



шолтфа

**А.С.Ахременко, А.П.Ч.Петров,  
Д.К.Стукал, С.А.Жеглов, М.В.Хавроненко**

## **НА ЧТО СПОСОБНЫ БОТЫ? МОДЕЛЬ ПРОТЕСТНОЙ И КОНТРПРОТЕСТНОЙ ПОЛИТИЧЕСКОЙ МОБИЛИЗАЦИИ<sup>1</sup>**

<sup>1</sup> Исследование выполнено при финансовой поддержке РФФИ и ЭИСИ в рамках научного проекта №20-011-31351.

Андрей Сергеевич Ахременко — доктор политических наук, профессор департамента политики и управления факультета социальных наук Национального исследовательского университета «Высшая школа экономики». Для связи с автором: aakhremenko@hse.ru.

Александр Пхоун Чжо Петров — доктор физико-математических наук, ведущий научный сотрудник Института прикладной математики им. М.В.Келдыша РАН. Для связи с автором: petrov.alexander.p@yandex.ru.

Денис Константинович Стукал — кандидат политических наук, зам. директора Института прикладных политических исследований Национального исследовательского университета «Высшая школа экономики». Для связи с автором: dstukal@hse.ru.

Сергей Александрович Жеглов — аспирант Аспирантской школы по политическим наукам Национального исследовательского университета «Высшая школа экономики». Для связи с автором: s\_zheglov@mail.ru.

Максим Викторович Хавроненко — студент факультета политологии МГУ им. М.В.Ломоносова. Для связи с автором: mxavronenko@mail.ru.

**Аннотация.** Несмотря на растущий интерес исследователей к влиянию автоматизированных аккаунтов социальных медиа (интернет-ботов) на процессы политической коммуникации и мобилизации в онлайн-среде, степень их эффективности и механизмы использования остаются слабо изученными. Одна из причин дефицита общего концептуального понимания и конкретных результатов состоит в стремлении исследователей решить проблему сугубо эмпирическим путем, без попыток соединить анализ данных с математическим и вычислительным моделированием.

Проанализировав существующий опыт построения такого рода моделей, авторы предлагают собственную разработку, базирующуюся на теории «спирали молчания». Ключевыми особенностями представленной ими модели, отличающими ее от имеющихся аналогов, являются: (а) фиксация

различий в мотивации и «стоимости» выражения протестной и лояльной позиций; (б) включение в механизм «спирали молчания» «эффекта партнера»; (в) использование нейрологической схемы принятия решений, согласно которой один и тот же стимул может побуждать к действию и быть сдерживающим фактором.

На основе серии численных экспериментов с моделью авторы демонстрируют, что при мобилизации оппозиционеров боты наиболее эффективны в ситуации, когда мотивированный к политическому участию индивид не проявляет активности, поскольку его локальное социальное окружение не разделяет его взглядов. В этом случае появление бота-единомышленника способно разрушить созданную этим окружением «спираль молчания», тем самым стимулировав данного индивида к открытому выражению своей позиции. Напротив, при мобилизации лоялистов боты наиболее эффективны применительно к слабо мотивированным индивидам.

Построенная авторами модель не только позволяет по-новому оценить эффекты ботов, но и проливает свет на то, как — в рамках политической коммуникации и мобилизации в социальных сетях — принимают решения люди.

**Ключевые слова:** боты, автоматизированные аккаунты, социальные медиа, политический протест, политическая мобилизация, агентное моделирование, «спираль молчания»

**Введение:  
боты повсеместные  
и неизвестные**

Изучение автоматизированных аккаунтов (ботов) в социальных сетях — новая междисциплинарная область, востребованная при исследовании механизмов политической мобилизации в интернете и находящаяся на стыке социальных и компьютерных наук. Ее развитие в последние годы происходило на фоне сообщений о влиянии ботов на выборы в Германии<sup>2</sup>, референдум о выходе Великобритании из ЕС<sup>3</sup> и обвинений руководства России в инициировании вмешательства в выборы президента США в 2016 г.<sup>4</sup> На этой волне появился массив работ под общим названием «вычислительная пропаганда» (computational propaganda), концентрирующих внимание на потенциальном использовании автоматизированных аккаунтов в целях продвижения повестки дня и интересов государств, не входящих в Североатлантический альянс<sup>5</sup>. Хотя такого рода идеологически предвзятый подход выглядит как минимум сомнительным, свой (и немалый) вклад в рост интереса академического сообщества к этой теме он внес. В 2021 г. первая статья, посвященная исследованию ботов, была опубликована в самом престижном международном журнале по политическим наукам — *American Political Science Review*<sup>6</sup>.

В более широком контексте изучение роли автоматизированных аккаунтов в социальных сетях осуществляется в рамках направления, известного как «вычислительные социальные науки» (computational social sciences), внутри которого выделяются два основных подхода к проблематике ботов. Один подход, развиваемый специалистами в области

<sup>2</sup> Brachten et al. 2017.

<sup>3</sup> Howard and Kollanyi 2016.

<sup>4</sup> Bessi and Ferrara 2016.

<sup>5</sup> Howard et al. 2018; Neudert, Howard, and Kollanyi 2019; Bradshaw and Howard 2019.

<sup>6</sup> Silva and Proksch 2021.

компьютерных наук и наук о данных, фокусируется на выявлении ботов в совокупности активных аккаунтов социальных сетей; другой, представленный работами обществоведов (в первую очередь политологов и социологов), ориентирован на исследование реального функционирования автоматизированных аккаунтов и их вклада в процессы поляризации и радикализации социальных сетей, распространение дискриминационных нарративов и предрассудков и т.п.<sup>7</sup> Диалог между этими двумя группами затрудняют не только различия в методологии, но и сама новизна этой области знания, обуславливающая большое число открытых вопросов, которые создают впечатление, что время междисциплинарного синтеза еще не пришло.

<sup>7</sup> См. *European Symposium Series 2018, 2019.*

Проблемы начинаются уже на уровне определения того, что считать автоматизированным аккаунтом. Авторы одной из первых работ по этой теме<sup>8</sup> предложили подразделять аккаунты на ботов, людей и киборгов, рассматривая последнюю категорию как промежуточную. Необходимость ее введения отчасти связана с многообразием форм автоматизации в социальных сетях, отчасти же объясняется попытками создателей автоматизированных аккаунтов обойти существующие методы детектирования ботов путем поддержания перемежающейся активности автоматизированных алгоритмов и обычных людей.

<sup>8</sup> *Chu et al. 2012.*

Многочисленные исследования показывают, что сложность и внешнее подобие автоматизированных аккаунтов тем, которые поддерживаются обычными людьми без обращения к автоматическим настройкам, возрастают<sup>9</sup>. Активность современных автоматизированных аккаунтов может даже описываться циркадными ритмами, схожими с человеческими<sup>10</sup>. Прогрессирующий рост сложности алгоритмов, используемых для автоматизации аккаунтов, обуславливает дискуссии по поводу самого содержания объекта исследований. Общее определение ботов как «компьютерных алгоритмов, разработанных для подражания человеческому поведению и взаимодействия с людьми на базе социальных онлайн-платформ»<sup>11</sup> строится на явных и не проверенных эмпирически допущениях о целях и типах активности автоматизированных аккаунтов и выводит за рамки ботов такую их разновидность, как аккаунты, созданные для обхода других алгоритмов (например, для продвижения некоего контента)<sup>12</sup>.

<sup>9</sup> См., напр. *Everett et al. 2016; Stieglitz et al. 2017.*

<sup>10</sup> *Stella et al. 2018.*

<sup>11</sup> *Schuchard et al. 2019.*

<sup>12</sup> *Stukal et al. 2019.*

Впрочем, темпы совершенствования и достигнутый уровень сложности интернет-ботов тоже неочевидны. Так, недавнее исследование рынка автоматизированных аккаунтов (включая «черный рынок» в DarkNet)<sup>13</sup> свидетельствует о преобладании на нем примитивных алгоритмов, очень далеких от даже самых скромных представлений об искусственном интеллекте.

<sup>13</sup> *Assenmacher et al. 2020.*

Спорным остается и вопрос о наиболее продуктивном подходе к выявлению автоматизированных аккаунтов. Дискуссия по этому поводу ведется в рамках компьютерных наук и наук о данных и представлена двумя основными подходами, один из которых опирается на использование размеченных данных и массивов уже выявленных ботов

<sup>14</sup> *Ratkiewicz et al. 2011; Oentaryo et al. 2016; Ferrara et al. 2016; Stukal et al. 2017; Yang et al. 2020.*

<sup>15</sup> *Chavoshi, Hamooni, and Mueen 2016.*

<sup>16</sup> *Bessi and Ferrara 2016.*

для поиска аналогичных действующих аккаунтов<sup>14</sup>, а другой — на поиск нетипичных либо в том или ином отношении странных паттернов в активности аккаунтов или их групп<sup>15</sup>.

Наконец, открытым является вопрос о последствиях активности автоматизированных аккаунтов. К чему приводит их существование? Как они воздействуют на процессы распространения информации? Как меняется под их влиянием содержание политической онлайн-коммуникации? Способствуют ли такие аккаунты маргинализации не поддерживаемых ими точек зрения? Исследования на эту тему касаются преимущественно предполагаемого вмешательства России в американские выборы и имеют четко выраженную идеологическую окраску<sup>16</sup>.

Одна из наиболее очевидных причин острой нехватки научных результатов, проливающих свет на влияние автоматизированных аккаунтов на различные аспекты политической коммуникации и мобилизации в онлайн-среде, — стремление исследователей решить проблему сугубо эмпирическим путем, без попыток соединить анализ данных с построением математических и вычислительных моделей. Математическое моделирование позволяет (как минимум) сформулировать в явном виде предположения о механизмах влияния автоматизированных аккаунтов на политическое поведение, что крайне важно в свете обозначенных выше проблем — дефицита концептуального понимания, сложностей с верифицированной эмпирической базой, методологией идентификации ботов и т.д. В настоящей статье мы проанализируем существующий опыт моделирования интернет-ботов в политической коммуникации и предложим собственную разработку, закрывающую некоторые лакуны в имеющихся исследованиях.

**Боты  
в политических  
дискуссиях:  
опыт  
моделирования**

Опыт моделирования активности ботов в политических дискуссиях на сегодняшний день довольно скуден. Речь идет о считанных публикациях, практически каждую из которых мы упомянем ниже. Вместе с тем есть все основания полагать, что эта область исследований будет активно развиваться. Почти все подобные публикации появились в самые последние годы (в основном в 2019—2020 гг.) на фоне интенсивного роста интереса к моделированию политических онлайн-коммуникаций в целом. И уже можно говорить по крайней мере об отдельных особенностях данного направления, которые и будут задавать тон последующего развития.

Прежде всего отметим складывание двух подходов, которые условно обозначим как «алгоритмически ориентированный» и «теоретически ориентированный». Первый нацелен на максимально точное, хотя и в рамках разумных модельных ограничений, воспроизводство свойств алгоритмов реальных социальных сетей. Модели этого типа адаптируют свою сетевую архитектуру под *de facto* существующие платформы (например, отношения подписки в Твиттере и Инстаграме отображаются через направленные связи между пользователями); вводится принцип

ограниченности внимания (limited attention), в соответствии с которым пользователь может обработать не более некоторого определенного количества сообщений за единицу времени; в модель включается прообраз рекомендательной системы. В результате получается достаточно универсальная модель, позволяющая исследовать широкий круг феноменов коммуникации в диапазоне от политической поляризации до диффузии протестных настроений. Характерными примерами такого подхода являются модель «информационного загрязнения» ботами интернет-среды, представленная Ксиаоданом Лоу, Алессандро Фламмини и Филиппо Менсцером<sup>17</sup>, и модель «дезинформационных маневров» ботов в Твиттере Дэвида Бескоу и Кэтлин Карли<sup>18</sup>. При этом каждая из названных моделей может служить инструментом решения гораздо большего числа исследовательских вопросов, нежели указано в заголовках публикаций<sup>19</sup>.

<sup>17</sup> Lou, Flammini, and Menczer 2019.

<sup>18</sup> Beskow and Carley 2019.

<sup>19</sup> Отметим, что наш коллектив тоже создал модель политической коммуникации в Твиттере (см. Akhremenko, Filippov, and Yureskul 2020), но заложенные в ней возможности симуляции активности ботов еще требуют систематического анализа.

<sup>20</sup> Ross et al. 2019.

<sup>21</sup> Cheng, Luo, and Yu 2020.

<sup>22</sup> Напомним, что суть этого феномена, впервые подробно описанного Элизабет Ноэль-Нойман (см. Noelle-Neumann 1974), заключается в том, что пребывание в меньшинстве побуждает индивида избегать публичного выражения своей позиции. Обзор литературы по этой теме см. Sohn 2019.

<sup>23</sup> Детальный обзор состояния и современных трендов агентного моделирования см. Akhremenko, Петров и Жеглов 2021.

Универсальность алгоритмически ориентированного подхода имеет, однако, и оборотную сторону, накладывая ограничения на возможности теоретического анализа конкретных политических явлений в онлайн-среде. Смещение фокуса с мотивации и принятия решений людьми на работу алгоритмов неизбежно обедняет содержательное осмысление политической коммуникации и мобилизации. В альтернативном — теоретически ориентированном — подходе модели исходно «заточены» под особенности конкретного мобилизационного процесса (скажем, распространения призывов к протесту), тогда как специфика функционирования конкретных платформ задается скорее эскизно. Применительно к моделированию активности ботов данный подход использован в разработках Бьёрна Росса и его коллег<sup>20</sup>, а также Чана Ченга с соавторами<sup>21</sup>, посвященных исследованию роли автоматизированных аккаунтов в таком феномене, как «спираль молчания»<sup>22</sup>.

Таким образом, если в рамках первого подхода центральное место занимает вопрос о тех возможностях для влияния на политическую коммуникацию, которые обеспечивают ботам алгоритмы социальных медиа, то второй подход делает акцент на том, какие особенности и механизмы политической коммуникации могут использовать боты в социальных медиа. Последний фокус, на наш взгляд, имеет существенно большее отношение к ключевым политологическим проблемам мобилизации, и в настоящей работе мы предложим модель, выдержанную в духе теоретически ориентированного подхода.

Общим для обоих подходов является то, что все модели выполнены в дизайне агентного моделирования (agent-based modeling, ABM)<sup>23</sup> и представляют собой объединения взаимосвязанных элементов (именуемых агентами), которые в дискретные моменты времени находятся в одном из возможных состояний. Этот набор состояний есть своего рода «поведенческий репертуар» агента, и в большинстве моделей он ограничен всего двумя состояниями: «находится в покое — протестует», «молчит — высказывает позицию», «неактивен — передает сообщение» и т.п. Переход агента из одного состояния в другое определяется

правилом типа: «если суммарная мотивация к протесту превышает сумму контрстимулов, переходит из состояния „в покое“ в состояние „протестует“», или «если более половины пользователей, на которых ты подписан, опубликовали некое сообщение, тоже опубликуй его». Эти состояния влияют на состояния других агентов, с которыми данный агент связан сетевыми структурами или отношениями пространственной близости, и, в свою очередь, зависят от них. Кроме состояний, агенты характеризуются внутренними (то есть не влияющими на окружающих) признаками, индивидуальные значения которых могут быть зафиксированы в момент инициализации модели, а могут меняться эндогенно в процессе симуляции.

<sup>24</sup> Ross et al. 2019. Отметим, что эта модель во многом опирается на предшествующие разработки Донганга Соны (см., напр. Sohn and Geidner 2016).

Проиллюстрируем реализацию этих принципов в модели влияния ботов на «спираль молчания», созданной Россом с коллегами<sup>24</sup>, тем более что ее особенности во многом послужили отправной точкой для наших размышлений. Основной характеристикой агентов в модели выступает позиция по некоему вопросу, представляющая собой качественную дихотомию «за» (+) — «против» (–). Соответствующие значения в некоторой пропорции (в простейшем случае поровну) определяются для всех агентов на этапе инициализации модели и далее не меняются. Эти экзогенно заданные позиции {+, –} могут быть скрытыми (и потому неспособными оказывать социальное воздействие) или открытыми, то есть выраженными публично и доступными для других индивидов. И здесь естественным образом выделяются два интересующих нас класса агентов — агенты-люди (индивиды) и агенты-боты. Принципиальная особенность последних заключается в том, что их позиции всегда открыты; автоматизированные аккаунты и созданы для того, чтобы влиять на общественное мнение публично транслируемыми сообщениями. В случае же агентов-людей ключевое решение, которое принимается в каждом временном такте развития модели, — это именно решение об артикуляции или сохранении в тайне своей позиции.

Решение это отчасти зависит от внутренней склонности индивидов к самоцензуре, варьирующей в интервале [0, 1]. Значение названной переменной, как и позиция, не меняется во времени. Вместе с тем на него влияют *открытые* позиции в окружении индивида — и здесь уже в полной мере проявляется специфика «спирали молчания». Росс и его соавторы вводят переменную «уверенность в своей позиции», индивидуальные значения которой меняются во времени в зависимости от позиций в сетевом окружении. Это тоже непрерывная переменная со значениями в интервале [0, 1]. Если открытые позиции большинства агентов (не только людей!), с которыми связан данный индивид, совпадают с его собственной позицией, его уверенность в ней будет возрастать, в противном случае — уменьшаться. Это и задает механизм влияния ботов на локальное общественное мнение (opinion climate): демонстрируя позиции, они воздействуют на соотношение точек зрения в коммуникативном пространстве.

Итак, открытость/закрытость позиции индивида определяется балансом между внутренней и неизменной склонностью к самооценке, с одной стороны, и динамической по своей природе и зависящей от окружения уверенностью индивида в собственной позиции — с другой. Со временем в такой модели устанавливается равновесие: соотношение открытых и скрытых позиций «за» и «против» перестает меняться. Это соотношение является макроскопической характеристикой исследуемой системы и основным «выходным» (output) параметром. Ключевым экспериментальным параметром выступает доля ботов (с одинаковой позицией) в модели. По заключению Росс с коллегами, всего 2–4% ботов в сети коммуникаций способны привести к доминированию продвигаемой ими позиции в 2/3 случаев, что, конечно, выглядит фантастически.

Из опубликованных результатов анализа других упомянутых нами моделей также следует, что применение ботов — чуть ли не решающий фактор динамики мнений. Так, Бескоу и Карли говорят о 12-процентной доле ботов в сети как о критическом пороге, при превышении которого начинается «игра в одни ворота»<sup>25</sup>; Лоу, Фламмини и Менсцера — о 10-процентной<sup>26</sup>; Ченг с коллегами — о 5–10-процентной<sup>27</sup> и т.д.

Подобные результаты хотя и не опровергаются прямо эмпирическими исследованиями (как уже отмечалось, полноценное эмпирическое тестирование эффективности ботов остается задачей на будущее), но все же не находят подтверждения. Более того, если указанные выше пороговые значения верны, то — исходя из существующих эмпирических оценок распространенности ботов в социальных сетях<sup>28</sup> — можно с уверенностью утверждать, что политическая коммуникация в социальных медиа полностью контролируется ботами. Такое утверждение кажется нам, однако, крайне сомнительным.

На наш взгляд, корни проблемы кроются в слишком прямолинейной трактовке заложенного в модели механизма воздействия ботов на локальное общественное мнение. Ведь принципиальны здесь не алгоритмические тонкости функционирования ботов, а те детали механизма, которые имеют отношение к принятию решений человеком. В связи с этим в предлагаемой ниже модели мы постараемся учесть несколько существенных параметров, касающихся как возможного влияния ботов, так и политической мобилизации в онлайн-среде в целом.

Во-первых, мы будем рассматривать ситуацию, когда позиции «за» и «против» несимметричны и позиция «за» означает поддержку действующей власти (или некоей принятой нормы — например, нормы «политкорректности» в западных странах), а позиция «против» — наоборот. В этих обстоятельствах «стоимость» публичного выражения данных позиций становится неодинаковой; более того, мы покажем различия и в системах мотивации индивидов, придерживающихся той или иной точки зрения. Во-вторых, мы включим в механизм «спирали

<sup>25</sup> *Beskow and Carley 2019.*

<sup>26</sup> *Lou, Flammini, and Menczer 2019.*

<sup>27</sup> *Cheng, Luo, and Yu 2020.*

<sup>28</sup> *См. Subrahmanian et al. 2016; Stukal et al. 2017.*

<sup>29</sup> *Rashevsky 1949.*

молчания» открыт еще в середине прошлого века «эффект партнера». Наконец, при оценке принятия индивидуальных решений мы будем опираться на нейробиологическую схему Николааса Рашевского<sup>29</sup>, согласно которой один и тот же стимул может ложиться одновременно на обе «чашы весов» — побуждать к действию и быть сдерживающим фактором. В целом мы предложим более нюансированную картину «спирали молчания», проливающую свет на конкретный механизм влияния автоматизированных аккаунтов.

**Математическая  
модель,  
или Мой партнер —  
бот**

<sup>30</sup> *Asch 1951.*

Предположим, что некий оппозиционер N учитывает позиции своих знакомых — лоялистов А, В и С — таким образом, что если они трое высказали свое (проправительственное) мнение по некоторому политическому вопросу, то он проявляет конформизм и воздерживается от высказывания своего. Собственно, этот феномен и известен в политологии и теории коммуникаций как «спираль молчания». Психологические основания конформизма были впервые изучены Соломоном Ашем<sup>30</sup>. В его экспериментах испытуемый включался в группу из нескольких человек, которым предлагалось ответить на вопрос, касающийся неких объективных вещей (например, их просили сопоставить по длине изображенные на карточке линии). Испытуемый отвечал последним, а остальные члены группы по указанию экспериментатора давали заведомо неправильные ответы. В результате выяснилось, что в  $\frac{3}{4}$  случаев испытуемый воспроизводит мнение большинства. Условия экспериментов менялись как самим Ашем, так и его многочисленными последователями<sup>31</sup>. В ходе их проведения был обнаружен, в частности, упомянутый выше «эффект партнера»: наличие хотя бы одного правильного ответа существенно повышало вероятность того, что индивид выскажет вслух то, что он в действительности думает.

<sup>31</sup> *См., напр. Morris and Miller 1975; Egebark and Ekström 2018.*

Применительно к нашему примеру это означает: если некий «партнер» (то есть кто-то из окружения N) выскажет оппозиционную точку зрения, этого может оказаться достаточно для того, чтобы разомкнуть «спираль молчания», созданную А, В и С, и мотивировать N высказать свое мнение. И таким «партнером» может быть не только человек, но и бот. Тем самым влияние ботов может заключаться в «партнерстве», а именно в том, что они могут помогать индивидам найти силы, чтобы артикулировать собственную позицию. Но насколько эффективно такое влияние? Чтобы оценить его, мы используем вычислительную модель, имитирующую протестную мобилизацию в социальной сети.

Узлами этой сети выступают индивиды, каждый из которых является либо лоялистом, либо оппозиционером. В каждый конкретный момент соответствующая политическая позиция либо закрыта, либо открыта для наблюдения соседям по сети. Упрощенно это можно представить так. Каждый день обсуждается новая политическая тема, причем все они ставят пользователей перед выбором — проявить



(может быть, в очередной раз) свою лояльность/оппозиционность или промолчать. Принимая решение об открытии/закрытии позиции, индивид исходит из ряда факторов, включающих влияние со стороны товарищей по сети. Скажем, если все сетевое окружение некоего оппозиционно настроенного индивида высказывает лоялистские взгляды, это мотивирует его скрывать свою оппозиционность. Таким образом, модель учитывает взаимное влияние участников коммуникации. И поскольку на пользователей-людей влияют и боты, это дает возможность анализировать и сравнивать между собой динамику процессов в сетях, содержащих разное количество таковых. Предположим, что к некоему сообществу, состоящему из 400 лоялистов и 400 оппозиционеров, подсоединились 200 ботов-лоялистов или, наоборот, ботов-оппозиционеров. Как будут различаться процессы в этих двух случаях? Позволят ли боты своей стороне одержать информационную победу, то есть заставить замолчать оппонентов? Чтобы ответить на эти вопросы, мы проводим с моделью численные эксперименты. Однако прежде чем переходить к их описанию, имеет смысл чуть подробнее остановиться на дизайне самой модели, так как именно в этих деталях проявляются главные различия между ней и рассмотренными выше разработками.

В каждый конкретный момент времени каждый пользователь принимает решение о том, держать ли свою политическую позицию закрытой или открыть ее так, что она будет видна соседям по сети. Открытость или закрытость зависит от латентной позиции пользователя  $\varphi + \psi(t)$ , представляющей собой сумму установки  $\varphi$ , характеризующей предрасположенность индивида к проявлению своей позиции, и динамической компоненты  $\psi(t)$ . Если латентная позиция  $\varphi + \psi(t)$  достаточно велика, то в момент времени  $t$  политическая позиция данного индивида открыта, то есть его соседи видят, что он проявляет себя как лоялист либо как оппозиционер. В противном случае его политическая позиция закрыта (заметим, однако, что это не означает, что соседи «забыли» о ней, если она была открыта ранее).

Установка  $\varphi$  ( $-1 \leq \varphi \leq 0$ ) полагается сформированной долгосрочными факторами (такими, как социальное положение индивида, опыт участия в дискуссиях, психологические особенности и т.д.) и неизменной на протяжении анализируемого процесса. Динамическая компонента  $\psi(t)$  вычисляется как разность мотивов к открытости и закрытости позиции:

$$\psi(t) = M_{open}(t) - M_{close}(t).$$

Мотивы оппозиционеров и лоялистов определяются по-разному. В первом случае сетевое окружение участвует в формировании мотивации, создавая локальную норму поведения. Если в таком окружении доминируют открытые оппозиционеры, это воспринимается индивидом как социальная норма и тем самым мотивирует

<sup>32</sup> Представление о такого рода мотиве было введено и эмпирически обосновано Бертом Кландермансом (Klandermans 1984). В его работе этот тип мотива назывался социальным; термин «нормативный» был предложен Стефаном Штурмером и Берндом Саймоном (см. Stürmer and Simon 2004).

его к открытию позиции. Соответствующий мотив можно назвать нормативным<sup>32</sup>.

Чтобы формализовать нормативный мотив, обозначим через  $x(t-1)$  количество оппозиционеров в окружении индивида, выразивших свое мнение хотя бы в один из моментов  $0, 1, 2... t-1$ ; через  $y(t-1)$  — аналогичное количество лоялистов. Положим для нормативного мотива оппозиционеров:

$$M_{norm}^{opposition}(t) = \begin{cases} \frac{x(t-1) - y(t-1)}{x(t-1) + y(t-1)} & \text{if } it \text{ is } > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Так, если все окружение оппозиционера состоит из его единомышленников, нормативный мотив принимает максимальное значение, равное единице:  $M_{norm}^{opposition}(t)=1$ . Если таковых лишь половина (или меньше),  $M_{norm}^{opposition}(t)=0$ ; если три четверти —  $M_{norm}^{opposition}(t)=0,5$  и т.д.

Другая компонента мотива оппозиционера к открытию позиции связана с гневом в отношении власти. В свою очередь этот гнев тоже имеет две компоненты. Первая проистекает из обстоятельства, положившего начало дискуссии. Скажем, если обсуждаемый вопрос — это коррупционные обвинения, то первая компонента отвечает возмущению в связи с этими обвинениями. Вторая же соответствует возмущению, обусловленному репрессиями за политическое участие (в данном случае — за открытие позиции). Обозначив эти компоненты через  $a_0$  и  $s$  соответственно и произведя нормировку таким образом, чтобы максимальное значение как для каждого элемента мотива, так и для мотива в целом было равно единице, получим:

$$\boxed{\text{мотив к открытию позиции}} = \left( \boxed{\text{мотив, связанный с гневом}} + \boxed{\text{нормативный мотив}} \right) / 2,$$

то есть в формальном выражении:

$$M_{open}^{opposition}(t) = \frac{1}{2} \left[ \frac{a_0 + s}{2} + M_{norm}^{opposition}(t) \right].$$

Применительно к мотиву оппозиционера закрыть позицию будем исходить из того, что он тем выше, чем выше уровень репрессий. Конкретная спецификация имеет вид:

$$M_{close}^{opposition}(t) = 1 - \exp[-k_s s]^{33}$$

<sup>33</sup> В численных экспериментах принято  $k_s=3$ .

В частности, если репрессии отсутствуют ( $s=0$ ), мотив к закрытию позиции будет равен нулю:  $M_{close}^{opposition}(t)=1-1=0$ .

Итак, положим, что оппозиционер с установкой  $\phi$  открывает свою позицию в момент времени  $t$ , если



то есть оппозиционер высказывает свое мнение, если суммарная мотивация превышает порог, создаваемый «спиралью молчания» и препятствующий открытию позиции. Этот порог тем выше, чем больше оппонентов и меньше единомышленников находятся в сетевом окружении данного индивида. В математической нотации эта формула имеет вид:

$$\phi + M_{open}^{opposition}(t) - M_{close}^{opposition}(t) > T(t),$$

а функциональная спецификация порога, создаваемого «спиралью молчания», выбирается так, чтобы учесть «эффект партнера»:

$$T(t) = e^{-x(t-1)} \left( 1 - e^{-y(t-1)/3} \right).$$

При такой спецификации наличие одного единомышленника в окружении оппозиционера уменьшает порог примерно в три раза по сравнению с ситуацией, когда окружение полностью состоит из лоялистов, а наличие трех — в 20 раз.

Итак, сетевое окружение влияет на открытие позиции оппозиционера следующим образом. В случае если в нем доминируют единомышленники, они создают локальную норму, которая мотивирует данного индивида к политическому участию путем открытия позиции. В свою очередь доминирование оппонентов формирует вокруг него «спираль молчания», препятствуя участию. Однако эта «спираль» разрушается, если у оппозиционера имеется партнер. Иначе говоря, результирующее влияние сетевого окружения на индивида незначительно, когда в его окружении преобладают открытые оппоненты, но имеется и некоторое количество единомышленников либо численность сторон сбалансирована.

Относительно лоялистов примем, что, в отличие от оппозиционеров, они не находятся под угрозой репрессий и не испытывают гнева в связи с обстоятельством, обусловившим начало дискуссии. Вместе с тем окружение тоже оказывает на них двойное влияние. В частности, «спираль молчания» действует в их случае так же, как в случае оппозиционеров.

Помимо оппозиционеров и лоялистов, в сети присутствуют оппозиционные и лоялистские боты. Они отличаются от пользователей-людей в двух аспектах. Во-первых, боты имеют меньше связей. Этот

параметр введен в модель потому, что в реальных сетях пользователи по большей части избегают установления связей с автоматизированными аккаунтами. Второе отличие состоит в том, что позиция бота всегда открыта. Эти две идеи не являются оригинальными: на тех же допущениях строятся и предшествующие модели, в частности модель Росса и его коллег.

Обратимся теперь к результатам вычислительных экспериментов с описанной моделью.

### **Численные эксперименты и их результаты**

<sup>34</sup> *Детальное его описание см. Ахременко, Петров и Жеглов 2021.*

Общая схема численного исследования<sup>34</sup> имела следующий вид. Эксперименты включали в себя контролируруемую и стохастическую составляющие. В качестве контролируемых параметров выступали уровень репрессий (11 градаций от 0 до 1 с шагом в 0,1) и численность лояльных и оппозиционных ботов. Во всех экспериментах общее количество агентов в сети составляло 1000: по 400 лоялистов и оппозиционеров-людей и 200 ботов. Применительно к последним рассматривались три варианта: все они — оппозиционеры; все они — лоялисты; 100 ботов — оппозиционеры, 100 — лоялисты.

<sup>35</sup> *Barabási and Albert 1999.*

Стохастическая составляющая отражает вариативность отклонений характеристик реальных объектов от математических идеализаций. Так, используемый в модели алгоритм преференциального присоединения Барабаши-Альберт<sup>35</sup> порождает случайную модельную сеть, архитектурно соответствующую реальным интернет-сетям. Несоответствия между модельной и реальными сетями в среднем невелики, но для одиночного случая они могут оказаться существенными и повлиять на результат. Чтобы нивелировать такие эффекты, для каждого набора контролируемых параметров проводится значительное число (в настоящей работе — 200) прогонов, в которых случайная модельная сеть реализуется заново; при этом при анализе результатов учитываются все прогоны. Помимо сети, стохастическим является также распределение индивидов по установке, извлекаемое из равномерного распределения на отрезке  $[-1, 0]$ .

Наблюдаемая в ходе каждого прогона динамика заключается в том, что с течением времени (в соответствии с представленными выше уравнениями, отражающими анализ мотивации) изменяются латентные позиции пользователей и, соответственно, численность открытых лоялистов и оппозиционеров. По прошествии некоего периода устанавливается равновесное состояние, при котором эта численность становится неизменной (для данной модели опытным путем выяснено, что во всех случаях для этого «с запасом» достаточно ста временных тактов). Эти равновесные состояния анализировались на предмет зависимости от контролируемых параметров.

Перед тем как представить основной результат, напомним, что, когда в социальном окружении пользователя доминируют индивиды с открытыми позициями, противоположными его собственной, порождаемая этим окружением «спираль молчания» создает повышенный

порог для его мотивации. Мы будем называть этот порог вторым; лишь при его преодолении пользователь открывает свою позицию и тем самым присоединяется к протестной или контрпротестной активности. В отсутствие оппонирующего окружения для участия достаточно преодолеть более низкий порог. Индивид, не преодолевший даже первого порога, недостаточно мотивирован для открытия позиции. Индивид, преодолевший оба, открывает позицию. Индивид, находящийся между порогами, мотивирован настолько, что мог бы открыться, но не делает этого ввиду оппонирующего окружения (эффект «спирали молчания»).

Главный вывод настоящей работы состоит в том, что оппозиционные боты помогают своим единомышленникам-людям преодолеть второй порог в большей степени, чем первый. Другими словами, боты отчасти увеличивают мотивацию оппозиционеров-людей через влияние на климат мнений, однако на оппозиционеров, которые находятся между порогами, боты влияют не только через климат, но также через разрушение «спирали молчания».

В табл. 1 приведены результаты одного из экспериментов (параметры модели:  $a_0=1$ ,  $s=0$ ,  $k_s=3$ ; нецелые значения численности возникают ввиду усреднения по 200 прогонам модели) применительно к оппозиционерам. Нетрудно увидеть, что с увеличением количества оппозиционных ботов происходит не только увеличение численности открытых оппозиционеров (что естественно), но и уменьшение численности оппозиционеров между порогами. Другими словами, процесс преодоления «спирали молчания» и открытия позиций является более интенсивным, нежели процесс преодоления первого порога (что особенно заметно по показателям относительного прироста числа оппозиционеров).

**Таблица 1** Влияние численности ботов на оппозиционеров<sup>36</sup>

<sup>36</sup> Здесь и в табл. 2 в скобках указаны минимальные уровни значимости (*p-value*) для сравнения средних двух независимых выборок по критерию Стьюдента с поправкой Беджамини-Хохберга на множественную проверку гипотез. Среднее из соответствующей ячейки сравнивается со средним из ячейки сверху.

Численность оппозиционных ботов	Средняя численность оппозиционеров		
	ниже первого порога / относительный прирост	между порогами / относительный прирост	открытые / относительный прирост
0	261,05	34,13	104,82
100	238,35 / -0,09 (<0,01)	23,1 / -0,32 (<0,01)	138,55 / 0,32 (<0,01)
200	207,19 / -0,13 (<0,01)	14,83 / -0,36 (<0,01)	177,98 / 0,28 (<0,01)

В табл. 2 представлены результаты того же эксперимента в части лоялистов. Здесь закономерность имеет противоположный знак: с увеличением количества ботов-лоялистов численность людей-лоялистов между порогами повышается. Причина такого различия кроется в том, что лоялисты, в противоположность оппозиционерам, испытывают недостаток как стимула к политическому участию в виде гнева, так и мотива отказаться от него в форме страха. Соответственно, дополнительный мотив, создаваемый ботом-единомышленником, оказывается для оппозиционера лишь небольшой добавкой к гневу, а для лоялиста — существенной компонентой мотивации.

**Таблица 2 Влияние численности ботов на лоялистов**

Численность лоялистских ботов	Средняя численность лоялистов		
	ниже первого порога / относительный прирост	между порогами / относительный прирост	открытые / относительный прирост
0	382,93	1,06	16,01
100	367,26 / -0,04 (<0,01)	1,49 / 0,41 (0,018)	31,25 / 0,95 (<0,01)
200	346,96 / -0,06 (<0,01)	2,08 / 0,40 (0,004)	50,96 / 0,63 (<0,01)

Указанные результаты продемонстрировали свою робастность при вариации параметров эксперимента (всего было проведено 396 экспериментов по 200 прогонов). Такие вариации касались, в частности, момента включения ботов в систему: (а) на момент старта симуляции боты уже активированы, (б) боты активируются в первом такте, (в) система достигает равновесного состояния без ботов, после чего они включаются и достигается новое равновесие; для контроля и сравнения изучался также вариант г: боты неактивны на протяжении всей симуляции. Варьировали также стартовые условия: (а) изначально позиции всех пользователей закрыты, (б) изначально открыты позиции 50 случайных пользователей, (в) изначально открыты позиции 50 пользователей, имеющих наибольшее число связей с другими пользователями.

### **Заключение**

Подводя итоги, постараемся взглянуть на проделанную работу под более широким углом зрения, нежели просто количественные результаты модельного эксперимента по проблематике влияния ботов.

Главной особенностью представленного исследования является использование математического и вычислительного моделирования в качестве основного инструмента, и стремление продемонстрировать его возможности было одним из важных мотивов написания этой статьи. Что же дало построение и исследование модели в этом конкретном случае?

Прежде всего следует отметить, что моделирование стало не просто полезным, но необходимым шагом к выработке теории. Используя наработки социальной психологии (концепция «спирали молчания» и теория социальной мотивации Берта Кландерманса) и опираясь на имеющиеся экспериментальные данные («эффект партнера»), мы предложили новый подход к процессу принятия решения о публичной артикуляции мнения по конфликтной политической теме. В его рамках социальное окружение воздействует на индивида двояко: с одной стороны, открытые сторонники создают локальную норму поведения, мотивирующую к высказыванию, с другой — открытые оппоненты формируют противоположные стимулы, порождая «спираль молчания». На качественном, вербальном уровне развести эти два эффекта довольно трудно, и здесь на помощь приходит формальная спецификация — каждый из них описывается своей формулой.

Еще одна «неоднозначность» связана с вводимой нами асимметрией публичных высказываний «за» и «против» с точки зрения возможных негативных санкций. Это имеет принципиальное значение, когда речь идет как о протесте против власти в автократиях, так и о несогласии с некоей неформальной, но «политкорректной» нормой в западных демократиях. Угрозы санкций для протестующих и несогласных в нашей конструкции вновь работают двояко, вызывая гнев и страх. Первый стимулирует людей к выражению позиции, второй — к молчанию. И поскольку эти факторы — в теории и в реальной жизни — действуют одновременно на одних и тех же людей, разумной стратегией их осмысления выступает математическое моделирование, позволяющее на теоретическом уровне развести эмпирически сложно идентифицируемые эффекты.

Следующий шаг состоит в переходе от теории к эмпирически проверяемым гипотезам. В нашем случае они связаны с мобилизационным влиянием автоматизированных аккаунтов социальных сетей. Мы показываем, что применительно к поддерживающим власть и протестующим против нее это влияние не только количественно, но и качественно различно. Если для первых боты играют бóльшую роль в усилении мотивации к высказыванию, участвуя в формировании локальной социальной нормы, то для вторых — в ослаблении барьеров к высказыванию, помогая преодолеть эффект «спирали молчания». Такое положение вещей отнюдь не вытекает из интуитивных соображений и может быть наглядно продемонстрировано только в рамках численного эксперимента — «фирменного» инструмента моделирования.

Эмпирическая проверка полученного модельного результата на материале больших данных социальных медиа — дело будущего. Пока

же нам остается надеяться, что проведенное исследование не только позволит по-новому оценить эффекты ботов или особенности конфликтной политической коммуникации в социальных сетях, но и внесет свой вклад в методологическую дискуссию более широкого плана. При этом мы полагаем, что именно в политической науке — работающей со сложными системами, контринтуитивными эффектами, неоднозначными концептуализациями и испытывающей дефицит надежного эмпирического материала — моделирование может стать одним из ключевых инструментов в арсенале как теоретика, так и эмпирически ориентированного исследователя.

## Библиография

- Ахременко А.С., А.П.Петров и С.А.Жеглов. (2021) «Как информационно-коммуникационные технологии меняют тренды в моделировании политических процессов: к агентному подходу» // *Политическая наука*, № 1: 12–45.
- Akhremenko A., I.Filippov, and E.Yureskul. (2020) «The Effects of Opinion Leader Radicalization under Different User Tolerance Levels: Simulating Political Communications on Twitter» // *2020 International Conference on Engineering Management of Communication and Technology (EMCTECH)*. IEEE: 1–7.
- Asch S.E. (1951) «Effects of Group Pressure on the Modification and Distortion of Judgments» // Guetzkow H., ed. *Groups, Leadership and Men*. Pittsburgh (PA): Carnegie Press: 177–190.
- Assenmacher D., L.Clever, L.Frischlich, T.Quandt, H.Trautmann, and C.Grimme. (2020) «Demystifying Social Bots: On the Intelligence of Automated Social Media Actors» // *Social Media and Society*, vol. 6, no. 3. URL: [https://www.researchgate.net/publication/344019850\\_Demystifying\\_Social\\_Bots\\_On\\_the\\_Intelligence\\_of\\_Automated\\_Social\\_Media\\_Actors](https://www.researchgate.net/publication/344019850_Demystifying_Social_Bots_On_the_Intelligence_of_Automated_Social_Media_Actors) (accessed on 20.03.2021).
- Barabási A.L. and R.Albert. (1999) «Emergence of Scaling in Random Networks» // *Science*, vol. 286, no. 5439: 509–512.
- Beskow D.M. and K.M.Carley. (2019) «Agent Based Simulation of Bot Disinformation Maneuvers in Twitter» // *2019 Winter Simulation Conference (WSC)*. IEEE: 750–761. <http://simulation.su/uploads/files/default/2019-beskow-carley.pdf> (accessed on 20.03.2021).
- Bessi A. and E.Ferrara. (2016) «Social Bots Distort the 2016 US Presidential Election Online Discussion» // *First Monday*, vol. 21, no. 11: 7.
- Brachten F., S.Stieglitz, L.Hofeditz, K.Kloppenborg, and A.Reimann. (2017) «Strategies and Influence of Social Bots in a 2017 German State Election: A Case Study on Twitter» // *ArXiv: 1710.07562*. URL: <https://arxiv.org/abs/1710.07562> (accessed on 21.02.2021).
- Bradshaw S. and P.N.Howard. (2019) *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation*. URL: <https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf> (accessed on 21.02.2021).



Chavoshi N., H.Hamooni, and A.Mueen. (2016) «Identifying Correlated Bots in Twitter» // Spiro E. and Y.Ahn, eds. *Social Informatics. Lecture Notes in Computer Science*, vol. 10047. Cham: Springer: 14–21. URL: [https://www.researchgate.net/publication/308021003\\_Identifying\\_Correlated\\_Bots\\_in\\_Twitter](https://www.researchgate.net/publication/308021003_Identifying_Correlated_Bots_in_Twitter) (accessed on 20.04.2021).

Cheng C., Y.Luo, and C.Yu. (2020) «Dynamic Mechanism of Social Bots Interfering with Public Opinion in Network» // *Physica A: Statistical Mechanics and Its Applications*, vol. 551(C): 124163.

Chu Z., S.Gianvecchio, H.Wang, and S.Jajodia. (2012) «Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?» // *IEEE Transactions on Dependable and Secure Computing*, vol. 9, no. 6: 811–824.

Cresci S., R. Di Pietro, M.Petrocchi, A.Spognardi, and M.Tesconi. (2016) «DNA-Inspired Online Behavioral Modeling and Its Application to Spambot Detection» // *IEEE Intelligent Systems*, vol. 31, no. 5: 58–64.

Edwards C., A.Edwards, P.R.Spence, and A.K.Shelton. (2014) «Is That a Bot Running the Social Media Feed? Testing the Differences in Perceptions of Communication Quality for a Human Agent and a Bot Agent on Twitter» // *Computers in Human Behavior*, vol. 33: 372–376.

Egebark J. and M.Ekström. (2018) «Liking What Others „Like“: Using Facebook to Identify Determinants of Conformity» // *Experimental Economics*, vol. 21, no. 4: 793–814.

*European Symposium Series on Societal Challenges in Computational Social Science: Bias and Discrimination*. (2018) URL: <http://symposium.computationalsocialscience.eu/2018/> (accessed on 21.02.2021).

*European Symposium Series on Societal Challenges in Computational Social Science: Polarization and Radicalization*. (2019) URL: <http://symposium.computationalsocialscience.eu/2019/> (accessed on 21.02.2021)

Everett R.M., J.R.Nurse, and A.Erola. (2016) «The Anatomy of Online Deception: What Makes Automated Text Convincing?» // *Proceedings of the 31st Annual ACM Symposium on Applied Computing*. New York: Association for Computing Machinery: 1115–1120. URL: [https://www.researchgate.net/publication/303773435\\_The\\_anatomy\\_of\\_online\\_deception\\_what\\_makes\\_automated\\_text\\_convincing](https://www.researchgate.net/publication/303773435_The_anatomy_of_online_deception_what_makes_automated_text_convincing) (accessed on 20.04.2021).

Ferrara E., O.Varol, C.Davis, F.Menczer, and A.Flammini. (2016) «The Rise of Social Bots» // *Communications of the ACM*, vol. 59, no. 7: 96–104.

Howard P. N. and B.Kollanyi. (2016) *Bots, #Strongerin, and #Brexit: Computational Propaganda during the UK-EU Referendum*. Working Paper 2016.1. Oxford: Project on Computational Propaganda. URL: [www.politicalbots.org](http://www.politicalbots.org) (accessed on 21.02.2021).

Howard P.N., B.Ganesh, D.Liotsiou, J.Kelly, and C.François. (2018) *The IRA, Social Media and Political Polarization in the United States, 2012–2018*. URL: <https://digitalcommons.unl.edu/senatedocs/1/> (accessed on 21.02.2021).

Klandermans B. (1984) «Mobilization and Participation: Social-psychological Expansions of Resource Mobilization Theory» // *American Sociological Review*, vol. 49, no. 5: 583–600.

Lou X., A.Flammini, and F.Menczer. (2019) «Information Pollution by Social Bots» // *ArXiv: 1907.06130*. URL: <https://arxiv.org/abs/1907.06130v1> (accessed on 21.02.2021).

Morris W.N. and R.S.Miller. (1975) «The Effects of Consensus-breaking and Consensus-preempting Partners on Reduction of Conformity» // *Journal of Experimental Social Psychology*, vol. 11, no. 3: 215—223.

Neudert L.M., P.Howard, and B.Kollanyi. (2019) «Sourcing and Automation of Political News and Information during Three European Elections» // *Social Media and Society*, vol. 5, no. 3: 2056305119863147.

Noelle-Neumann E. (1974) «The Spiral of Silence a Theory of Public Opinion» // *Journal of Communication*, vol. 24, no. 2: 43—51.

Oentaryo R.J., A.Murdopo, P.K.Prasetyo, and E.P.Lim. (2016) «On Profiling Bots in Social Media» // Spiro E. and Y.Ahn, eds. *Social Informatics. Lecture Notes in Computer Science*, vol. 10046. Cham: Springer: 92—109.

Rashevsky N. (1949). «Mathematical Theory of Human Relations: An Approach to a Mathematical Biology of Social Phenomena» // *Bulletin of the American Mathematical Society*, vol. 55: 722—724.

Ratkiewicz J., M.Conover, M.Meiss, B.Gonçalves, A.Flammini, and F.Menczer. (2011) «Detecting and Tracking Political Abuse in Social Media» // *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. Barcelona: 297—304. URL: <https://aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2850/3274> (accessed on 20.04.2021).

Ross B., L.Pilz, B.Cabrera, F.Brachten, G.Neubaum, and S.Stieglitz. (2019) «Are Social Bots a Real Threat? An Agent-based Model of the Spiral of Silence to Analyse the Impact of Manipulative Actors in Social Networks» // *European Journal of Information Systems*, vol. 28, no. 4: 394—412.

Schuchard R., A.Crooks, A.Stefanidis, and A.Croituru. (2018) «Bots in Nets: Empirical Comparative Analysis of Bot Evidence in Social Networks» // *International Conference on Complex Networks and Their Applications*. Cham: Springer: 424—436. URL: [https://www.researchgate.net/publication/329425138\\_Bots\\_in\\_Nets\\_Empirical\\_Comparative\\_Analysis\\_of\\_Bot\\_Evidence\\_in\\_Social\\_Networks](https://www.researchgate.net/publication/329425138_Bots_in_Nets_Empirical_Comparative_Analysis_of_Bot_Evidence_in_Social_Networks) (accessed on 20.04.2021).

Silva B.C. and S.O.Proksch. (2021) «Fake It 'Til You Make It: A Natural Experiment to Identify European Politicians' Benefit from Twitter Bots» // *American Political Science Review*, vol. 115, no. 1: 316—322.

Sohn D. (2019) «Spiral of Silence in the Social Media Era: A Simulation Approach to the Interplay between Social Networks and Mass Media» // *Communication Research*, 23.06. URL: <https://journals.sagepub.com/doi/abs/10.1177/0093650219856510> (accessed on 21.02.2021).

Sohn D. and N.Geidner. (2016) «Collective Dynamics of the Spiral of Silence: The Role of Ego-network Size» // *International Journal of Public Opinion Research*, vol. 28, no. 1: 25—45.

Stella M., E.Ferrara, and M. de Domenico. (2018) «Bots Increase Exposure to Negative and Inflammatory Content in Online Social Systems» // *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 49: 12435—12440.

Stieglitz S., F.Brachten, D.Berthelé, M.Schlaus, C.Venetopoulou, and D.Veutgen. (2017) «Do Social Bots (Still) Act Different to Humans? Comparing Metrics of Social Bots with Those of Humans» // Meiselwitz G., ed. *Social Computing and Social Media: Human Behavior. Lecture Notes in Computer Science*, vol. 10282. Cham: Springer: 379—395. URL: [https://www.researchgate.net/publication/317172975\\_Do\\_Social\\_Bots\\_Still\\_Act\\_Different\\_to\\_Humans\\_-\\_Comparing\\_Metrics\\_of\\_Social\\_Bots\\_with\\_Those\\_of\\_Humans](https://www.researchgate.net/publication/317172975_Do_Social_Bots_Still_Act_Different_to_Humans_-_Comparing_Metrics_of_Social_Bots_with_Those_of_Humans) (accessed on 20.04.2021).

Stukal D., S.Sanovich, R.Bonneau, and J.A.Tucker. (2017) «Detecting Bots on Russian Political Twitter» // *Big Data*, vol. 5, no. 4: 310—324.

Stukal D., S.Sanovich, J.A.Tucker, and R.Bonneau. (2019) «For Whom the Bot Tolls: a Neural Networks Approach to Measuring Political Orientation of Twitter Bots in Russia» // *Sage Open*, vol. 9, no. 2: 2158244019827715.

Stürmer S. and B.Simon. (2004) «The Role of Collective Identification in Social Movement Participation: A Panel Study in the Context of the German Gay Movement» // *Personality and Social Psychology Bulletin*, vol. 30, no. 3: 263—277.

Subrahmanian V., A.Azaria, and S.Durst. (2016) «The DARPA Twitter Bot Challenge» // *Computer*, vol. 49, no. 6: 38—46.

Yang K.C., O.Varol, P.M.Hui, and F.Menczer. (2020) «Scalable and Generalizable Social Bot Detection through Data Selection» // *ArXiv: 1911.09179*. URL: <https://arxiv.org/pdf/1911.09179.pdf> (accessed on 20.04.2021).



ωοιηθε

**A.S.Akhremenko, A.P.Ch.Petrov,  
D.K.Stukal, S.A.Zheglov, M.V.Khavronenko**  
**WHAT BOTS CAN (AND CAN'T) DO?**  
**MODEL OF PROTEST AND COUNTER-PROTEST**  
**POLITICAL MOBILIZATION<sup>37</sup>**

<sup>37</sup> The research was funded by the Russian Foundation for Basic Research (RFBR) and Expert Institute of Social Studies (EISS). The number of the research project is 20-011-31351.

Andrei S. Akhremenko — Doctor of Political Science; Professor at the Department of Politics and Governance, Faculty of Social Sciences, HSE University. Email: [aakhremenko@hse.ru](mailto:aakhremenko@hse.ru).

Alexander P.Ch. Petrov — Doctor of Applied Mathematics; Senior Researcher at the Keldysh Institute for Applied Mathematics (Russian Academy of Sciences). Email: [petrov.alexander.p@yandex.ru](mailto:petrov.alexander.p@yandex.ru).

Denis K. Stukal — Ph.D. in Political Science; Deputy Director at the Institute for Applied Political Studies, HSE University. Email: [dstukal@hse.ru](mailto:dstukal@hse.ru).

Sergey A. Zheglov — Ph.D. Student, Doctoral School of Political Science, HSE University. Email: s\_zheglov@mail.ru.

Maxim V. Khavronenko — BA Student, Faculty of Politics, Lomonosov Moscow State University. Email: mxavronenko@mail.ru.

**Abstract.** Despite the increasing interest among scholars in the effect of Internet bots, or automated social media accounts, on the processes of political communication and mobilization in the online sphere, the extent of bots' effectiveness and the specific mechanisms of their use remain largely understudied. The deficit of the overarching conceptual understanding and concrete results is arguably due to researchers' aspiration to solve a problem in the empirical way, without attempting to combine data analysis with mathematical and computational modeling.

Having analyzed the existing models on the topic, the authors offer their own model that is based on the spiral-of-silence theory. The key features of the model that set it apart from the existing ones are the following: a) taking into account differences in the types of motivation and costs associated with expressing protest and loyalist sentiments; b) including "partner effect" into the spiral-of-silence mechanism; c) employing a neurological decision-making scheme according to which the same stimulus can prompt action and be a deterrent.

On the basis of a series of computational experiments with the model, the authors demonstrate that bots are more effective in mobilizing opposition members when an individual motivated for political participation refrains from it because his local social community does not share his views. In this case, the emergence of a like-minded partner bot can destroy the spiral of silence created by this community and encourage this individual to openly express his position. On the contrary, when mobilizing loyalists, bots are most effective in relation to poorly motivated individuals.

The model elaborated by the authors not only allows us to evaluate bots' effects in a new way, but it also sheds light on how people make decisions in the framework of political communication and mobilization in social networks.

**Keywords:** bots, automated accounts, social media, political protest, political mobilization, agent-based modeling, spiral of silence

## References

Akhremenko A., I.Filippov, and E.Yureskul. (2020) "The Effects of Opinion Leader Radicalization under Different User Tolerance Levels: Simulating Political Communications on Twitter" // *2020 International Conference on Engineering Management of Communication and Technology (EMCTECH)*. IEEE: 1–7.

Akhremenko A.S., A.P.Petrov, and S.A.Zheglov. (2021) "Kak informatsionno-kommunikatsionnye tekhnologii menjajut trendy v modelirovanii politicheskikh protsessov: k agentnomu podkhodu" [How Information and Communication Technologies Change Trends in Modeling Political Processes:

towards an Agent-based Approach] // *Politicheskaja nauka* [Political Science], no. 1: 12–45. (In Russ.)

Asch S.E. (1951) “Effects of Group Pressure on the Modification and Distortion of Judgments” // Guetzkow H., ed. *Groups, Leadership and Men*. Pittsburgh (PA): Carnegie Press: 177–190.

Assenmacher D., L.Clever, L.Frischlich, T.Quandt, H.Trautmann, and C.Grimme. (2020) “Demystifying Social Bots: On the Intelligence of Automated Social Media Actors” // *Social Media and Society*, vol. 6, no. 3. URL: [https://www.researchgate.net/publication/344019850\\_Demystifying\\_Social\\_Bots\\_On\\_the\\_Intelligence\\_of\\_Automated\\_Social\\_Media\\_Actors](https://www.researchgate.net/publication/344019850_Demystifying_Social_Bots_On_the_Intelligence_of_Automated_Social_Media_Actors) (accessed on 20.03.2021).

Barabási A.L. and R.Albert. (1999) “Emergence of Scaling in Random Networks” // *Science*, vol. 286, no. 5439: 509–512.

Beskow D.M. and K.M.Carley. (2019) “Agent Based Simulation of Bot Disinformation Maneuvers in Twitter” // *2019 Winter Simulation Conference (WSC)*. IEEE: 750–761. <http://simulation.su/uploads/files/default/2019-beskow-carley.pdf> (accessed on 20.03.2021).

Bessi A. and E.Ferrara. (2016) “Social Bots Distort the 2016 US Presidential Election Online Discussion” // *First Monday*, vol. 21, no. 11: 7.

Brachten F., S.Stieglitz, L.Hofeditz, K.Kloppenborg, and A.Reimann. (2017) “Strategies and Influence of Social Bots in a 2017 German State Election: A Case Study on Twitter” // *ArXiv: 1710.07562*. URL: <https://arxiv.org/abs/1710.07562> (accessed on 21.02.2021).

Bradshaw S. and P.N.Howard. (2019) *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation*. URL: <https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf> (accessed on 21.02.2021).

Chavoshi N., H.Hamooni, and A.Mueen. (2016) “Identifying Correlated Bots in Twitter” // Spiro E. and Y.Ahn, eds. *Social Informatics. Lecture Notes in Computer Science*, vol. 10047. Cham: Springer: 14–21. URL: [https://www.researchgate.net/publication/308021003\\_Identifying\\_Correlated\\_Bots\\_in\\_Twitter](https://www.researchgate.net/publication/308021003_Identifying_Correlated_Bots_in_Twitter) (accessed on 20.04.2021).

Cheng C., Y.Luo, and C.Yu. (2020) “Dynamic Mechanism of Social Bots Interfering with Public Opinion in Network” // *Physica A: Statistical Mechanics and Its Applications*, vol. 551(C): 124163.

Chu Z., S.Gianvecchio, H.Wang, and S.Jajodia. (2012) “Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?” // *IEEE Transactions on Dependable and Secure Computing*, vol. 9, no. 6: 811–824.

Cresci S., R. Di Pietro, M.Petrocchi, A.Spognardi, and M.Tesconi. (2016) “DNA-Inspired Online Behavioral Modeling and Its Application to Spambot Detection” // *IEEE Intelligent Systems*, vol. 31, no. 5: 58–64.

Edwards C., A.Edwards, P.R.Spence, and A.K.Shelton. (2014) “Is That a Bot Running the Social Media Feed? Testing the Differences in Perceptions of Communication Quality for a Human Agent and a Bot Agent on Twitter” // *Computers in Human Behavior*, vol. 33: 372–376.

Egebark J. and M.Ekström. (2018) “Liking What Others „Like“: Using Facebook to Identify Determinants of Conformity” // *Experimental Economics*, vol. 21, no. 4: 793–814.

*European Symposium Series on Societal Challenges in Computational Social Science: Bias and Discrimination.* (2018) URL: <http://symposium.computationalsocialscience.eu/2018/> (accessed on 21.02.2021).

*European Symposium Series on Societal Challenges in Computational Social Science: Polarization and Radicalization.* (2019) URL: <http://symposium.computationalsocialscience.eu/2019/> (accessed on 21.02.2021)

Everett R.M., J.R.Nurse, and A.Erola. (2016) “The Anatomy of Online Deception: What Makes Automated Text Convincing?” // *Proceedings of the 31st Annual ACM Symposium on Applied Computing*. New York: Association for Computing Machinery: 1115–1120. URL: [https://www.researchgate.net/publication/303773435\\_The\\_anatomy\\_of\\_online\\_deception\\_what\\_makes\\_automated\\_text\\_convincing](https://www.researchgate.net/publication/303773435_The_anatomy_of_online_deception_what_makes_automated_text_convincing) (accessed on 20.04.2021).

Ferrara E., O.Varol, C.Davis, F.Menczer, and A.Flammini. (2016) “The Rise of Social Bots” // *Communications of the ACM*, vol. 59, no. 7: 96–104.

Howard P. N. and B.Kollanyi. (2016) *Bots, #Strongerin, and #Brexit: Computational Propaganda during the UK-EU Referendum*. Working Paper 2016.1. Oxford: Project on Computational Propaganda. URL: [www.politicalbots.org](http://www.politicalbots.org) (accessed on 21.02.2021).

Howard P.N., B.Ganesh, D.Liotsiou, J.Kelly, and C.François. (2018) *The IRA, Social Media and Political Polarization in the United States, 2012–2018*. URL: <https://digitalcommons.unl.edu/senatedocs/1/> (accessed on 21.02.2021).

Klandermans B. (1984) “Mobilization and Participation: Social-psychological Expansions of Resource Mobilization Theory” // *American Sociological Review*, vol. 49, no. 5: 583–600.

Lou X., A.Flammini, and F.Menczer. (2019) “Information Pollution by Social Bots” // *ArXiv: 1907.06130*. URL: <https://arxiv.org/abs/1907.06130v1> (accessed on 21.02.2021).

Morris W.N. and R.S.Miller. (1975) “The Effects of Consensus-breaking and Consensus-preempting Partners on Reduction of Conformity” // *Journal of Experimental Social Psychology*, vol. 11, no. 3: 215–223.

Neudert L.M., P.Howard, and B.Kollanyi. (2019) “Sourcing and Automation of Political News and Information during Three European Elections” // *Social Media and Society*, vol. 5, no. 3: 2056305119863147.

Noelle-Neumann E. (1974) “The Spiral of Silence a Theory of Public Opinion” // *Journal of Communication*, vol. 24, no. 2: 43–51.

Oentaryo R.J., A.Murdopo, P.K.Prasetyo, and E.P.Lim. (2016) “On Profiling Bots in Social Media” // Spiro E. and Y.Ahn, eds. *Social Informatics. Lecture Notes in Computer Science*, vol. 10046. Cham: Springer: 92–109.

Rashevsky N. (1949). “Mathematical Theory of Human Relations: An Approach to a Mathematical Biology of Social Phenomena” // *Bulletin of the American Mathematical Society*, vol. 55: 722–724.

Ratkiewicz J., M.Conover, M.Meiss, B.Gonçalves, A.Flammini, and F.Menczer. (2011) “Detecting and Tracking Political Abuse in Social Media” // *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. Barcelona: 297–304. URL: <https://aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2850/3274> (accessed on 20.04.2021).

Ross B., L.Pilz, B.Cabrera, F.Brachten, G.Neubaum, and S.Stieglitz. (2019) “Are Social Bots a Real Threat? An Agent-based Model of the Spiral of Silence to Analyse the Impact of Manipulative Actors in Social Networks” // *European Journal of Information Systems*, vol. 28, no. 4: 394–412.

Schuchard R., A.Crooks, A.Stefanidis, and A.Croitoru. (2018) “Bots in Nets: Empirical Comparative Analysis of Bot Evidence in Social Networks” // *International Conference on Complex Networks and Their Applications*. Cham: Springer: 424–436. URL: [https://www.researchgate.net/publication/329425138\\_Bots\\_in\\_Nets\\_Empirical\\_Comparative\\_Analysis\\_of\\_Bot\\_Evidence\\_in\\_Social\\_Networks](https://www.researchgate.net/publication/329425138_Bots_in_Nets_Empirical_Comparative_Analysis_of_Bot_Evidence_in_Social_Networks) (accessed on 20.04.2021).

Silva B.C. and S.O.Proksch. (2021) “Fake It ’Til You Make It: A Natural Experiment to Identify European Politicians’ Benefit from Twitter Bots” // *American Political Science Review*, vol. 115, no. 1: 316–322.

Sohn D. (2019) “Spiral of Silence in the Social Media Era: A Simulation Approach to the Interplay between Social Networks and Mass Media” // *Communication Research*, 23.06. URL: <https://journals.sagepub.com/doi/abs/10.1177/0093650219856510> (accessed on 21.02.2021).

Sohn D. and N.Geidner. (2016) “Collective Dynamics of the Spiral of Silence: The Role of Ego-network Size” // *International Journal of Public Opinion Research*, vol. 28, no. 1: 25–45.

Stella M., E.Ferrara, and M. de Domenico. (2018) “Bots Increase Exposure to Negative and Inflammatory Content in Online Social Systems” // *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 49: 12435–12440.

Stieglitz S., F.Brachten, D.Berthel , M.Schlaus, C.Venetopoulou, and D.Veutgen. (2017) “Do Social Bots (Still) Act Different to Humans? Comparing Metrics of Social Bots with Those of Humans” // Meiselwitz G., ed. *Social Computing and Social Media: Human Behavior. Lecture Notes in Computer Science*, vol. 10282. Cham: Springer: 379–395. URL: [https://www.researchgate.net/publication/317172975\\_Do\\_Social\\_Bots\\_Still\\_Act\\_Different\\_to\\_Humans\\_-\\_Comparing\\_Metrics\\_of\\_Social\\_Bots\\_with\\_Those\\_of\\_Humans](https://www.researchgate.net/publication/317172975_Do_Social_Bots_Still_Act_Different_to_Humans_-_Comparing_Metrics_of_Social_Bots_with_Those_of_Humans) (accessed on 20.04.2021).

Stukal D., S.Sanovich, R.Bonneau, and J.A.Tucker. (2017) “Detecting Bots on Russian Political Twitter” // *Big Data*, vol. 5, no. 4: 310–324.

Stukal D., S.Sanovich, J.A.Tucker, and R.Bonneau. (2019) “For Whom the Bot Tolls: a Neural Networks Approach to Measuring Political Orientation of Twitter Bots in Russia” // *Sage Open*, vol. 9, no. 2: 2158244019827715.

St rmer S. and B.Simon. (2004) “The Role of Collective Identification in Social Movement Participation: A Panel Study in the Context of the German Gay Movement” // *Personality and Social Psychology Bulletin*, vol. 30, no. 3: 263–277.

Subrahmanian V., A.Azaria, and S.Durst. (2016) “The DARPA Twitter Bot Challenge” // *Computer*, vol. 49, no. 6: 38–46.

Yang K.C., O.Varol, P.M.Hui, and F.Menczer. (2020) “Scalable and Generalizable Social Bot Detection through Data Selection” // *ArXiv: 1911.09179*. URL: <https://arxiv.org/pdf/1911.09179.pdf> (accessed on 20.04.2021).