фейков на ранних стадиях, пометку их и блокировку (Aldwairi, Alwahedi, 2018). Подход, основанный на знаниях, предполагает использования искусственного интеллекта и экспертных оценок (Baly, 2018).

При опровержении фейков важное значение может иметь источник сообщения. Ряд исследований показывают, что сообщения, которые передаются из источников, которые считаются заслуживающими доверия, с большей вероятностью будут переданы другим людям (Buchanan and Benson, 2019). Например, в Англии есть примеры того, что во время пандемии COVID-19 увеличивался трафик в традиционно пользующиеся доверием СМИ, таким как BBC, NHS и WOS. Многими исследователями также указывается на большое влияние при распространении информации (дезинформации или опровержений) заявлений, сделанных известными деятелями культуры, шоу-бизнеса и политиками, у которых большая аудитория последователей, поклонников (Allington and Dhavan, 2021).

В маркетинговых исследованиях также рассматривается доверие к позитивным или негативным высказываниям, сделанное потенциальным, действующим или бывшим потребителем, посредством Интернет в ходе коммуникации из «уст в уста» (electronic word-of-mouth communication (eWOM)). Проблематика скептического отношения к коммуникации из «уст в уста», ставшей в последние годы важным фактором, влияющим на восприятие и поведение пользователей Интернет, связана с появлением поддельной (фальшивой) коммуникации eWOM, в том числе осуществляемой ботами. Доверие пользователей к опровержению в этом контексте зависит от воспринимаемых мотивов отправителя сообщения и идентификационных признаков отправителя (Zhang et al., 2016).

Многие инициативы в области противодействия дезинформации в социальных сетях, носят просветительский характер и направлены на формирование негативного отношения к дезинформации и повышение бдительности интернет-пользователей в отношении фейковых сообщений. Однако, до сих пор остается не ясным, в какой степени отношение к дезинформации как к социальному явлению влияет на способность к распознаванию фейковых сообщений и на уровень доверия к опровержениям дезинформации.

Одним из способов опровержения может быть его «маскировка» под конспирологическую теорию. Известно, что конспирологическая информация, как и обычная информация, распространяются одинаковыми путями (Mosanu, 2015), но при этом пользователи более предрасположены к восприятию именно конспирологических новостей (Bessi, 2015). В этой связи можно предположить о более высоком уровне эффективности опровержения, представленного как конспирологические слухи. Данное предположение требует проведения дополнительных исследований.

Проведенные ранее исследования показывают, что наиболее эффективными тактиками опровержения являются 1) факт-чекинг (ссылка на специальные ресурсы, занимающиеся проверкой достоверности сообщений), 2) объяснение происхождения

фейков, в том числе того, кому они выгоды; 3) апелляции к компетентности, то есть приведения точки зрения лица или организации, вызывающих доверие и считающихся компетентными в данном вопросе. Несколько менее эффективны, но также действенны 1) указание на консенсус экспертов по поводу данной проблемы; а также 2) предупреждение о ложном или сомнительном характере новости (Walter, Murphy, 2018). Однако, до сих пор нет ясности в отношении того, как психологическая эффективность опровержения дезинформации зависит от степени персонализации источника опровержения, конспирологичности опровергающего сообщения и его эмоциональности. Проведенное нами исследование позволяет уточнить роль этих факторов в противодействии фейкам.

## ЭМПИРИЧЕСКОЕ ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ ПРОТИВОДЕЙСТВИЯ ФЕЙКОВОЙ ИНФОРМАЦИИ В ЗАВИСИМОСТИ ОТ УРОВНЯ ЭМОЦИОНАЛЬНОСТИ И КОНСПИРОЛОГИЧНОСТИ ОПРОВЕРЖЕНИЙ

Для изучения факторов устойчивости пользователей социальных сетей к воздействию фейков, а также эффективности методов противодействия дезинформации в социальных сетях нами в январе 2022 г. был проведен массовый опрос пользователей социальных сетей в виде онлайн-анкетирования с использованием панели Анкетолог.ру среди пользователей, имеющих аккаунт в одной или нескольких социальных сетях: Вконтакте, Одноклассники, LiveJournal, Facebook, Twitter, Мой Мир, LinkedIn, Instagram, TikTok, Telegram (N = 3007; 49% - мужчины; 51% - женщины, Mage = 41,9; SD=10,8; ), в том числе 1208 респондентов, проявляющих интерес к фейковым сообщениям (N = 1208, 51% - мужчины; 49% - женщины; Mage = 42; SD=10,7).

Для выявления респондентов, проявляющих интерес к фейковым сообщениям, нами использовались: разработанная нами шкала готовности к перепосту сомнительных новостных сообщений в социальных сетях (4, альфа Кронбаха = 0,801), опросник "Подверженность воздействию дезинформации в социальных сетях" (Т.А.Нестик, Е.А. Михеев), измеряющий доверие к новостям в социальных сетях (альфа Кронбаха =

0,785) и готовность делать репосты и комментировать сообщения в социальных сетях (альфа Кронбаха = 0,825); шкала способности к распознаванию ложных сообщений (5, альфа Кронбаха = 0,605). Кроме того, для измерения отношения респондентов к дезинформации в социальных сетях в инструментарий были включены дополнительные шкалы: шкала готовности к критическому и некритическому распространению информации о пандемии (Солдатова и др., 2021). Также использовалась субшкала вовлеченности в использование социальных сетей из опросника "Вовлеченность в использование ИКТ" (Татарко и др., 2020). Для измерения социально-психологических характеристик пользователей социальных сетей использовался ряд методик, в том числе шкалы социального доверия из опросника World Values Survey (Crepaz et al., 2014); шкалы доверия к ученым (3;  $\alpha$  Кронбаха=0,719) и доверия к науке (Hartman et al., 2017;  $\alpha$  Кронбаха=0,873) и опросник конспирологического мировоззрения (Bruder et al., 2013).

Для измерения отношения интернет-пользователей к дезинформации как социальному явлению в рамках данного исследования нами был разработан и апробирован опросник «Отношение личности к дезинформации в социальных сетях» (Т.А. Нестик, Е.А. Михеев), измеряющий значимость угрозы дезинформации в социальных сетях, веру в преднамеренность распространения ложных сообщений, убеждение в необходимости сохранения свободы слова в социальных сетях, готовность защищать себя и других от дезинформации в социальных сетях. Испытуемым предлагается оценить степень своего согласия с 14 утверждениями со следующей инструкцией: «В социальных сетях сегодня много фейковых новостей — вымышленных, ложных слухов, непроверенной информации, которая кажется правдой, но потом не подтверждается. Следующие вопросы - о том, как Вы к этому относитесь. Пожалуйста, оцените степень своего согласия или несогласия со следующими утверждениями по шкале от 1 — совершенно не согласен до 5 — совершенно согласен». Конфирматорный факторный анализ, проведенный в программе Jamovi, подтвердил

четырёхфакторную структуру методики со связями между ошибками пп. 1 и 9, 4 и 5, 5 и 6, 7 и 11, 7 и 13, 7 и 14 ( $N=3007$ ; хи-квadrat= $251,496$ ;  $DF=58$ ;  $p<0,001$ ;  $CFI=0,979$ ;  $TLI=0,967$ ;  $SRMR=0,035$ ;  $RMSEA=0,039$ ; 90% CI [ $0,034$ ;  $0,044$ ]). Вклад всех пунктов в факторы значим на уровне  $p<0,001$ . Внутренняя согласованность всех шкал удовлетворительная: для шкалы «Значимость угрозы дезинформации в социальных сетях» омега Макдональда =  $0,750$ ;  $M=3,52$ ;  $SD=0,939$ ; для шкалы «Убеждение в преднамеренности распространения ложных сообщений» омега Макдональда =  $0,810$ ;  $M=3,64$ ;  $SD=0,886$ ; для шкалы «Убеждение в необходимости сохранения свободы слова в социальных сетях» омега Макдональда =  $0,711$ ;  $M=3,33$ ;  $SD=0,895$ ; для шкалы «Убеждение в преднамеренности распространения ложных сообщений» омега Макдональда =  $0,613$ ;  $M=3,23$ ;  $SD=0,729$ . Показатели куртозиса пунктов находятся в диапазоне от  $-0,998$  до  $0,178$ , а показатели скошенности распределения — в пределах от  $0,751$  до  $0,259$ , при том, что значения в интервале  $\pm 1$  считаются превосходными, а в интервале  $\pm 2$  — допустимыми (George, Mallery, 2011). Для оценки конвергентной валидности мы прокоррелировали показатели субшкал методики с шкалой готовности к критическому и некритическому распространению информации о пандемии (Солдатова и др., 2021). Как и ожидалось, нами была выявлена прямая связь готовности к распространению сообщений после критического анализа с значимостью угрозы дезинформации в социальных сетях ( $r=0,260^{**}$ ), убеждением в преднамеренности распространения ложных сообщений ( $r=0,146^{**}$ ) и готовностью защищать себя и других от дезинформации в социальных сетях ( $r=0,324^{**}$ ). Текст методики приведен в Приложении 1.

***Связь воспринимаемой достоверности опровержений фейковых сообщений с отношением интернет-пользователей к дезинформации как социальному явлению.***

Нами было выдвинуто предположение о том, что оценка интернет-пользователями достоверности опровержений фейковой информации связана с их отношением к дезинформации как социальному явлению. В качестве показателя воспринимаемой

достоверности опровержений фейковой информации нами были использованы средние значения оценки достоверности опровержений, предъявленных нами участникам экспериментальных групп (N=955), при этом оценки участников контрольных групп не учитывались (подробнее см. ниже в табл. 2-4). Как показали результаты корреляционного анализа с использованием коэффициента Спирмана, доверие к сообщениям, опровергающим дезинформацию в социальных сетях, оказалось прямо связанным с значимостью угрозы дезинформации в социальных сетях ( $r=0,101^{**}$ ), убеждением в преднамеренности распространения ложных сообщений ( $r=0,099^{**}$ ) и готовностью защищать себя и других от дезинформации в социальных сетях ( $r=0,096^{**}$ ). Способность распознавать ложные сообщения в социальной сети, измерявшаяся нами при предъявлении испытуемым 5 фейковых новостей, отрицательно связана с убеждением в необходимости сохранения свободы слова в социальных сетях ( $r=-0,206^{**}$ ).

Таблица 1

**Связь воспринимаемой достоверности опровержений фейковых сообщений и способности к распознаванию ложных сообщений с отношением интернет-пользователей к дезинформации**

	Доверие к сообщениям, опровергающим дезинформацию в социальных сетях	Способность распознавать ложные сообщения в социальной сети	Значимость угрозы дезинформации в социальных сетях	Убеждение в преднамеренности распространения ложных сообщений	Убеждение в необходимости сохранения свободы слова в социальных сетях	Готовность защищать себя и других от дезинформации в социальных сетях
Значимость угрозы дезинформации в социальных сетях	,101**	,091**	1	,592**	-,400**	,561**
Убеждение в преднамеренности распространения ложных сообщений	,099**	,096**	,592**	1	-,243**	,425**

Убеждение в необходимости сохранения свободы слова в социальных сетях	0,027	-,206**	-,400**	-,243**	1	-,163**
Готовность защищать себя и других от дезинформации в социальных сетях	,096**	,068*	,561**	,425**	-,163**	1

Примечание: \*\* -  $p < 0,01$ ; \*  $p < 0,05$ .

В целом полученные нами результаты указывают на то, что характеристики отношения интернет-пользователей к дезинформации как социальному явлению и тем или иным способам борьбы с ней, действительно, связаны с доверием к опровержениям и способностью к распознаванию фейков. Однако, линейный регрессионный анализ показывает, что вклад отношения к дезинформации в данные показатели крайне незначительный (соответственно,  $R^2=0,017$  и  $R^2=0,055$ ). Для уточнения роли отношения личности к дезинформации в борьбе с фейковыми новостями в социальных сетях необходимы дальнейшие исследования.

**Факторы эффективности опровержения фейковых новостей.** Нами были выдвинуты гипотезы о влиянии на эффективность опровержения фейковых новостей трех параметров: 1) степени персонализации источника (предполагалось, что опровержение более эффективно, если оно исходит от обезличенной организации, а не конкретного лица или сообщества людей); 2) степени конспирологичности опровержения (предполагалось, что опровержение более эффективно, если оно облечено в форму конспирологической теории); 3) позитивной, нейтральной или негативной эмоциональной окраски (было выдвинуто предположение о том, что опровержение более эффективно, если оно носит позитивную окраску, и менее эффективно, если оно вызывает страх).

Для проверки гипотез использовался онлайн-эксперимент с А/Б-тестированием, при котором случайно отобранным испытуемым предъявлялись разные варианты опровержений, предварительно оцененных экспертами. Всех испытуемых мы просили прочитать новость на тему коронавирусной инфекции и оценить ее достоверность, а затем, после предъявления опровержения, - еще раз оценить, насколько она кажется достоверной при повторном прочтении. Наличие статистически значимых различий в оценке новости до и после опровержения являлось критерием подтверждения или опровержения гипотезы. Все эксперименты проводились с контрольной группой, которой вместо опровержения предъявлялись ложные или достоверные сообщения на другую тему, не связанную с первоначальной новостью.

Для проверки гипотезы №1 испытуемым предлагалось оценить вероятность правдивости новости «Курение и коронавирус. Никотин блокирует передачу COVID-19, тем самым помогает защититься от инфекции. Источник: сайт «MD Yurtaev». Первой группе испытуемых (N=94, 53% - мужчины, 47% - женщины) затем предлагалось прочитать и оценить правдивость опровержения от имени неформального сообщества: «Врачи из сообщества «Вбросам.нет» предупредили, что курильщики чаще страдают от инфекций из-за удара по иммунитету, кроме того, у пациентов с хронической обструктивной болезнью лёгких вырабатывается избыточное количество белка, через который COVID-19 проникает в клетки. Источник: газета «Московские новости». Второй группе (N=105, 44% - мужчины, 57% - женщины) предлагалось опровержение от государственной организации: «В Министерстве здравоохранения РФ опровергли устойчивость курильщиков к коронавирусу. В ведомстве напомнили, что употребление табачной продукции в ситуации с пандемией смертельно. Источник: сайт «РИАМО». Третьей группе (N=72, 59% - мужчины, 41% - женщины) предлагалось опровержение от имени конкретного эксперта: «О связи между сигаретами и устойчивостью к COVID-19 говорят не врачи, а блогеры и журналисты альтернативных СМИ. Доверять этой информации не стоит. Курение повышает риск возникновения осложнений при

заболевании ковидом», - говорит врач-пульмонолог высшей категории Игнатов И.В. Источник: сайт РБК». Наконец, контрольной группе (N=94, 40% - мужчины, 60% - женщины) предлагалось прочитать и оценить ложную новость на другую тему: «В России вернут смертную казнь. Источник: социальная сеть "Одноклассники"».

Результаты серии экспериментов свидетельствуют в пользу гипотезы о том, что опровержение из деперсонализированных официальных источников наиболее эффективно ( $p=0,05$ ). Опровержение от неформального сообщества и персонализированное опровержение от конкретного эксперта не оказали значимого эффекта ( $p>0,05$ ).

Таблица 2

### Зависимость эффективности опровержений от степени их персонализации

	До опровержения		После опровержения		Значимость различий	
	M1	SD1	M2	SD2	Z	p
Опровержение от неформального сообщества (Врачи из сообщества «Вбросам.нет»...; N=288)	2,43	1,73	2,40	1,70	-1,025	0,306
Опровержение от государственной организации ("В Министерстве здравоохранения РФ..."; N=312)	2,77	1,92	2,68	1,85	-1,991	0,053
Опровержение от эксперта ("Врач-пульмонолог высшей категории Игнатов И.В."; N=306)	2,44	1,73	2,40	1,75	-0,614	0,539
Контрольная группа (фейковая новость на другую тему: "В России вернут смертную казнь"; N=302)	2,40	1,79	2,32	1,67	-1,227	0,22

Для проверки второй гипотезы участникам эксперимента предлагалось оценить вероятность правдивости новости «Эксперт: Только своевременная профилактика и лечение с использованием антибактериальных лекарственных препаратов помогают в борьбе с коронавирусной инфекцией. Источник: сайт «Тайны вселенной». При этом первой группе испытуемых (N=113, 55% - мужчины, 45% - женщины) затем предъявлялось официальное опровержение: «Применение антибиотиков при коронавирусной инфекции не только не защитит от болезни, но и может привести к

серьезным последствиям — проблемам с желудочно-кишечным трактом, спазмам бронхов и аритмии. Кроме того, организм, привыкая к данным препаратам, теряет иммунитет к инфекциям, из-за чего люди могут начать погибать даже от простых ОРВИ. Источник: сайт Росздрава (Федерального агентства по здравоохранению и социальному развитию)». Второй экспериментальной группе (N=132, 38% - мужчины, 62% - женщины) предъявлялось конспирологическое опровержение: «Значительная часть смертей во время «испанского гриппа» 1918 года была вызвана передозировкой модного тогда аспирина, а не самим вирусом. Информированные источники в медицинских кругах говорят, что за мифом о пользе аспирина, анальгина и антибиотиков при заболевании COVID-19 стоят чьи-то коммерческие интересы. Известны случаи, когда антибиотики отпускались без рецепта, если клиент покупал еще и пробиотики для восстановления микрофлоры после "курса лечения". Не отдавайте свои деньги аптечным сетям, не вредите себе. Источник: телеграм-канал «Совершенно секретно». Контрольной группе (N=120, 54% - мужчины, 46% - женщины) после первой новости предъявлялась достоверная новость на другую тему: «В России планируют сделать предновогоднюю пенсионную выплату ежегодной. Источник: МК.RU».

Результаты серии экспериментов свидетельствуют в пользу гипотезы о том, что для пользователей, проявляющих интерес к фейковым новостям, опровержение, облеченное в форму конспирологического предупреждения, является более эффективным, чем опровержение от официальной организации (соответственно,  $\rho=0,05$  и  $\rho=0,08$ ).

Таблица 3

**Зависимость эффективности опровержений от степени их конспирологичности**

<b>Новость 2. Только своевременная профилактика и лечение с использованием антибактериальных лекарственных препаратов помогают в борьбе с коронавирусной инфекцией. Источник: сайт «Тайны вселенной».</b>						
	До опровержения		После опровержения		Значимость различий	
	M1	SD1	M2	SD2	Z	p
Официальное опровержение (сайт Росздрава; N=398)	3,96	1,92	3,85	1,90	-1,762	0,079
Конспирологическое	4,13	1,92	4,01	1,85	-2,253	0,055

опровержение (телеграм-канал «Совершенно секретно»; N=401)						
Контрольная группа (достоверная новость на другую тему: "В России планируют сделать предновогоднюю пенсионную выплату ежегодной". Источник: МК.RU; N=1012)	3,99	2,00	3,91	2,02	-1,175	0,24

Для проверки третьей гипотезы испытуемым предлагалось оценить вероятность правдивости новости «COVID-19 не опаснее обычной пневмонии». Источник: ИА «Примавера». Затем первой экспериментальной группе (N=70, 42% - мужчины, 58% - женщины) предъявлялось юмористическое опровержение: "Иных уж нет, а те далече — Они любили covid-встречи". После пандемии останутся только глупые. Потому что умные знают, что коронавирус — это все ерунда, и они не заболеют. Источник: телеграм-канал "Красная зона". Второй группе (N=80, 45% - мужчины, 55% - женщины) предъявлялось пугающее опровержение: «Вид легких шокировал»: вскрытие умерших от COVID показало, что это не пневмония. Патологоанатом опроверг слух о безобидности вируса и призвал к принятию жестких мер в отношении ковид-диссидентов. Источник: Телеканал «НТВ». Третьей экспериментальной группе (N=76, 49% - мужчины, 51% - женщины) предъявлялось эмоционально нейтральное опровержение: «Установлено, что при пневмонии, вызванной COVID-19, причиной воспаления является вирус. Он не ограничивается отдельными участками, как при бактериальной пневмонии. Воспаление распространяется, захватывая оба лёгких. Источник: официальный сайт Российской академии наук». Наконец, контрольной группе (N=73, 49% - мужчины, 51% - женщины) предъявлялось достоверное сообщение на другую тему: «Москва и Санкт-Петербург оказались самыми тревожными регионами России в 2021 году. Источник: РИА Новости».

Проведенные нами эксперименты свидетельствуют в пользу гипотезы о том, что позитивно (юмористически) окрашенное опровержение более эффективно, чем

нейтрально окрашенное (соответственно,  $\rho=0,05$  и  $\rho=0,06$ ). При этом негативно окрашенное (пугающее) опровержение не оказало статистически значимого эффекта ( $\rho=0,205$ ).

Таблица 4

**Зависимость эффективности опровержений от их позитивной, нейтральной или негативной эмоциональной окраски**

Новость 3. "COVID-19 не опаснее обычной пневмонии". Источник: ИА «Примавера».						
	До опровержения		После опровержения		Значимость различий	
	M1	SD1	M2	SD2	Z	$\rho$
Юмористическое опровержение (сетевая шутка на тему ковидскептицизма; N=244)	3,31	1,96	3,17	1,95	-2,135	0,048
Пугающее опровержение («Вид легких шокировал»...; N=242)	3,45	1,87	3,38	1,86	-1,203	0,229
Эмоционально нейтральное опровержение (сайт РАН; N=248)	3,42	2,02	3,28	1,96	-1,894	0,062
Контрольная группа: достоверное сообщение на другую тему (N=230)	3,59	2,00	3,43	1,93	-1,666	0,096

Известно, что если пользователи сталкиваются с опровержением, которое противоречит их картине мира и системе ценностей, то такое опровержение еще больше усиливает веру в ранее прочитанную ложную новость (Nyhan, Reifler, 2010; Wittenberg, Berinsky, 2020). Вместе с тем, если опровержение поддерживает высокую самооценку адресата, то оно вызывает меньшее сопротивление (Nyhan, Reifler, 2019). Поэтому при составлении опровержения ложного сообщения важно убедиться, что формулировка не оскорбительна для читателя, уже поверившего первоначальной ложной новости. Опровержение должно поддерживать позитивную групповую идентичность и соответствовать мировоззрению адресата (Ecker et al., 2022). Например, опровергая утверждение об опасности прививок против коронавируса, можно начать со слов: «Безопасность вакцин — важное условие готовности людей делать прививки...». Учитывая выявленные нами особенности мировоззрения пользователей, проявляющих интерес к фейкам, в опровержениях на актуальные темы социально значимых

происшествий, социальной защиты, образования, уровня благосостояния граждан и политики можно подчеркивать важность ответственности государства перед народом, заботы о людях, необходимость контроля за элитами, защиты общества от внешних и внутренних угроз. Объясняя происхождение фейков, можно ссылаться на борьбу тех или иных сил, имеющих политические и экономические интересы.

Как показывают полученные нами данные, подавляющее большинство пользователей социальных сетей, проявляющих интерес к фейкам, характеризуются выраженной верой в конспирологические теории ( $p < 0,001$ ). Известно, что конспирологическая информация распространяется теми же путями, что и обычная информация, но при этом пользователи более предрасположены к восприятию именно конспирологических новостей (Bessi, 2015). Исследования также показывают, что вместо подчеркивания надежности источника опровержения следует дискредитировать надежность источника самих слухов (Walter, Tukachinsky, 2020). Поэтому опровержение фейков может быть само облечено в форму конспирологической теории. Вместе с тем, такого рода опровержения должны завершаться ссылками на достоверные источники, чтобы избежать формирования эффекта «выученной информационной беспомощности», когда пользователи уже не верят ни одному каналу информации и не считают себя способными разобраться, где правда, а где ложь.

Исследование показало, что для пользователей социальных сетей, интересующихся фейками, характерно признание отсутствия консенсуса между экспертами, а также низкое доверие к науке как социальному институту ( $p < 0,001$ ). Значительная часть среди представителей данной группы пользователей отметили известных ученых в качестве предпочитаемого источника получения опровержений. Это может объясняться стремлением разрешить противоречия в информации через обращение к наиболее компетентным экспертам на фоне низкого институционального доверия. Вместе с тем результаты эксперимента показали, что разоблачение от имени официальной обезличенной организации более действенно, чем от имени конкретного, но не известного

общественности эксперта. Это значит, что опровержение должно исходить от всем известной федеральной организации или от известных ученых, представляющих официальную позицию научного учреждения.

Как показали результаты проведенных нами экспериментов, позитивно (юмористически) окрашенное опровержение более эффективно, чем нейтрально окрашенное и негативно окрашенное (пугающее). Это хорошо согласуется с проведенными ранее исследованиями, показавшими, что при нагнетании тревоги включаются защитные психологические механизмы, приводящие к недооценке угрозы и закреплению первоначальных заблуждений (Witte, Allen, 2000). Опровержения более эффективны, если они провоцируют юмор или сопереживание другим людям, и менее эффективны, если они внушают страх.

Следует признать наличие целого ряда *ограничений*, которые нужно учитывать при использовании результатов проведенного нами исследования. Во-первых, интерес к фейковым сообщениям определялся нами на основании оценки интернет-пользователями своей готовности пересылать те или иные предъявляемые им непроверенные сообщения. Безусловно, такая оценка может быть неточной и должна в будущих исследованиях дополняться результатами анализа реальной активности пользователей, в том числе наличием подписок на телеграм-каналы, распространяющие фейковые сообщения. Во-вторых, случайно отобранные выборки испытуемых для участия в экспериментах посредством автоматизированной процедуры А/Б-тестирования не были сбалансированы по полу и возрасту. Очевидно, что вклад демографических факторов следует контролировать в последующих экспериментальных исследованиях. В-третьих, предъявляемые испытуемым сообщения были связаны преимущественно с пандемией COVID-19 и ее последствиями. Между тем, эффективность опровержения фейковых новостей может зависеть от тематики сообщений, а также от психологического состояния всего общества в условиях глобальных кризисов (Нестик и др., 2020).

Опираясь на анализ литературы и полученные нами эмпирические результаты, можно выделить несколько *перспективных направлений дальнейших исследований*. К таким направлениям относится, на наш взгляд, исследование вклада индивидуальных и коллективных переживаний на доверие к дезинформации и эффективность ее опровержения. Уязвимость к воздействию фейков изучалась до сих пор преимущественно в связи с индивидуальными эмоциональными состояниями (тревогой, депрессией, стрессом), однако мы все еще очень мало знаем о том, как на действенность опровержений влияют политические эмоции (Redlawsk, Mattes, 2022) и связанные с групповой идентичностью коллективные переживания, такие как тревога за будущее своей страны. Еще одним перспективным направлением исследований является изучение психологических предпосылок противодействия фейкам в условиях выученной информационной беспомощности (Nisbet, Kamenchuk, 2021). Крайне мало изученным остается влияние на уязвимость к дезинформации и успешность противодействия ей со стороны таких факторов, как социальный пессимизм и краткосрочная ориентация. Наконец, все более актуальным становится изучение психологических факторов эффективности опровержения дипфейков, основанных на синтезе человеческих изображений с использованием технологий искусственного интеллекта (Köbis et al., 2021). Подводя итоги, хотелось бы выразить надежду на то, что исследования противодействия ложным новостным сообщениям будет в ближайшие годы будут в большей степени опираться на инициативы гражданского общества, научное волонтерство и самоорганизацию пользователей социальных сетей.

## ЛИТЕРАТУРА

- Нестик Т. А., Михеев Е. А.* Информационные войны с использованием систем искусственного интеллекта: анализ психологических механизмов воздействия // Институт психологии Российской академии наук. Организационная психология и психология труда. 2019. Т. 4. № 4. С. 148-174.
- Нестик Т.А., Дейнека О.С., Максименко А.А.* Социально-психологические предпосылки веры в конспирологические теории происхождения COVID-19 и

вовлеченность в сетевые коммуникации // Социальная психология и общество. 2020. Том 11. № 4. С. 87–104. DOI: 10.17759/sps.2020110407.

*Нестик Т.А., Селезнева А.В., Шестопап Е.Б., Юревич А.В.* Проблема психологического состояния общества и политических процессов в современной России // Вопросы психологии. 2021. Т. 67. № 5. С. 3-14.

*Солдатова Г. У., Рассказова Е.И., Неякина Ю.Ю., Ширяева О. С.* Потребность в информации и отношение к цифровым технологиям как факторы критичного и некритичного распространения новостей о пандемии // Вестник Московского Университета. Серия 14. Психология. 2021. № 1. С.170-195. DOI: 10.11621/vsp.2021.01.07

*Татарко А.Н., Макласова Е.В., Лепшокова Э.Х., Галяпина В.Н., Ефремова М.В., Дубров Д.И., Бульцева М.А., Бушина Е.В., Миронова А.А.* Методика оценки вовлеченности в использование информационно-коммуникационных технологий // Социальная психология и общество. 2020. Т. 11. № 1. С. 159—179. DOI:10.17759/sps.2020110110

*Aldwairi M., Alwahedi A.* Detecting fake news in social media networks // Procedia Computer Science. 2018. V.141. P.215–222. DOI: 10.1016/j.procs.2018.10.171

*Allington, D., Duffy, B., Wessely, S., Dhavan, N., & Rubin, J.* Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency // Psychological Medicine. 2021. V.51. T.10. P.1763–1769. DOI:10.1017/S0033291721000593

*Baly R., Karadzhov G., Alexandrov D., Glass J., Nakov P.* Predicting Factuality of Reporting and Bias of News Media Sources // Factuality of Reporting and Bias of News Media Sources. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Belgium: Association for Computational Linguistics, 2018. P. 3528–3539. DOI:10.18653/v1/D18-1389

*Bessi A., Coletto M., Davidescu G., Scala A., Caldarelli G., Quattrociocchi W.* Science vs Conspiracy: Collective Narratives in the Age of Misinformation // PLoS ONE. 2015. V. 10. №2. DOI: 10.1371/journal.pone.0118093. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0118093> (Accessed: 27.06.2022).

*Bode L., Vraga E. K.* See something, say something: Correction of global health misinformation on social media // Health Communication. 2018. V.33. №9. P. 1131–1140. DOI:10.1080/10410236.2017.1331312

*Buchanan T., Vladlena B.* Spreading Disinformation on Facebook: Do Trust in Message Source, Risk Propensity, or Personality Affect the Organic Reach of ‘Fake News’? // Social Media and Society. 2019. V. 5. №4. DOI:10.1177/2056305119888654

- Chan M.P.S., Jones C.R., Hall J.K., Albarracín D.* Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation // *Psychological Science*. 2017. V. 28. P. 1531–1546. DOI:10.1177/0956797617714579
- Ecker U.K.H., Lewandowsky S., Cook J.* The psychological drivers of misinformation belief and its resistance to correction// *Nature Reviews Psychology*. 2022. № 1. P. 13–29. DOI: 10.1038/s44159-021-00006-y.
- Featherstone J. D., Davis C. A., Zhang J.* Correcting vaccine misinformation on social media using fact-checking labels// In APHA's 2019 annual meeting and expo. 2019. Nov. 2–Nov. 6.
- Gao M., Xiao Z., Karahalios K., Fu W.-T.* To label or not to label: The effect of stance and credibility labels on readers' selection and perception of news articles // *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW). 2018. P.1–16. DOI:10.1145/3274324
- George M., Mallery P.* Descriptive Statistics. // *SPSS for Windows step by step. A simple guide and reference 18.0 update/ S. Hartman (Ed.)*. Boston: Pearson, 2011. P. 95–104.
- Gilani Z., Kochmar E., Crowcroft J.* Classification of twitter accounts into automated agents and human users. In: *Proceedings of the 2017 IEEE//ACM International Conference on Advances in Social Networks Analysis and Mining 2017*. ACM. 2017. P. 489–496. DOI:10.1145/3110025.3110091
- Grimmelmann J.* The Law and Ethics of Experiments on Social Media Users // *Colorado Technology Law Journal*. 2015. V. 13. P. 219-271. DOI:10.31228/osf.io/cdt7y
- Kahan D., Peters E., Dawson E., Slovic P.* Motivated numeracy and enlightened self-government // *Behavioural Public Policy*. 2017. № 1. P. 54–86. DOI:10.1017/bpp.2016.2
- Köbis N. C., Bonnefon J.-F., Rahwan I.* Bad machines corrupt good morals. *Nature Human Behaviour*. 2021. V.5. №6. P.679–685. DOI:10.1038/s41562-021-01128-2
- Kramer A.D., Guillory J.E., Hancock J.T.* Experimental evidence of massive-scale emotional contagion through social networks // *Proceedings of the National Academy of Sciences*. 2014. V. 111. №24. P. 8780-8790. DOI:10.1073/pnas.1320040111
- Kuran T., Sunstein C. R.* Availability Cascades and Risk Regulation // *Stanford Law Review*.1999. 51(4), 683–768. DOI: 10.2307/1229439.
- Mocanu D., Rossi L., Zhang Q., Karsai M., Quattrociocchi W.* Collective attention in the age of (mis)information // *Computers in Human Behavior*. 2015. Vol. 51, Part B, P. 1198-1204. DOI:10.1016/j.chb.2015.01.024

- Nisbet E.C., Kamenchuk O.* Russian News Media, Digital Media, Informational Learned Helplessness, and Belief in COVID-19 Misinformation, *International//Journal of Public Opinion Research*. 2021.V. 33. P. 571–590. DOI:10.1093/ijpor/edab011.
- Nyhan B.* Why the backfire effect does not explain the durability of political misperceptions // *Proceedings of the National Academy of Sciences*. 2021. V.118. №15. DOI: 10.1073/pnas.1912440117
- Nyhan B., Reifler J.* When corrections fail: The persistence of political misperceptions // *Political Behavior*. 2010. V. 32. №2. P.303–330. DOI:10.1007/s11109-010-9112-2
- Nyhan, B., Reifler, J.* The roles of information deficits and identity threat in the prevalence of misperceptions // *Journal of Elections, Public Opinion, and Parties*. 2019. V. 29. №2. P. 222–244. DOI:10.1080/17457289.2018.1465061
- Redlawsk D., Mattes K.* Emotions and Politics. In D. Osborne & C. Sibley (Eds.) // *The Cambridge Handbook of Political Psychology (Cambridge Handbooks in Psychology)*. Cambridge: Cambridge University Press. 2022. P. 139-158. DOI:10.1017/9781108779104.010
- Trevors G. J., Muis K. R., Pekrun R., Sinatra G. M., Winne P.H.* Identity and Epistemic Emotions During Knowledge Revision: A Potential Account for the Backfire Effect // *Discourse Processes*. 2016. V. 53. P.5-6, P. 339–370. DOI: 10.1080/0163853X.2015.1136507
- Walter N., Murphy S. T.* How to unring the bell: A meta-analytic approach to correction of misinformation // *Communication Monographs*. 2018. V.85. №3. P. 423–441. DOI:10.1080/03637751.2018.1467564
- Walter N., Tukachinsky R.* A Meta-Analytic Examination of the Continued Influence of Misinformation in the Face of Correction: How Powerful Is It, Why Does It Happen, and How to Stop It? // *Communication Research*. 2020. V.47. №2. P. 155-177. DOI:10.1177/0093650219854600
- Witte K., Allen M.* A Meta-Analysis of Fear Appeals: Implications for Effective Public Health Campaigns // *Health Education & Behavior*. 2000. V.27. №5. P.591–615. DOI:10.1177/109019810002700506
- Wittenberg C., Berinsky A.* Misinformation and Its Correction // *Social Media and Democracy: The State of the Field, Prospects for Reform* / N. Persily, J. Tucker (Eds.). Cambridge: Cambridge University Press, 2020. P. 163-198. DOI:10.1017/9781108890960.009
- Yang K., Niven T., Kao H.* Fake News Detection as Natural Language Inference // *WSDM '19*. February 11–15, 2019. DOI: 10.1145/1122445.1122456

Zhang X., Ko M., Carpenter D. Development of a scale to measure skepticism toward electronic word-of-mouth // Computers in Human Behavior. 2016. V. 56. P. 198-208. DOI:10.1016/j.chb.2015.11.042

## ПРИЛОЖЕНИЕ

### Опросник «Отношение личности к дезинформации в социальных сетях»

В социальных сетях сегодня много фейковых новостей — вымышленных, ложных слухов, непроверенной информации, которая кажется правдой, но потом не подтверждается. Следующие вопросы - о том, как Вы к этому относитесь. Пожалуйста, оцените степень своего согласия или несогласия со следующими утверждениями по шкале от 1 — совершенно не согласен до 5 — совершенно согласен»

	1 — совершенно не согласен(на)	2 — скорее не согласен(на)	3 — ни то, ни другое / трудно сказать	4 — скорее согласен(на)	5 — полностью согласен(на)
1. Распространение лживой информации в соцсетях представляет серьезную опасность					
2. Большинство недостоверных слухов — это целенаправленные информационные вбросы, у которых всегда есть заказчик					
3. Я легко пожертвую свободой общения в социальных сетях, чтобы защитить себя и близких от их негативного влияния					
4. Борьба с фейками в социальных сетях ведет к ограничению свободы слова					
5. Государство должно пресекать распространение недостоверной информации в социальных сетях					

6. Борьба с дезинформацией в социальных сетях — это всего лишь предлог для ужесточения контроля над рядовыми гражданами					
7. Меня очень тревожит распространение фейков в социальных сетях и мессенджерах					
8. Я стараюсь заблокировать сообщения от людей, которые пересылают мне непроверенные новости в социальных сетях или мессенджерах					
9. У россиян должно быть право свободно выражать свое мнение в социальных сетях					
10. Распространение ложных сообщений и конспирологических теорий — это результат чей-то преступной деятельности					
11. Я предпринимаю меры, чтобы обезопасить себя и своих близких от недостоверных слухов и обмана в социальных сетях					
12. Большинство фейковых новостей распространяются преднамеренно, чтобы посеять панику и настроить людей против друг друга					
13. Если я сталкиваюсь с фейковыми сообщениями, я жалуюсь на них администрации социальной сети или мессенджера					

14. Совместными усилиями интернет-пользователи могут остановить распространение дезинформации в социальных сетях					
--	--	--	--	--	--

При обработке результатов подсчитываются средние значения по каждой из четырех субшкал опросника:

- 1) значимость угрозы дезинформации в социальных сетях: 1, 5, 7;
- 2) убеждение в преднамеренности распространения ложных сообщений: 2, 10, 12;
- 3) убеждение в необходимости сохранения свободы слова в социальных сетях: 4, 6, 9;
- 4) готовность защищать себя и других от дезинформации в социальных сетях: 3, 8, 11, 13, 14.

Статья поступила в редакцию: 25.06.2022. Статья опубликована: 04.07.2022.

## PSYCHOLOGICAL FACTORS OF THE EFFECTIVENESS OF REFUTING DISINFORMATION IN SOCIAL NETWORKS<sup>2</sup>

© 2021 Т. А. Nestik \*, Е. А. Mikheev\*\*

*\* doctor of psychology, Professor of the Russian Academy of Sciences, head of the laboratory of social and economic psychology, Institute of psychology of the Russian Academy of Sciences, Moscow; E-mail: nestik@ipras.ru*

*\*\* postgraduate student, Institute of psychology of the Russian Academy of Sciences, Moscow, E-mail: mih-news@mail.ru*

The article analyzes the psychological factors of the effectiveness of counteracting the spread of fake information in social networks. Psychological mechanisms are identified that reduce the effectiveness of refuting disinformation (backward effect, etc.). Existing approaches to countering disinformation are considered. The results of testing the questionnaire "Attitude towards disinformation in social networks" are presented. It is shown that the characteristics of the attitude of Internet users to disinformation as a social phenomenon and one way or another

<sup>2</sup> The article was written with the support of the RFBR grant №18-29-22046.

to combat it are indeed associated with trust in refutations and the ability to recognize fakes. The results of a series of experiments conducted among users of social networks prone to forwarding fake messages (N = 1208) are presented in order to study the role of personalization of the source of refutation, the conspiracy of the refuting message and its emotionality in countering disinformation. It has been shown that a denial in the form of a conspiracy warning is more effective than a denial from an official organization; and a positively (humorously) colored rebuttal is more effective than a neutrally colored one. Recommendations are proposed to improve the effectiveness of refuting fakes distributed in social networks.

*Keywords:* disinformation, fakes, refutation, boomerang effect, conspiracy theories, humor.

## REFERENCES

- Nestik, T.A., & Mikheev, E.A. (2019). Informacionnye vojny s ispol'zovaniem sistem iskusstvennogo intellekta: analiz psikhologicheskikh mehanizmov vozdeystviya [Information wars using artificial intelligence systems: analysis of psychological mechanisms of influence]. *Institut Psikhologii Rossiyskoy Akademii Nauk. Organizatsionnaya Psikhologiya i Psikhologiya Truda [Institute of Psychology of the Russian Academy of Sciences. Organizational Psychology and Psychology of Labor]*, 4 (4), 148-174. (in Russian).
- Nestik, T.A., Deyneka, O.S., & Maksimenko, A.A. (2020). Sotsial'no-psikhologicheskie predposylki very v konspiologicheskie teorii proiskhozhdeniya COVID-19 i вовлеченност' v setevye kommunikatsii [Socio-Psychological Predictors of Belief in Conspiracy Theories of the Origin of COVID-19 and Involvement in Social Media] *Sotsial'naya psikhologiya i obshchestvo [Social Psychology and Society]*, 11(4), 87–104. DOI: 10.17759/sps.2020110407 (in Russian).
- Nestik, T.A., Selezneva, A.V., Shestopal, E.B., & Yurevich, A.V. (2021). Problema psikhologicheskogo sostoyaniya obshchestva i politicheskikh protsessov v sovremennoi Rossii [The problem of the psychological state of society and political processes in modern Russia]. *Voprosy psikhologii [Questions of psychology]*, 67 ( 5), 3-14. (in Russian).
- Soldatova, G. U., Rasskazova, E.I., Nejaskina, Ju.Ju., & Shirjaeva, O. S. (2021). Potrebnost' v informacii i otnoshenie k cifrovym tehnologijam kak faktory kritichnogo i nekritichnogo rasprostraneniya novostej o pandemii [The need for information and attitudes towards digital technologies as factors in the critical and non-critical spread of news about the pandemic]. *Vestnik Moskovskogo Universiteta. Serija 14. Psihologija [Bulletin of Moscow University. Issue 14]*. 1. 170-195. DOI: 10.11621/vsp.2021.01.07(in Russian).

- Tatarko, A.N., Maklasova, E.V., Lepshokova, Z.H., Galjapina, V.N., Efremova, M.V., Dubrov, D.I., Bul'ceva, M.A., Bushina, E.V., & Mironova, A.A. (2020). Metodika ocenki вовлеченности v ispol'zovanie informacionno-kommunikacionnyh tehnologii [Methodology for assessing involvement in the use of information and communication technology]. *Social'naja psihologija i obshhestvo [Social psychology and society]*, 11(1), 159—179. DOI:10.17759/sps.2020110110 (in Russian).
- Aldwairi, M., Alwahedi, A. (2018). Detecting fake news in social media networks. *Procedia Computer Science*, 141, 215—222. DOI: 10.1016/j.procs.2018.10.171
- Allington, D., Duffy, B., Wessely, S., Dhavan, N., & Rubin, J. (2021). Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. *Psychological Medicine*, 51(10), 1763—1769. DOI:10.1017/S0033291721000593
- Baly, R., Karadzhov, G., Alexandrov, D., Glass, J., & Nakov, P. (2018). Predicting Factuality of Reporting and Bias of News Media Sources. *Factuality of Reporting and Bias of News Media Sources. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Belgium: Association for Computational Linguistics*, 3528—3539. DOI:10.18653/v1/D18-1389
- Bessi, A., Coletto, M., Davidescu, G., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Science vs Conspiracy: Collective Narratives in the Age of Misinformation. *PLoS ONE*, 10(2). DOI:10.1371/journal.pone.0118093. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0118093> (Accessed: 27.06.2022).
- Bode, L., & Vraga, E. K. (2018). See something, say something: Correction of global health misinformation on social media. *Health Communication*, 33(9), 1131—1140. DOI:10.1080/10410236.2017.1331312
- Buchanan, T., Vladlena, B. (2019). Spreading Disinformation on Facebook: Do Trust in Message Source, Risk Propensity, or Personality Affect the Organic Reach of 'Fake News'? *Social Media and Society*, 5(4). DOI:10.1177/2056305119888654
- Chan, M.P.S., Jones, C.R., Hall, J.K., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, 28, 1531—1546. DOI:10.1177/0956797617714579
- Ecker, U.K.H., Lewandowsky, S., & Cook, J. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1, 13—29. DOI: 10.1038/s44159-021-00006-y.

- Featherstone, J. D., Davis, C. A., & Zhang, J. (2019). Correcting vaccine misinformation on social media using fact-checking labels. *APHA's 2019 annual meeting and expo. (2- 6 November)*.
- Gao, M., Xiao, Z., Karahalios, K., & Fu, W.-T. (2018). To label or not to label: The effect of stance and credibility labels on readers' selection and perception of news articles. *Proceedings of the ACM on Human-Computer Interaction, 2(CSCW)*, 1–16. DOI:10.1145/3274324
- George, M., Mallery, P. (2011). Descriptive Statistics. SPSS for Windows step by step. A simple guide and reference 18.0 update. S. Hartman (Ed.), (pp.95-104). Boston: Pearson.
- Gilani, Z., Kochmar, E., & Crowcroft, J. (2017). Classification of twitter accounts into automated agents and human users. In: Proceedings of the 2017 IEEE. *ACM International Conference on Advances in Social Networks Analysis and Mining*. ACM, 489–496. DOI:10.1145/3110025.3110091
- Grimmelmann, J. (2015). The Law and Ethics of Experiments on Social Media Users. *Colorado Technology Law Journal*, 13, 219-271. DOI:10.31228/osf.io/cdt7y
- Kahan, D., Peters, E., Dawson, E., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, 1, 54–86. DOI:10.1017/bpp.2016.2
- Köbis, N. C., Bonnefon, J.-F., & Rahwan, I. (2021). Bad machines corrupt good morals. *Nature Human Behaviour*, 5(6), 679–685. DOI:10.1038/s41562-021-01128-2
- Kramer, A.D., Guillory, J.E., & Hancock, J.T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8780-8790. DOI:10.1073/pnas.1320040111
- Kuran, T., & Sunstein, C. R. (1999). “Availability Cascades and Risk Regulation”. *Stanford Law Review*, 51(4), 683 DOI: 10.2307/1229439.
- Mocanu, D., Rossi, L., Zhang, Q., & Karsai, M., Quattrociocchi W. (2015) Collective attention in the age of (mis)information. *Computers in Human Behavior*. DOI: 10.1016/j.chb.2015.01.024
- Nisbet, E.C., & Kamenchuk, O. (2021). Russian News Media, Digital Media, Informational Learned Helplessness, and Belief in COVID-19 Misinformation, International. *Journal of Public Opinion Research*, 33, 571–590. DOI:10.1093/ijpor/edab011
- Nyhan, B. (2021). Why the backfire effect does not explain the durability of political misperceptions. *Proceedings of the National Academy of Sciences*, 118(15). DOI: 10.1073/pnas.1912440117

- Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303–330. DOI:10.1007/s11109-010-9112-2
- Nyhan, B., & Reifler, J. (2019). The roles of information deficits and identity threat in the prevalence of misperceptions. *Journal of Elections, Public Opinion, and Parties*, 29(2), 222–244. DOI:10.1080/17457289.2018.1465061
- Redlawsk, D., & Mattes, K. (2022). Emotions and Politics. *The Cambridge Handbook of Political Psychology*. D. Osborne & C. Sibley (Eds.) (Cambridge Handbooks in Psychology), (pp. 139-158). Cambridge: Cambridge University Press. DOI:10.1017/9781108779104.010
- Trevors, G. J., Muis, K. R., Pekrun, R., Sinatra, G. M., & Winne, P.H. (2016). Identity and Epistemic Emotions During Knowledge Revision: A Potential Account for the Backfire Effect. *Discourse Processes*, 53, 5-6, 339–370. DOI: 10.1080/0163853X.2015.1136507
- Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, 85(3), 423–441. DOI:10.1080/03637751.2018.1467564
- Walter, N., & Tukachinsky, R. (2020). A Meta-Analytic Examination of the Continued Influence of Misinformation in the Face of Correction: How Powerful Is It, Why Does It Happen, and How to Stop It?. *Communication Research*, 47 (2), 155-177. DOI:10.1177/0093650219854600
- Witte, K., Allen, M. (2000). A Meta-Analysis of Fear Appeals: Implications for Effective Public Health Campaigns. *Health Education & Behavior*, 27(5), 591–615. DOI:10.1177/109019810002700506
- Wittenberg, C., & Berinsky, A. (2020). Misinformation and Its Correction. *Social Media and Democracy: The State of the Field, Prospects for Reform*. N. Persily & J. Tucker (Eds.), (pp. 163-198). Cambridge: Cambridge University Press DOI:10.1017/9781108890960.009
- Yang, K., Niven, T., & Kao, H. Fake News Detection as Natural Language Inference. *WSDM '19*. (February 11–15, 2019). DOI: 10.1145/1122445.1122456
- Zhang, X., Ko, M., & Carpenter, D. (2016). Development of a scale to measure skepticism toward electronic word-of-mouth. *Computers in Human Behavior*, 56, 198-208. DOI:10.1016/j.chb.2015.11.042

Библиографическая ссылка на статью:

Нестик Т.А., Михеев Е.А. Психологические факторы эффективности опровержения дезинформации в социальных сетях // Институт психологии Российской академии наук. Организационная психология и психология труда. 2022. Т. 7. № 2. С. 65 - 94. DOI: 10.38098/ipran.opwp\_2022\_23\_2\_003

Nestik, T. A., Mikheev, E. A. (2022). Psihologicheskie faktory jeffektivnosti oproverzhenija dezinformacii v social'nyh setjah [Psychological factors of the effectiveness of refuting disinformation in social networks]. *Institut Psikhologii Rossiyskoy Akademii Nauk. Organizatsionnaya Psikhologiya i Psikhologiya Truda [Institute of Psychology of the Russian Academy of Sciences. Organizational Psychology and Psychology of Labor]*, 7 (2), С. 65 - 94. DOI: 10.38098/ipran.opwp\_2022\_23\_2\_003

Адрес ссылки <http://work-org-psychology.ru/engine/documents/document783.pdf>