



К.А.Толокнев

## НЕВИДИМЫЙ ПОЛИТРУК: КАК АЛГОРИТМЫ ПЕРСОНАЛИЗАЦИИ ФОРМИРУЮТ ОБЩЕСТВЕННОЕ МНЕНИЕ

Кирилл Александрович Толокнев — аспирант аспирантской школы по политическим наукам Национального исследовательского университета «Высшая школа экономики». Для связи с автором: kirtoloknev@gmail.com.

**Аннотация.** Социальные медиа давно и прочно вошли в повседневную жизнь современного человека. Тем не менее их влияние на формирование общественного мнения все еще недостаточно изучено. Важная черта социальных медиа заключается в том, что они не являются нейтральной средой. Не только люди взаимодействуют друг с другом на платформах социальных медиа, но и сами социальные медиа активно взаимодействуют с людьми, подбирая для них персонализированный контент, базирующийся на информации об их интересах и поведении. В 2011 г. Эли Прейзер выдвинул гипотезу, согласно которой персонализация контента должна вести к формированию своеобразных «информационных коконов», или «пузырей фильтров», — гомогенных групп пользователей, придерживающихся схожих взглядов. Однако фрагментация интернет-сообщества на «пузыри фильтров» — не единственная угроза, которой чревато использование алгоритмов персонализации. Еще более опасно, что социальные медиа обладают возможностью манипулировать алгоритмами, подбирающими для пользователей контент, с тем чтобы воздействовать на их взгляды.

В статье предпринята попытка протестировать реальность этих угроз посредством вычислительного моделирования. Для решения этой задачи задействована простая агентно-ориентированная модель, симулирующая влияние алгоритмов персонализации на коммуникацию в социальных медиа. Продемонстрировано, что, вопреки предположению Прейзера, алгоритмы, подбирающие максимально близкий к установкам пользователей контент, только в редких случаях способствуют возникновению «пузырей фильтров». Показано также, что манипулирование алгоритмами персонализации позволяет стабильно воздействовать на формирование общественного мнения лишь в двух ситуациях: (1) когда манипуляции подвергаются все пользователи и при этом они открыты к внешнему влиянию; (2) когда манипуляция направлена на условных «центристов», не обладающих четко выраженным мнением по какому-то вопросу.

**Ключевые слова:** эхо-камера, пузырь фильтров, социальные медиа, вычислительное моделирование, агентно-ориентированная модель, общественное мнение, политические коммуникации

К началу 2010-х годов социальные медиа прочно вошли в нашу жизнь. Они стали важным средством не только потребления новостей и развлекательного контента, но и поддержания социальных связей. Чтобы сохранить интерес пользователей, социальные медиа задействуют специальные алгоритмы, которые на основе данных о поведении и интересах аудитории персонализируют доступный ей контент. По мнению критиков<sup>1</sup>, подобная практика ведет к формированию своего рода «информационных коконов», состоящих из единомышленников, — «пузырей фильтров». Будучи сама по себе достаточно спорной, эта точка зрения обрела популярность среди активистов, журналистов и политиков<sup>2</sup>. Но использование алгоритмов персонализации несет с собой и другую угрозу: поскольку нейтральность не является их имманентной характеристикой, они могут продвигать определенный политический контент, ограничивая доступ к другому.

<sup>1</sup> Pariser 2011.

<sup>2</sup> Zuiderveen  
Borgesius et al.  
2016.

В связи с этим закономерно возникает ряд вопросов. Во-первых, действительно ли использование алгоритмов персонализации способно привести к расколу общества на отдельные «информационные коконы»? Во-вторых, насколько эффективными могут быть манипуляции алгоритмами персонализации? В-третьих, на какие сегменты общества следует нацеливать такие манипуляции, чтобы они дали желаемый результат?

### **«Пузыри фильтров» и социальные медиа**

У коммуникаций в социальных медиа имеются две важные черты. Одна из них — это неравномерное распределение влияния. На первый взгляд может показаться, что коммуникации в социальных медиа устроены весьма демократично. Однако это не так: в социальных медиа выделяется «цифровая элита», обладающая диспропорционально большим, по сравнению с рядовыми пользователями, количеством «друзей» и/или «подписчиков». Социальные медиа представляют собой «безмасштабную сеть», тип графа, в котором распределение степеней вершин (числа соседей) подчинено степенному закону: доля вершин со степенью  $k$  приблизительно равна  $k^{-\gamma}$ , где  $2 < \gamma < 3$ .

<sup>3</sup> McPherson,  
Smith-Lovin, and  
Cook 2001.

<sup>4</sup> Huber and  
Malhotra 2017.

<sup>5</sup> Cinelli et al. 2021.

Вторая черта — гомофилия. Схожие люди, идет ли речь о сходстве социально-экономического положения, этнического происхождения<sup>3</sup> или политических предпочтений<sup>4</sup>, легче вступают в контакт. Отсюда следует, что и в повседневном общении, и в контексте коммуникации в социальных медиа схожие люди с большей вероятностью окажутся связаны друг с другом. Кроме того, люди более склонны воспринимать информацию, которая согласуется с их картиной мира. Это может вести к формированию относительно гомогенных групп<sup>5</sup>, в рамках которых

приверженцы той или иной политической позиции будут в основном получать подтверждающую эту позицию информацию, та же, что не вписывается в нее, просто не сможет проникнуть в этот своеобразный «информационный кокон» — «эхо-камеру».

Впрочем, гомофилия — не единственная возможная причина возникновения «информационных коконов» в социальных медиа. Социальные медиа не являются некоей нейтральной средой, где взаимодействие ведется исключительно между пользователями, — они сами активно взаимодействуют с ними.

Еще в 1995 г. Николас Негропonte прогнозировал, что со временем каждый сможет потреблять тот медиа-контент, который захочет: «Представьте себе будущее, где некий интерфейс будет иметь доступ ко всем газетам, всем телевизионным каналам и всем радиопередачам на земле, а затем составлять персонализированную подборку: своего рода газету, созданную специально для вас»<sup>6</sup>.

<sup>6</sup> *Negroponte 1995: 153.*

Уже после появления современных социальных медиа и поисковых систем Эли Прейзер назвал такую ситуацию «пузырем фильтров»<sup>7</sup>. Под «фильтрами» он понимал алгоритмы, нацеленные на индивидуализацию предлагаемого пользователю контента, с тем чтобы как можно дольше удерживать его внимание и увеличивать время, проводимое им на сайте. По мнению Прейзера, применение таких алгоритмов ведет к тому, что разные пользователи получают разную информацию, при этом поступающая к ним информация соответствует изначально присутствующим им установкам. В итоге пользователи с разными установками оказываются изолированы в «информационных коконах» — собственно «пузырях».

<sup>7</sup> *Pariser 2011.*

Рассуждая о «пузырях фильтров», Прейзер в первую очередь имел в виду поисковую систему Google. Однако в дальнейшем термин «пузыри фильтров» начал чаще применяться к социальным медиа, таким как Twitter и Facebook\*, которые тоже задействовали информацию о пользователях, чтобы индивидуализировать получаемый ими контент. Отчасти это было связано с ростом популярности социальных медиа, отчасти — с повышением интереса как со стороны исследователей, так и со стороны широкой общественности к последствиям политической коммуникации в социальных медиа, а именно к роли социальных медиа как потенциального источника политической поляризации<sup>8</sup>, усиления поддержки правого популизма<sup>9</sup> и распространения дезинформации<sup>10</sup>. Одновременно концепция «пузырей фильтров» срослась с концепцией «эхо-камер»<sup>11</sup>. Подобное слияние негативно сказалось на качестве проводимых исследований, так как возникновение «информационных коконов» в ходе естественного процесса, обусловленного особенностями межличностной коммуникации, принципиально иной феномен, нежели формирование их в результате действий социальных медиа.

<sup>8</sup> *Flaxman, Goel, and Rao 2016; Kitchens, Johnson, and Gray 2020.*

<sup>9</sup> *Hameleers et al. 2021.*

<sup>10</sup> *Au, Ho, and Chiu 2021.*

<sup>11</sup> *Bruns 2019.*

<sup>12</sup> *Munn 2019.*

<sup>13</sup> *Lewis 2018.*

Мнения исследователей относительно последствий использования алгоритмов персонализации расходятся. Согласно исследованиям Люка Мунна<sup>12</sup> и Ребекки Льюис<sup>13</sup>, они действительно могут способст-

- 14 *Bakshy, Messing, and Adamic 2015.* вовать радикализации части пользователей, формируя для них «пузырь фильтров», и тем самым усиливать поляризацию общества. В свою очередь исследование Итана Бакши и его коллег<sup>14</sup> демонстрирует, что алгоритмическая лента новостей играет существенно меньшую роль в снижении потребления контента, созданного представителями «чужой» идеологической группы, и, соответственно, в формировании «информационных коконов», чем собственный выбор пользователей.
- 15 *Fletcher and Nielsen 2018.* По заключению Ричарда Флетчера и Расмуса Нильсена<sup>15</sup>, не погружают своих пользователей в «пузыри фильтров» и поисковые системы: напротив, идеологическое разнообразие получаемых теми новостей выше, чем у тех, кто не задействует поисковые системы. Наконец, исследование Мэтью Барниджа говорит о том, что пользователи социальных медиа получают в них более разнообразную идеологически информацию, нежели в повседневном общении<sup>16</sup>.
- 16 *Barnidge 2017.*

Противоречивые результаты дают и исследования воздействия рекомендательных систем вне социальных медиа: некоторые из них показывают, что использование этих систем ведет не к фрагментации онлайн-сообщества, а, наоборот, к гомогенизации потребляемого пользователями контента, не к расколу сообщества на несколько «информационных коконов», а к формированию общего. Именно такую картину обнаружили, в частности, Тянь Нгуен и его коллеги<sup>17</sup>, проанализировав, как алгоритм персонализации влияет на предпочтения пользователей сайта, посвященного кинематографу. К сходным выводам пришли Картик Хосанагер с коллегами, но уже на данных о работе системы, рекомендовавшей музыку<sup>18</sup>. Правда, следует отметить, что упомянутые исследования не касались собственно социальных медиа — последние практически недоступны для подробного изучения теми, кто не связан с самими социальными медиа.

- 17 *Nguyen et al. 2014.*
- 18 *Hosanagar et al. 2014.*
- 19 *Geschke, Lorenz, and Holtz 2019.* Результаты моделирования воздействия рекомендательных систем тоже оказались неоднозначными. Например, в работе Дэниэла Гешке и его коллег в большей части рассмотренных сценариев их использование вело к фрагментации сообщества, то есть к возникновению отдельных «пузырей»<sup>19</sup>, тогда как в работе Аллисон Чени с коллегами алгоритмы персонализации способствовали гомогенизации поведения пользователей<sup>20</sup>.
- 20 *Chaney, Stewart, and Engelhardt 2018.*

Таким образом, можно выдвинуть две альтернативные гипотезы:  
 (H1) использование алгоритмов персонализации ведет к фрагментации сообщества;  
 (H2) использование алгоритмов персонализации ведет к гомогенизации сообщества.

- Нейтральны ли алгоритмы персонализации? По заключению Вен Чена и его коллег, новостная лента Twitter политически нейтральна, хотя предлагаемая пользователю информация в значительной мере зависит от его политических предпочтений<sup>21</sup>. В то же время, как показывает исследование Анны Кардинал с коллегами, в случае Google и Facebook\* дело обстоит иначе, по крайней мере в Испании: испанские
- 21 *Chen et al. 2021.*

пользователи Facebook\*, придерживающиеся левых взглядов, с большей вероятностью получают идеологически близкий им контент, чем приверженцы правых идеологий. При использовании поиска в Google картина противоположная<sup>22</sup>. Впрочем, это отнюдь не обязательно указывает на намеренную манипуляцию: такое положение вещей может объясняться тем, что левые более активны в Facebook\*, чем правые, и, следовательно, производят больше контента, который может быть рекомендован другим пользователям.

<sup>22</sup> Cardenal et al. 2019.

Вместе с тем есть свидетельства того, что алгоритмы фильтрации могут служить инструментом целенаправленной манипуляции. В 2012 г. Facebook\* провел эксперимент, изменив для части пользователей алгоритмы фильтрации таким образом, чтобы в ленте новостей одним поступало больше постов с позитивным содержанием, а другим — с негативным. В результате обнаружилось, что эмоциональным состоянием пользователей можно манипулировать. Те, в чьей ленте преобладали позитивные посты, сами начинали выкладывать больше позитивных и меньше негативных постов; у тех же, кто видел преимущественно негативные посты, среди их собственных постов стали доминировать негативные<sup>23</sup>.

<sup>23</sup> Kramer, Guillory, and Hancock 2014.

Правомерно предположить, что механизм воздействия алгоритмических манипуляций на общественное мнение сходен с механизмом онлайн-цензуры: завывсив количество желательных для манипулятора точек зрения, можно сдвинуть общественное мнение в их сторону<sup>24</sup>.

<sup>24</sup> Roberts 2020.

Соответственно, можно выдвинуть третью гипотезу:

*(Н3) манипуляция алгоритмами персонализации может эффективно влиять на общественное мнение.*

Для того чтобы протестировать эти гипотезы, обратимся к вычислительному моделированию. Выбор этой методологии обусловлен несколькими соображениями. Во-первых, манипулирование алгоритмами персонализации в социальных медиа доступно только для самих социальных медиа. К тому же этичность подобных экспериментов как минимум спорна. Во-вторых, эта сфера анализа еще относительно молода, имеющиеся эмпирические исследования дают разнонаправленные результаты. То же относится и к существующим теориям. Использование же методов вычислительного моделирования позволит сформулировать более точные гипотезы, предварительно рассмотрев поведение интересующей нас системы в контролируемой модельной среде, чтобы в дальнейшем уже проверять их на эмпирическом материале.

## Модель

Агентно-ориентированное моделирование — один из методов вычислительного моделирования. Исследователь создает модель, состоящую из некоторого конечного числа автономных агентов, обладающих заранее установленным числом параметров и взаимодействующих по определенным и заранее установленным правилам. Агенты могут находиться в ограниченном числе состояний<sup>25</sup>. Поведение модели анали-

<sup>25</sup> Ахременко, Петров и Желтов 2021.

зируется при помощи множества симуляций (вычислительных экспериментов), проводимых с варьируемыми значениями параметров. Каждая отдельная симуляция состоит из этапа инициализации, когда устанавливаются значения параметров агентов, и определенного числа шагов — повторяющихся взаимодействий агентов по заданным правилам. Поскольку некоторые значения параметров могут извлекаться из случайных распределений, для каждой комбинации параметров проводится множество симуляций, после чего анализируются агрегированные данные.

Предлагаемая в данной статье модель состоит из 2000 агентов, объединенных в безмасштабную сеть. Параметр *гомофилия* ( $H$ ), принимающий значение от 0 до 1, определяет, насколько сильно могут различаться мнения агентов, связанных друг с другом в сети: чем он выше, тем ниже уровень гомофилии и тем сильнее могут различаться агенты, вступающие в связь между собой.

<sup>26</sup> Sarup, Suchner, and Gaylord 1991.

Как уже говорилось, социальное влияние происходит тогда, когда люди уже обладают относительно схожими установками<sup>26</sup>. Поэтому до начала симуляции задается параметр  $\alpha$  — *интервал доверия*, определяющий, как сильно могут различаться мнения агентов, способных в ходе взаимодействия повлиять друг на друга. Чем выше значения  $\alpha$ , тем больше агенты открыты для влияния со стороны других агентов.

Кроме того, агенты обладают такими параметрами, как *мнение* ( $O$ ) и *внутренняя мотивация* ( $m$ ). Внутренняя мотивация определяет готовность агента продемонстрировать свою позицию. Мнение отражает отношение агента к какой-либо политической проблеме. Предположим, что при  $O = 0$  агент не поддерживает некую политическую позицию, а при  $O = 1$  полностью поддерживает. Общественное мнение в модели — распределение всех мнений всех агентов.

С наибольшей вероятностью свои мнения в социальных медиа высказывают люди, которые, во-первых, достаточно высоко мотивированы, а во-вторых, обладают достаточно отчетливо выраженным мнением по какому-либо политическому вопросу. Формально это можно выразить следующим образом:

$$(m * |O - 0.5| > U[0,1]),$$

где  $m$  — внутренняя мотивация агента,  $O$  — мнение агента.

По приведенной выше формуле выбираются агенты, активные на момент инициализации. Дальнейшая симуляция включает в себя три шага: (1) для каждого агента подбирается некоторое число других агентов, с которыми он на данном шаге сможет взаимодействовать; (2) новые агенты получают возможность вступить в коммуникацию и стать активными; (3) активные агенты меняют свои мнения. Рассмотрим их по порядку.

Поскольку нас интересует воздействие алгоритмов персонализации, активными агентами, с которыми может происходить обмен мнен-

ниями, будут не непосредственные соседи, а подобранные алгоритмом агенты. Для каждого агента  $i$  случайным образом составляются две выборки максимум по  $N$  агентов, которые являются соседями его соседей, но не его собственными. При этом агенты из одной выборки будут иметь мнение  $O > O_i$ , а из второй —  $O < O_i$ . Далее из обоих множеств случайным образом выделяется максимум по  $n$  агентов. Такой вариант можно назвать *случайным подбором*. Однако он не похож на процесс персонализации контента, когда подбираются мнения, наиболее близкие к мнению агента  $i$ . Поэтому обратимся к еще одному варианту.

Теперь из обоих множеств выделяется максимум по  $n$  агентов, чье различие во мнении с агентом  $i$  минимально. Тем самым симулируется подбор информации, наиболее близкой к мнению агента  $i$ , то есть тот самый процесс, который, согласно Прейзеру, должен вести к формированию сегрегированных «пузырей фильтров». Назовем этот вариант *подбором ближайшего мнения*.

В случаях, когда речь идет о манипуляции, в алгоритм вносятся следующие изменения: для каждого агента  $i$  подбирается не максимум  $n$  агентов с мнением  $O > O_i$ , а  $(n + 1)$  агентов, и не максимум  $n$  агентов с мнением  $O < O_i$ , а  $(n - 1)$  агентов. Соответственно, при манипуляции для каждого подвергнутого ей агента  $i$  будет завышаться число агентов с мнениями  $O > O_i$ . Таким образом, можно ожидать, что успешная манипуляция будет увеличивать медианное значение  $O$ . В условиях отсутствия манипуляций правомерно ожидать, что в среднем медианное мнение будет равно 0,5.

Возможность вступить в коммуникацию получают агенты, в чьем окружении уже достаточно много активных агентов и чья мотивация к вступлению в коммуникацию достаточно высока. Если число активных агентов, подобранных алгоритмом для агента  $i$ , с позицией, входящей в интервал доверия, обозначить как  $n_s$ , а общее число агентов, подобранных алгоритмом для агента  $i$ , — как  $n$ , то формально это можно записать так:

$$\frac{(m + \frac{n_s}{n})}{2} \leq U[0,1].$$

Далее происходит обмен мнениями между активными агентами. Здесь используется модель динамики мнений *ограниченного доверия*<sup>27</sup>. Также предполагается, что влияние оказывается только на тех агентов, чье мнение на момент взаимодействия не изменилось относительно исходного более чем на значение  $\alpha$ , и только стороны его соседей по сети, чьи мнения попадают в диапазон  $\alpha$ .

Взаимодействия с подобранными алгоритмом агентами (обмен мнениями) происходят следующим образом.

Для агента  $i$  если  $[O_{i0} - O_i] < \alpha$ , то

$$\Delta O_i = (1 - m_i) * \mu * w * (\bar{O}_j - O_i),$$

иначе 0,

<sup>27</sup> *Большое семейство моделей динамики мнений, которые строятся на допущении, что мнения меняются только при взаимодействии с достаточно схожими соседями.*



где  $\mu$  — сила воздействия;  $w$  — доля активных «соседей», способных оказывать ассимилятивное влияние, среди всех активных «соседей»;  $\bar{O}_j$  — среднее мнение активных агентов, чье мнение отличается от мнения агента  $i$  не более чем на  $\alpha$ ;  $O_i$  — мнение агента  $i$ .

Соответственно, мнение агента  $i$  после обмена мнениями меняется так:

$$O_{i+1} = O_i + \Delta O_i.$$

На этом шаг модели заканчивается, описанные выше процедуры повторяются, пока не возникнет стабильное распределение мнений агентов.

## Результаты моделирования

Моделирование выполнено в программе Netlogo версии 6.2.0. При симуляции варьировались следующие параметры: уровень гомофилии (0,2; 0,4; 0,6; 0,8; 1), интервал доверия (0,15; 0,2; 0,25; 0,3), сочетания параметров  $N$  и  $n$  ( $N = 20$  и  $n = 4$ ;  $N = 8$  и  $n = 4$ ). Значения параметров  $m$  и  $O$  извлекались из равномерного распределения с минимальным значением 0 и максимальным — 1. Для каждого сочетания параметров проводилось по 30 симуляций протяженностью в 250 шагов, далее анализировались агрегированные результаты по множеству симуляций.

Прежде чем перейти к рассмотрению результатов вычислительных экспериментов, выдвинем несколько дополнительных предположений. Во-первых, следует предположить, что, если гипотеза Прейзера верна, сегрегированные «пузыри фильтров» должны возникать в тех случаях, когда интернет-пользователям предлагается наиболее близкая к их убеждениям информация. В модели это происходит при  $N = 20$  и  $n = 4$ , так как при выборе четырех ближайших мнений из 20 эти мнения с большей вероятностью попадут в интервал  $\alpha$ , чем при выборе из восьми. Во-вторых, «пузыри фильтров» должны возникать при более высоком уровне гомофилии.

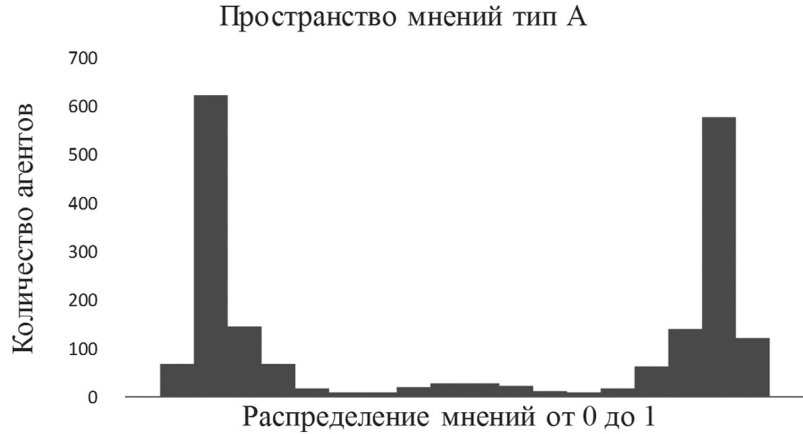
Для начала рассмотрим поведение модели без манипуляций. При использовании алгоритма случайного подбора сообщество раскалывается на две гомогенные группы при  $\alpha > 0,2$  и  $H > 0,2$ . Эти группы возникают в пространстве мнений в промежутках от 0,05 до 0,25 и от 0,75 до 0,95. При  $\alpha = 0,2$  или  $\alpha = 0,15$  образуется от трех до пяти групп. При этом две крупнейшие из них формируются в тех же сегментах пространства мнений. Назовем подобное распределение *пространством мнений типа А*. Пример такого пространства приведен на рис. 1.

Только при условии крайне высокой гомофилии ( $H = 0,2$ ) и высокой открытости внешнему влиянию ( $\alpha = 0,3$ ) происходит гомогенизация сообщества и образуется одна группа (при всех значениях  $N$  и  $n$ ). Впрочем, несмотря на формирование одной группы, максимальное число агентов концентрируется в пространстве мнений в промежутках от 0,3 до 0,35 и от 0,65 до 0,7. Назовем подобное распределение *пространством мнений типа Б*. Пример такого пространства приве-

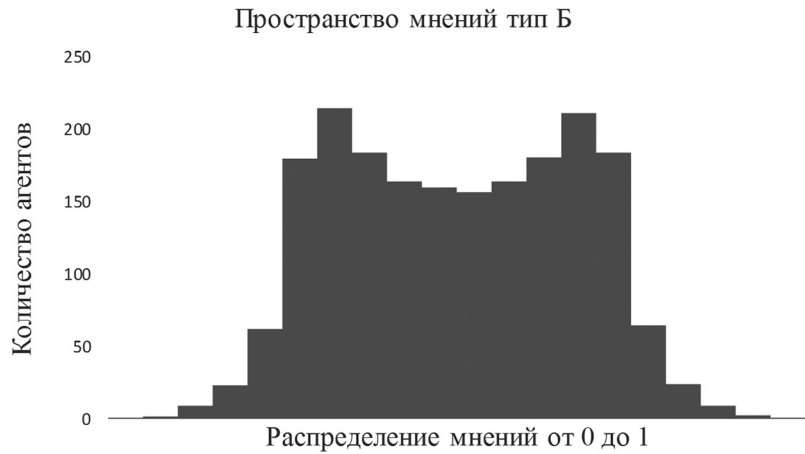


ден на рис. 2. При  $\alpha = 0,25$  и  $\alpha = 0,2$  возникают две крупные группы, при  $\alpha = 0,15$  — три отдельные группы.

**Рисунок 1** Пример формирования пространства мнений типа А  
( $N = 1$ ;  $\alpha = 0,25$ ;  $N = 8$ ;  $n = 4$ ; алгоритм случайного подбора)



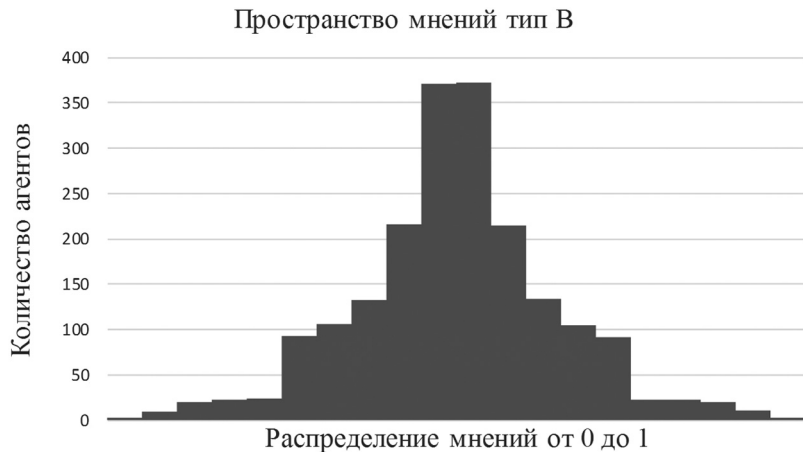
**Рисунок 2** Пример формирования пространства мнений типа Б  
( $N = 0,2$ ;  $\alpha = 0,3$ ;  $N = 8$ ;  $n = 4$ ; алгоритм случайного подбора)



При использовании алгоритма подбора ближайшего мнения раскол сообщества на две сегрегированные гомогенные группы наблюдается в двух случаях: (1)  $\alpha = 0,15$ ,  $N = 8$ ,  $n = 4$ ,  $H = 0,4$ ; (2)  $\alpha = 0,2$ ,  $N = 8$ ,  $n = 4$  и  $H$  от 0,4 до 0,8. При  $\alpha = 0,15$ ,  $N = 8$ ,  $n = 4$ ,  $H > 0,4$  образуются три отдельные группы, как и при сочетании параметров  $\alpha = 0,2$ ,

$N = 8, n = 4, H = 1$ . В иных случаях возникает одна группа, при этом, если  $H = 0,2$ , максимальное число агентов оказывается сконцентрировано в пространстве мнений в промежутках от 0,15 до 0,3 и от 0,7 до 0,85. При уменьшении уровня гомофилии до 0,4 одна группа с двумя «пиками» мнений формируется при  $\alpha = 0,25$  и  $\alpha = 0,3$  и максимальной точности подбора мнений алгоритмом ( $N = 20, n = 4$ ). При дальнейшем снижении гомофилии начинает складываться одна группа с максимальным числом агентов в промежутке от 0,45 до 0,55 пространства мнений. Назовем подобное распределение *пространством мнений типа В*. Пример такого пространства приведен на *рис. 3*. Результаты симуляций при использовании алгоритма подбора ближайшего мнения представлены в *табл. 1*. Как и ожидалось, среднее медианное мнение в сообществе для всех рассмотренных случаев равно 0,5.

**Рисунок 3** Пример формирования пространства мнений типа В ( $H = 1; \alpha = 0,25; N = 8; n = 4$ , алгоритм подбора ближайшего мнения)



Таким образом, предположение Прейзера, что использование алгоритмов персонализации должно вести к образованию сегрегированных «пузырей фильтров», подтверждается лишь в весьма специфических условиях — при высокой гомофилии, относительно низкой открытости агентов к внешнему влиянию и довольно ограниченной точности самого алгоритма подбора. Примечательно, что при очень высокой гомофилии ( $H = 0,2$ ) формирования сегрегированных групп агентов (пространства мнений типа А) не происходит в принципе.

Теперь рассмотрим возможные манипуляции. При наиболее простой для манипулятора форме манипуляции ее объектами оказываются все агенты. При использовании алгоритма подбора ближайшего мнения

**Таблица 1** Формирование различных типов пространства мнений при использовании алгоритма подбора ближайшего мнения

Сочетание параметров $\alpha$ , $N$ и $n$	Значение $H$				
	$H = 0,2$	$H = 0,4$	$H = 0,6$	$H = 0,8$	$H = 1$
$\alpha = 0,15; N = 20; n = 4$	Тип Б	Тип Б	Тип В	Тип В	Тип В
$\alpha = 0,15; N = 8; n = 4$	Тип Б	Тип А	Тип А	Тип А	Тип А
$\alpha = 0,2; N = 20; n = 4$	Тип Б	Тип Б	Тип В	Тип В	Тип В
$\alpha = 0,2; N = 8; n = 4$	Тип Б	Тип А	Тип А	Тип А	Тип А
$\alpha = 0,25; N = 20; n = 4$	Тип Б	Тип Б	Тип В	Тип В	Тип В
$\alpha = 0,25; N = 8; n = 4$	Тип Б	Тип В	Тип В	Тип В	Тип В
$\alpha = 0,3; N = 20; n = 4$	Тип Б	Тип Б	Тип В	Тип В	Тип В
$\alpha = 0,3; N = 8; n = 4$	Тип Б	Тип В	Тип В	Тип В	Тип В

манипуляция эффективна во всех случаях, когда задействован наиболее точный алгоритм подбора, то есть  $N = 20$  и  $n = 4$ . При этом, чем выше значение  $\alpha$ , тем сильнее эффект манипуляции: при  $\alpha = 0,15$  среднее медианное мнение равно 0,62; при  $\alpha = 0,3$  — 0,66. Если подбор менее точен ( $N = 8$  и  $n = 4$ ), а уровень гомофилии высок ( $H = 0,2$ ), манипуляция эффективна; ее эффективность тем больше, чем выше значение  $\alpha$ . При снижении уровня гомофилии поведение более разнообразно. При низкой открытости внешнему влиянию ( $\alpha = 0,15$ ) и снижении уровня гомофилии ( $H \geq 0,4$ ) значение среднего медианного мнения уменьшается (хотя различия в средних показателях по сравнению с ситуацией, когда манипуляция не осуществляется, статистически незначимы). То же самое происходит при чуть большей открытости внешнему влиянию ( $\alpha = 0,2; H > 0,4$ ), но в этом случае различия в средних показателях уже статистически значимы: при  $H = 0,6$  среднее медианное мнение равно 0,46; при  $H = 0,8$  — 0,44; при  $H = 1$  — 0,42. При дальнейшем росте открытости к внешнему влиянию ( $\alpha = 0,25$  и  $\alpha = 0,3$ ) манипуляция эффективна, среднее медианное мнения увеличивается. Интересно, что в обоих случаях неэффективности манипуляции, как и в ситуации отсутствия таковой, сообщество раскалывается на две сегрегированные группы, то есть возникает пространство мнений типа А. Таким образом, получается, что эффективность манипуляции зависит от того, как выглядело бы пространство мнений в «естественном состоянии», то есть без манипуляций.

При использовании алгоритма случайного подбора манипуляция эффективна при крайне высоком уровне гомофилии ( $H = 0,2$ ) и всех значениях  $\alpha$ ,  $N$  и  $n$ . При снижении уровня гомофилии ( $H > 0,2$ ) степень эффективности манипуляции начинает зависеть от значения  $\alpha$ : при  $\alpha = 0,15$  манипуляция эффективна при  $H = 0,4$  и  $H = 0,6$  (при дальнейшем снижении гомофилии различия статистически незначимы); при

$\alpha = 0,2$  и  $\alpha = 0,3$  она эффективна при всех уровнях гомофилии; при  $\alpha = 0,25$  различия статистически незначимы.

Предположим, что манипуляции подвергаются не все агенты, а только те, чьи мнения находятся в промежутке от 0,6 до 1. Для начала рассмотрим алгоритм подбора ближайшего мнения. При высокой точности подбора ( $N = 20$  и  $n = 4$ ) среднее медианное мнение увеличивается (хотя при  $\alpha = 0,15$ ,  $H < 0,6$  различия в средних показателях статистически незначимы), но не столь существенно, как в случае, когда манипуляция затрагивает всех агентов. При менее точном подборе ( $N = 8$ ,  $n = 4$ ) мы видим более разнообразное поведение. При низкой открытости внешнему влиянию ( $\alpha = 0,15$ ) и высокой гомофилии ( $H = 0,2$ ) среднее медианное мнение уменьшается; то же происходит при  $\alpha = 0,2$  и всех уровнях гомофилии и  $\alpha = 0,25$  и низкой гомофилии ( $H > 0,6$ ). Но особенно примечательно, что меняется сам процесс формирования групп. Для примера: если в ситуации отсутствия манипуляции при использовании алгоритма подбора ближайшего мнения и высокой гомофилии ( $H = 0,2$ ) формировалась одна группа, то при данном типе манипуляции образуются уже две относительно сегрегированные группы.

Если используется алгоритм случайного подбора, единственное сочетание параметров, при котором манипуляция эффективна (значение медианного мнения увеличивается), — крайне высокая гомофилия ( $H = 0,2$ ) и высокая открытость внешнему влиянию ( $\alpha = 0,3$ ) (впрочем, различия в средних показателях в этом случае статистически незначимы). При снижении открытости внешнему влиянию и сохранении столь же высокой гомофилии поведение меняется: теперь в результате манипуляции среднее медианное мнение стабильно снижается. При более низкой гомофилии ( $H \geq 0,4$ ) и достаточно высокой открытости внешнему влиянию ( $\alpha = 0,25$  и  $\alpha = 0,3$ ) наблюдается та же картина: среднее медианное мнение стабильно снижается, при этом чем выше значение  $\alpha$ , тем сильнее снижение. При более низкой открытости внешнему влиянию ( $\alpha = 0,2$  и  $\alpha = 0,15$ ) манипуляция не дает эффекта.

Теперь ограничим число агентов, подвергающихся манипуляции, теми, чьи мнения лежат в промежутке от 0,4 до 0,6. В этом случае манипуляции эффективны при обоих используемых алгоритмах и всех значениях  $\alpha$ ,  $H$ ,  $N$ ,  $n$ . Условия, при которых манипуляции стабильно эффективны, приведены в *табл. 2*.

Что касается масштаба изменений, вызванных манипуляциями, то прослеживается следующая закономерность: чем выше значение  $\alpha$ , тем больше изменения. Наиболее ярко это проявляется в ситуации, когда объектом манипуляции становятся все агенты. Как видно из *табл. 3*, при  $\alpha = 0,15$  медианное мнение агентов составляет 0,54 или 0,57 (в зависимости от используемого алгоритма), а при  $\alpha = 0,3$  — уже 0,68 или 0,71<sup>28</sup>.

<sup>28</sup> При усреднении по всем значениям  $H$ ,  $N$ ,  $n$  для заданного значения  $\alpha$ .

Подведем предварительные итоги. Во-первых, можно утверждать, что возникновение сегрегированных «информационных коконов» вследствие использования социальными медиа алгоритмов персонализа-

**Таблица 2 Условия, при которых манипуляции стабильно эффективны**

<i>Манипуляции подвергаются агенты с мнениями в промежутке</i>	<i>Подбор ближайшего мнения</i>	<i>Случайный подбор</i>
от 0 до 1	$\alpha \geq 0,25$ или $H = 0,2$	$\alpha = 0,3$ или $H = 0,2$
от 0,6 до 1	$\alpha = 0,3$	Неэффективны при любом сочетании параметров
от 0,6 до 0,4	Все значения $\alpha$ и $H$	Все значения $\alpha$ и $H$

**Таблица 3 Масштаб воздействия манипуляций на медианное мнение**

<i>Манипуляция в промежутке от 0 до 1</i>		
<i>Значение <math>\alpha</math></i>	<i>Случайный подбор</i>	<i>Подбор ближайшего мнения</i>
$\alpha = 0,15$	0,54	0,57
$\alpha = 0,2$	0,55	0,58
$\alpha = 0,25$	0,57	0,67
$\alpha = 0,3$	0,71	0,68

ции маловероятно. Этого не происходит ни при очень высоком уровне гомофилии ( $H = 0,2$ ), ни при применении максимально точной версии алгоритма подбора ближайшего мнения. Наоборот, к формированию сегрегированных групп ведет случайный или близкий к случайному подбор мнений. Во-вторых, манипуляции при использовании алгоритма подбора ближайшего мнения являются эффективным инструментом влияния на формирование общественного мнения, однако для того, чтобы манипуляция была результативной, обычно требуется либо высокий уровень гомофилии, либо высокая открытость внешнему влиянию.

### **Заключение**

Более 10 лет назад Прейзер высказал предположение, что использование интернет-сайтами алгоритмов персонализации, подбирающих для каждого пользователя близкий ему контент, повлечет за собой фрагментацию общества на «пузыри фильтров» — отдельные «информационные коконы», состоящие из единомышленников. С формированием подобных «пузырей» связывали такие негативные тенденции последнего десятилетия, как углубление политической поляризации и рост поддержки популистских движений. Однако эмпирические исследования так и не смогли однозначно подтвердить взаимосвязь между этими явлениями.

Результаты моделирования позволяют пролить свет на причины неоднозначности заключений. Во-первых, эффект использования ал-

горитмов персонализации зависит от уровня гомофилии, то есть степени близости убеждений людей, объединенных в социальную сеть. Во-вторых, на действенность применения таких алгоритмов влияет степень открытости людей внешнему влиянию. При высокой открытости использование алгоритмов персонализации ведет не к расколу общества на отдельные «пузыри», а к формированию одной гомогенной группы. В случае более низкой открытости (например, когда речь идет о чувствительных политических вопросах) отдельные «пузыри фильтров» действительно образуются, но только при условии, что алгоритм подбирает контент недостаточно точно. Это противоречит гипотезе Прейзера, согласно которой именно точный подбор близкого пользователю контента должен вести к возникновению «пузырей фильтров».

Более того, моделирование показало, что отдельные «пузыри фильтров» стабильно формируются не тогда, когда пользователю предлагается максимально сходный с его убеждениями контент, а тогда, когда тот подбирается случайно. Таким образом, надежды на то, что проблему поляризации общества можно решить посредством «создания алгоритмов, продвигающих более разнообразный контент»<sup>29</sup>, не просто необоснованны, но и контрпродуктивны. Вспомним также, что, согласно некоторым эмпирическим исследованиям, в социальных медиа пользователи получают доступ к более широкому спектру мнений, чем в обычном общении<sup>30</sup>. Это позволяет предположить, что к образованию «информационных коконов» ведет не меньший, а больший объем доступной идеологически разнообразной информации. Впрочем, проверка этого предположения выходит за рамки настоящей статьи.

Что касается основного объекта нашего внимания — потенциальной эффективности манипуляций алгоритмами персонализации, — то моделирование подтвердило, что такие манипуляции действительно могут влиять на формирование общественного мнения. Однако выявился и ряд ограничений. Во-первых, эффективность манипуляций выше в условиях высокой гомофилии и открытости внешнему влиянию. Во-вторых, наиболее стабильно «работают» манипуляции, ориентированные на «неопределившихся» участников сети, равноудаленных от крайних позиций в пространстве мнений. В-третьих, эффективность манипуляции зависит от того, насколько точно алгоритм подбирает контент: при недостаточной точности подбора контента в результате манипуляции общественное мнение может измениться не так, как хотелось манипулятору, и популярность точек зрения, которые он пытался скрыть от пользователей, увеличится.

Как уже говорилось, достоверно известен лишь один прецедент намеренной манипуляции алгоритмами персонализации<sup>31</sup>. Разумеется, не все изменения, вносимые социальными медиа в эти алгоритмы, нацелены на манипулирование установками пользователей: речь может идти, например, о тестировании различных вариантов алгоритма с тем,

<sup>29</sup> *Bozdag and Van Den Hoven 2015: 256—257.*

<sup>30</sup> *Barnidge 2017.*

<sup>31</sup> *Kramer, Guillory, and Hancock 2014.*

чтобы выявить тот, который позволит максимально долго удерживать внимание пользователей на предлагаемой социальными медиа рекламе. Тем не менее было бы наивным полагать, что ситуация, когда социальные медиа либо сами, либо под внешним давлением воспользуются находящимся в их руках инструментом влияния на общественное мнение, полностью исключена. Чтобы избежать такого поворота событий, необходим не только больший контроль над социальными медиа, но и дальнейшее изучение закономерностей их функционирования.

## 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.
- Au C.H., K.K.W.Ho, and D.K.Chiu. (2021) «The Role of Online Misinformation and Fake News in Ideological Polarization: Barriers, Catalysts, and Implications» // *Information Systems Frontiers*, April: 1—24. URL: <https://link.springer.com/article/10.1007/s10796-021-10133-9> (accessed on 28.06.2022).
- Bakshy E., S.Messing, and L.Adamic. (2015) «Exposure to Ideologically Diverse News and Opinion on Facebook» // *Science*, vol. 348, no. 6239: 1130—1132.
- Barnidge M. (2017) «Exposure to Political Disagreement in Social Media Versus Face-to-Face and Anonymous Online Settings» // *Political Communication*, vol. 34, no. 2: 302—321.
- Bozdag E. and J.Van Den Hoven. (2015) «Breaking the Filter Bubble: Democracy and Design» // *Ethics and Information Technology*, vol. 17, no. 4: 249—265.
- Bruns A. (2019) «Filter Bubble» // *Internet Policy Review: Journal on Internet Regulation*, vol. 8, no. 4: 1—14. URL: <https://snurb.info/files/2019/Filter%20Bubble.pdf> (accessed on 28.06.2022).
- Cardenal A.S., C.Aguilar-Paredes, C.Galais, and M.Pérez-Montoro. (2019) «Digital Technologies and Selective Exposure: How Choice and Filter Bubbles Shape News Media Exposure» // *The International Journal of Press/Politics*, vol. 24, no. 4: 465—486.
- Chaney A.J.B., B.M.Stewart, and B.E.Engelhardt. (2018) «How Algorithmic Confounding in Recommendation Systems Increases Homogeneity and Decreases Utility» // *Rec.Sys '18: Proceedings of the 12th ACM Conference on Recommender Systems*. Vancouver: 224—232. URL: <https://arxiv.org/pdf/1710.11214.pdf> (accessed on 28.06.2022).
- Chen W., D.Pacheco, K.C.Yang, and F.Menczer. (2021) «Neutral Bots Probe Political Bias on Social Media» // *Nature Communications*, vol. 12, no. 1: 1—10.
- Cinelli M., G. De Francisci Morales, A.Galeazzi, W.Quattrociocchi, and M.Starnini. (2021) «The Echo Chamber Effect on Social Media» // *Proceedings of the National Academy of Sciences*, vol. 118, no. 9: e2023301118.



Flaxman S., S.Goel, and J.M.Rao. (2016) «Filter Bubbles, Echo Chambers, and Online News Consumption» // *Public Opinion Quarterly*, vol. 80, no. S1: 298—320.

Fletcher R. and R.K.Nielsen. (2018) «Automated Serendipity: The Effect of Using Search Engines on News Repertoire Balance and Diversity» // *Digital Journalism*, vol. 6, no. 8: 976—989.

Geschke D., J.Lorenz, and P.Holtz. (2019) «The Triple-Filter Bubble: Using Agent-Based Modeling to Test a Meta-Theoretical Framework for the Emergence of Filter Bubbles and Echo Chambers» // *British Journal of Social Psychology*, vol. 58, no 1: 129—149.

Hameleers M., D.Schmuck, L.Bos, and S.Ecklebe. (2021) «Interacting with the Ordinary People: How Populist Messages and Styles Communicated by Politicians Trigger Users' Behaviour on Social Media in a Comparative Context» // *European Journal of Communication*, vol. 36, no. 3: 238—253.

Hosanagar K., D.Fleder, D.Lee, and A.Buja. (2014) «Will the Global Village Fracture into Tribes? Recommender Systems and Their Effects on Consumer Fragmentation» // *Management Science*, vol. 60, no. 4: 805—823.

Huber G.A. and N.Malhotra. (2017) «Political Homophily in Social Relationships: Evidence from Online Dating Behavior» // *The Journal of Politics*, vol. 79, no. 1: 269—283. URL: <https://www.journals.uchicago.edu/doi/pdf/10.1086/687533> (accessed on 28.06.2022).

Kitchens B., S.L.Johnson, and P.Gray. (2020) «Understanding Echo Chambers and Filter Bubbles: The Impact of Social Media on Diversification and Partisan Shifts in News Consumption» // *MIS Quarterly*, vol. 44, no. 4: 1619—1649. URL: [https://www.darden.virginia.edu/sites/default/files/inline-files/05\\_16371\\_RA\\_KitchensJohnsonGray%20Final\\_0.pdf](https://www.darden.virginia.edu/sites/default/files/inline-files/05_16371_RA_KitchensJohnsonGray%20Final_0.pdf) (accessed on 28.06.2022).

Kramer A.D.I., J.E.Guillory, and J.T.Hancock. (2014) «Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks» // *Proceedings of the National Academy of Sciences*, vol. 111, no. 24: 8788—8790. URL: <https://digitalwellbeing.org/wp-content/uploads/2014/06/PNAS-2014-Kramer-8788-90.pdf> (accessed on 28.06.2022).

Lewis R. (2018) *Alternative Influence: Broadcasting the Reactionary Right on YouTube*. URL: [https://datasociety.net/wp-content/uploads/2018/09/DS\\_Alternative\\_Influence.pdf](https://datasociety.net/wp-content/uploads/2018/09/DS_Alternative_Influence.pdf) (accessed on 28.06.2022).

McPherson M., L.Smith-Lovin, and J.M.Cook. (2001) «Birds of a Feather: Homophily in Social Networks» // *Annual Review of Sociology*, vol. 27: 415—444.

Munn L. (2019) «Alt-Right Pipeline: Individual Journeys to Extremism Online» // *First Monday*, vol. 24, no. 6. URL: <https://firstmonday.org/ojs/index.php/fm/article/download/10108/7920> (accessed on 28.06.2022).

Negroponte N. (1995) *Being Digital*. New York: Knopf.

Nguyen T.T., P.M.Hui, F.M.Harper, L.Terveen, and J.A.Konstan. (2014) «Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity» // *WWW '14: Proceedings of the 23rd International Conference on World Wide Web*. New York: 677—686.

Pariser E. (2011) *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. New York: Penguin.

Roberts M.E. (2020) «Resilience to Online Censorship» // *Annual Review of Political Science*, vol. 23: 401—419.

Sarup G., R.W.Suchner, and G. Gaylord. (1991) «Contrast Effects and Attitude Change: A Test of the Two-Stage Hypothesis of Social Judgment Theory» // *Social Psychology Quarterly*, vol. 54, no. 4: 364—372.

Zuiderveen Borgesius F.J., D.Trilling, Ju.Möller, B.Bodó, C.H. de Vreese, and N.Helberger. (2016) «Should We Worry about Filter Bubbles?» // *Internet Policy Review: Journal on Internet Regulation*, vol. 5, no. 1: 1—16. URL: <https://www.econstor.eu/bitstream/10419/214006/1/IntPolRev-2016-1-401.pdf> (accessed on 28.06.2022).

\* *Meta, компания, владеющая Facebook, признана в России экстремистской организацией и запрещена на территории страны.*



K.A.Toloknev

## THE INVISIBLE POLITICAL OFFICER: HOW PERSONALIZATION ALGORITHMS SHAPE PUBLIC OPINION

Kirill A. Toloknev — Ph.D. Student at the Doctoral School of Political Science, HSE University. E-mail: kirtoloknev@gmail.com.

**Abstract.** Social media have been firmly entrenched in the modern everyday life. Still, their influence on the formation of public opinion is not well understood. An important feature of social media is that they are not neutral. Not only do people interact with each other on social media platforms, but social media themselves actively interact with people, selecting personalized content for them based on the information about their interests and behavior. In 2011, Eli Pariser hypothesized that content personalization should lead to the formation of a kind of “information cocoons”, or “filter bubbles” — homogeneous groups of users who hold similar views. However, the fragmentation of the Internet community into “filter bubbles” is not the only threat posed by the use of personalization algorithms. Even more dangerously, social media possess the ability to manipulate content selection algorithms in order to influence users’ views.

The article attempts to test the reality of these threats through computational modeling. To solve this task, the author employs a simple agent-based model that simulates the impact of personalization algorithms on communication in social media. The article demonstrates that, contrary to Pariser's hypothesis, algorithms that select content as close as possible to user preferences result in the emergence of "filter bubbles" rather rarely. The author also finds that manipulation of personalization algorithms makes it possible to influence the formation of public opinion on a stable basis only under two conditions: (1) when all users are manipulated and at the same time they are open to external influence; (2) when manipulation aims at the so called "centrists" who do not possess a clear-cut opinion on some issue.

**Keywords:** echo chamber, filter bubble, social media, computational modeling, agent-based model, public opinion, political communications

## References

- 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.)
- Au C.H., K.K.W.Ho, and D.K.Chiu. (2021) "The Role of Online Misinformation and Fake News in Ideological Polarization: Barriers, Catalysts, and Implications" // *Information Systems Frontiers*, April: 1—24. URL: <https://link.springer.com/article/10.1007/s10796-021-10133-9> (accessed on 28.06.2022).
- Bakshy E., S.Messing, and L.Adamic. (2015) "Exposure to Ideologically Diverse News and Opinion on Facebook" // *Science*, vol. 348, no. 6239: 1130—1132.
- Barnidge M. (2017) "Exposure to Political Disagreement in Social Media Versus Face-to-Face and Anonymous Online Settings" // *Political Communication*, vol. 34, no. 2: 302—321.
- Bozdag E. and J.Van Den Hoven. (2015) "Breaking the Filter Bubble: Democracy and Design" // *Ethics and Information Technology*, vol. 17, no. 4: 249—265.
- Bruns A. (2019) "Filter Bubble" // *Internet Policy Review: Journal on Internet Regulation*, vol. 8, no. 4: 1—14. URL: <https://snurb.info/files/2019/Filter%20Bubble.pdf> (accessed on 28.06.2022).
- Cardenal A.S., C.Aguilar-Paredes, C.Galais, and M.Pérez-Montoro. (2019) "Digital Technologies and Selective Exposure: How Choice and Filter Bubbles Shape News Media Exposure" // *The International Journal of Press/Politics*, vol. 24, no. 4: 465—486.
- Chaney A.J.B., B.M.Stewart, and B.E.Engelhardt. (2018) "How Algorithmic Confounding in Recommendation Systems Increases Homogeneity and Decreases Utility" // *Rec.Sys '18: Proceedings of the 12th ACM Con-*

*ference on Recommender Systems*. Vancouver: 224–232. URL: <https://arxiv.org/pdf/1710.11214.pdf> (accessed on 28.06.2022).

Chen W., D.Pacheco, K.C.Yang, and F.Menczer. (2021) “Neutral Bots Probe Political Bias on Social Media” // *Nature Communications*, vol. 12, no. 1: 1–10.

Cinelli M., G. De Francisci Morales, A.Galeazzi, W.Quattrociocchi, and M.Starnini. (2021) “The Echo Chamber Effect on Social Media” // *Proceedings of the National Academy of Sciences*, vol. 118, no. 9: e2023301118.

Flaxman S., S.Goel, and J.M.Rao. (2016) “Filter Bubbles, Echo Chambers, and Online News Consumption” // *Public Opinion Quarterly*, vol. 80, no. S1: 298–320.

Fletcher R. and R.K.Nielsen. (2018) “Automated Serendipity: The Effect of Using Search Engines on News Repertoire Balance and Diversity” // *Digital Journalism*, vol. 6, no. 8: 976–989.

Geschke D., J.Lorenz, and P.Holtz. (2019) “The Triple-Filter Bubble: Using Agent-Based Modeling to Test a Meta-Theoretical Framework for the Emergence of Filter Bubbles and Echo Chambers” // *British Journal of Social Psychology*, vol. 58, no 1: 129–149.

Hameleers M., D.Schmuck, L.Bos, and S.Ecklebe. (2021) “Interacting with the Ordinary People: How Populist Messages and Styles Communicated by Politicians Trigger Users’ Behaviour on Social Media in a Comparative Context” // *European Journal of Communication*, vol. 36, no. 3: 238–253.

Hosanagar K., D.Fleder, D.Lee, and A.Buja. (2014) “Will the Global Village Fracture into Tribes? Recommender Systems and Their Effects on Consumer Fragmentation” // *Management Science*, vol. 60, no. 4: 805–823.

Huber G.A. and N.Malhotra. (2017) “Political Homophily in Social Relationships: Evidence from Online Dating Behavior” // *The Journal of Politics*, vol. 79, no. 1: 269–283. URL: <https://www.journals.uchicago.edu/doi/pdf/10.1086/687533> (accessed on 28.06.2022).

Kitchens B., S.L.Johnson, and P.Gray. (2020) “Understanding Echo Chambers and Filter Bubbles: The Impact of Social Media on Diversification and Partisan Shifts in News Consumption” // *MIS Quarterly*, vol. 44, no. 4: 1619–1649. URL: [https://www.darden.virginia.edu/sites/default/files/inline-files/05\\_16371\\_RA\\_KitchensJohnsonGray%20Final\\_0.pdf](https://www.darden.virginia.edu/sites/default/files/inline-files/05_16371_RA_KitchensJohnsonGray%20Final_0.pdf) (accessed on 28.06.2022).

Kramer A.D.I., J.E.Guillory, and J.T.Hancock. (2014) “Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks” // *Proceedings of the National Academy of Sciences*, vol. 111, no. 24: 8788–8790. URL: <https://digitalwellbeing.org/wp-content/uploads/2014/06/PNAS-2014-Kramer-8788-90.pdf> (accessed on 28.06.2022).

Lewis R. (2018) *Alternative Influence: Broadcasting the Reactionary Right on YouTube*. URL: [https://datasociety.net/wp-content/uploads/2018/09/DS\\_Alternative\\_Influence.pdf](https://datasociety.net/wp-content/uploads/2018/09/DS_Alternative_Influence.pdf) (accessed on 28.06.2022).

McPherson M., L.Smith-Lovin, and J.M.Cook. (2001) “Birds of a Feather: Homophily in Social Networks” // *Annual Review of Sociology*, vol. 27: 415–444.

- Munn L. (2019) “Alt-Right Pipeline: Individual Journeys to Extremism Online” // *First Monday*, vol. 24, no. 6. URL: <https://firstmonday.org/ojs/index.php/fm/article/download/10108/7920> (accessed on 28.06.2022).
- Negroponte N. (1995) *Being Digital*. New York: Knopf.
- Nguyen T.T., P.M.Hui, F.M.Harper, L.Terveen, and J.A.Konstan. (2014) “Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity” // *WWW '14: Proceedings of the 23rd International Conference on World Wide Web*. New York: 677—686.
- Pariser E. (2011) *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. New York: Penguin.
- Roberts M.E. (2020) “Resilience to Online Censorship” // *Annual Review of Political Science*, vol. 23: 401—419.
- Sarup G., R.W.Suchner, and G. Gaylord. (1991) “Contrast Effects and Attitude Change: A Test of the Two-Stage Hypothesis of Social Judgment Theory” // *Social Psychology Quarterly*, vol. 54, no. 4: 364—372.
- Zuiderveen Borgesius F.J., D.Trilling, Ju.Möller, B.Bodó, C.H. de Vreese, and N.Helberger. (2016) “Should We Worry about Filter Bubbles?” // *Internet Policy Review: Journal on Internet Regulation*, vol. 5, no. 1: 1—16. URL: <https://www.econstor.eu/bitstream/10419/214006/1/IntPolRev-2016-1-401.pdf> (accessed on 28.06.2022).